# Investigating individual differences in native and non-native speech sound processing:

# Insights from behaviour, electroencephalography, and neurostimulation


Claire Tamako Honda

Integrated Program in Neuroscience

McGill University, Montreal


September 2024

A thesis submitted to McGill University in partial fulfillment of

the requirements of the degree of Doctor of Philosophy

## Table of Contents

**Abstract**

Speech sounds are the building blocks of spoken language, enabling us to communicate and connect with others. Although understanding speech might feel like a relatively effortless and universal human capacity, speech perception is in fact a complex cognitive process, and adults show reliable individual differences in how they perceive both native and non-native speech sounds. The overarching goal of the current thesis was to better understand the nature of these individual differences and to investigate how to assist adults in learning difficult non-native speech sounds. To this end, we employed a multidisciplinary approach, harnessing behavioural and neurophysiological measures as well as neurostimulation.

In Chapter 1, we begin by characterizing individual differences in performance on various behavioural measures of native and non-native speech perception. This work provides new insight into how individual differences relate across tasks, revealing behavioural response patterns associated with successful speech perception. Specifically, we demonstrate that adults who show more accurate identification of non-native speech sounds also tend to show more consistent responses to native speech sounds across two tasks.

Chapter 2 builds upon this work by studying how the behavioural measures from Chapter 1 might relate to the frequency following response (FFR), a neural measure that represents the consistency of the brain's responses to speech sounds. We strengthen our previous findings by replicating the behavioural results from Chapter 1, and we find

possible (though inconclusive) evidence for a relationship between the consistency of neural and behavioural responses to speech sounds.

Finally, Chapter 3 investigates the potential for a non-invasive neurostimulation technique, transcutaneous auricular vagus nerve stimulation (taVNS), to improve the learning of the non-native speech sounds tested in the previous chapters. Our results do not reveal benefits of taVNS for non-native speech sound learning, though it appears that stimulation may improve the subjective language learning experience by reducing feelings of tension and pressure.

Overall, this research provides novel contributions to our understanding of how adults differ in native and non-native speech perception and of what these differences might signify. It also opens the door for future work to develop ways of making non-native speech sound learning easier and more enjoyable for adults.

**Résumé**

Les sons de la parole constituent les éléments de base du langage parlé, nous permettant de communiquer et de tisser des liens avec les autres. Bien que la compréhension de la parole puisse sembler être une capacité humaine relativement naturelle et universelle, la perception de la parole est en réalité un processus cognitif complexe. En fait, les adultes présentent des différences individuelles fiables dans la façon dont ils perçoivent les sons de la parole, aussi bien les sons natifs que les sons non natifs. L'objectif principal de cette thèse était de mieux comprendre la nature de ces différences individuelles et d'étudier comment on pourrait aider les adultes à apprendre les sons de langues étrangères considérées difficiles à acquérir. Pour ce faire, nous avons utilisé une approche multidisciplinaire exploitant des mesures comportementales et neurophysiologiques, ainsi que de la neurostimulation.

Dans le chapitre 1, nous commençons par caractériser des différences individuelles sur diverses mesures comportementales de la perception de sons de la langue maternelle et d'une langue étrangère. Cette étude permet de mieux comprendre l'impact des différences individuelles sur différentes tâches, en révélant des modes de réponse comportementale associés à une perception réussie de la parole. Plus précisément, nous démontrons que les adultes qui identifient avec plus de précision les sons d'une langue étrangère ont également tendance à donner des réponses plus cohérentes aux sons de leur langue maternelle dans le cadre de deux tâches.

Le chapitre 2 s'appuie sur ces travaux et étudie la manière dont les mesures comportementales du chapitre 1 pourraient être liées à la réponse d'adoption de

fréquence (RAF), une mesure neurale qui représente la cohérence des réponses du cerveau aux sons de la parole. Nous renforçons nos conclusions précédentes en reproduisant les résultats comportementaux du chapitre 1, et nous trouvons des preuves potentielles (bien que non concluantes) d'une relation entre la cohérence des réponses neuronales et comportementales aux sons de la parole.

Enfin, dans le chapitre 3 nous étudions le potentiel d'une technique de neurostimulation non invasive, la stimulation transcutanée auriculaire du nerf vague (STANV), pour améliorer l'apprentissage des sons de langues étrangères utilisés dans les chapitres précédents. Nos résultats ne révèlent pas de bénéfices de la STANV pour l'apprentissage des sons de langues étrangères, bien qu'il semble que la stimulation puisse améliorer l'expérience subjective de l'apprentissage des langues en réduisant les sentiments de tension et de pression.

Dans l'ensemble, cette recherche apporte de nouvelles contributions à notre compréhension de la manière dont les adultes varient dans leur capacité à percevoir la parole dans leur langue maternelle et dans des langues étrangères, et sur l'impact de cette variation. Elle ouvre également la voie à de futurs travaux visant à développer des moyens de rendre l'apprentissage des sons de langues étrangères plus facile et plus agréable pour les apprenants adultes.

**Acknowledgements**

First, I must of course thank Shari Baum and Meghan Clayards for their support throughout this journey. I am fortunate to have had such knowledgeable, responsive, and caring supervisors who were there when I needed guidance and who also gave me space and freedom to develop independently as a researcher. Shari, thank you for replying to all my emails within five minutes—I still don't know how you do it. Meghan, thank you for organizing lab gatherings and for diving into analysis details with me when I was drowning in mixed-effects models. Morgan Sonderegger was also a huge help in deciding which analysis approaches to use and in aiding me to understand both the theory and the R code behind those analyses. I am also very grateful to Annie Gilbert, who trained me on how to use EEG and how to be a rigorous researcher from the first day that I arrived at the lab (and introduced me to various aspects of Quebecois culture). In addition, I extend a big thank you to Denise Klein and Emily Coffey for being on my supervisory committee and for providing encouragement and insightful suggestions during the development of my experiments.

I am also extremely grateful to the Centre for Research on Brain, Language and Music (CRBLM). They have provided EEG equipment, facilitated networking, organized fascinating events at which everyone is well-fed, funded my conference travels, enabled me to take on a leadership role within their Student Engagement Committee, given opportunities to showcase the neuroscience board game that Gaby Von and I created, and more…all of this has enriched my grad student experience on so many levels. Denise, Heather, Inbal, Jasmine, and Amelia, you're the best! Gaby, I'm so so glad that we worked together on our Neurotropolis game, and I really look forward to whatever the future has in store for us (not just as game creators, but also as friends with shared values about promoting interdisciplinary collaboration and empathy). Speaking of interdisciplinary work and empathy promotion, I feel so fortunate to have met Naila Kuhlmann early in my degree and to have become involved in her beautiful Piece of Mind project. Thank you to the whole Piece of Mind team, which is made up of some of the most wonderful humans that I know. Naila, getting to dance and play music in our co-created artsci shows has been a highlight of the past few years and has brought me so much inspiration, hope, discovery, fulfillment, and joy (it's a chocolate-melting-on-the-tongue kind of feeling).

Additionally, I would like to thank the teachers and professors who inspired and supported me along my academic path since high school. Ms. Thompson, thank you for sharing your love of writing and of literature with us. Ms. EA and Mme. Chapiteau, thank you for encouraging me to pursue my passion for languages. Mr. Humphries, thank you for making PW Mini a space in which science was exciting and important, and in which respect and community were fundamental values. Thank you to Anna Taylor for first piquing my interest in psychology. Thank you to Stanka Fitneva, Niko Troje (and his grad student at the time, Sophie Kenny, who was a lovely mentor), Jordan Poppenk, Jiaying Zhao, and Jason Gallivan for enabling me to gain research experience in their labs as an undergrad, which motivated me to pursue research in grad school. Thank you as well to my incredible friends and housemates Julia, Tara, Somi, and Nadia for bringing so much laughter, sharing, growth, and compassion to our high school and undergrad adventures.

I must also express how essential the students and postdocs around me at McGill have been to my success and my perseverance in this program. Their support is largely what has kept me motivated through the good and less-than-good times (including a global pandemic, during which we somehow still had fun with Zoom game nights). So thank you Jasmine, Don, Alvaro, Maël, Angela, Yondu, Jonathan, Max, Alexandre, Alexandra, Haruka, Kaija, Marcos, Mila, Dominique, Nicky, Fernanda, Serte, Pri, Marcelo, Connie, Xuanda, Wei, Alex, Samin, Samira, Zahraa, Luisa, Eve, Stephanie, Nariman, Jovia, Haining, Aaron, Renee, Shuyi, Yufang, Deirdre, Peter, Sarah, Mitch, Josh, Gio, Nils, and Shannon. It is also important for me to recognize the valuable work of the McGill Writing Centre, and particularly of Graphos and their thesis writing support programming. Three Months to Advance Your Thesis and the Monday Meet and Write sessions were hugely helpful as I waded through the creation of this document. Donetta, your people skills and your nurturing energy are incredible, and I will miss getting to see you every Monday and having someone to lead me through stretching breaks.

The international tap community has also kept me going throughout this degree. Tap dance is such a rich art form with such a special close-knit and loving community. Getting to share the joy of percussive movement with dancers from around the world is always refreshing and energizing, and it has helped me to stay balanced through the years.

In addition, I would like to thank Dan for always keeping things in perspective, for grounding me, for encouraging me to care for myself in a holistic way, and for believing that my work was interesting and important even when I did not. Thank you to my dog Snowy for entertaining me and for keeping my lap warm during so many Zoom classes and meetings.

Finally, an especially huge thank you to my family members for their unconditional love and support that have brought me to where I am today and that I am so privileged to carry forward into the next chapter. It is impossible to express my feelings in one or two sentences, but my parents' determination, curiosity, love of learning, rigour, care, compassion, contentiousness, and pursuit of excellence are at the heart of who I am and of why I undertook this thesis.

To quote one of my favourite shows, "What's next?"

## Contribution to Original Knowledge

The present thesis provides a variety of novel contributions to the field. In Chapter 1, we administered two different native speech perception tasks (two-alternative forced choice [2AFC] and visual analog scaling [VAS]) to better understand how adults differ in performance on these tasks and whether individual differences are related across tasks. This is the first work to directly compare performance on 2AFC and VAS tasks by using identical stimuli and identical analysis methods for both tasks. Through direct comparison, we were able to clarify that the two tasks measure separate constructs and that they are related through consistency in responses. In addition, Chapter 1 is the first work to combine native 2AFC and VAS measures with non-native perception measures. In doing so, we revealed that participants who respond more consistently to native speech sounds on a VAS task tend to show better identification of non-native sounds. These results advance our understanding of individual differences in native and non-native speech perception. This chapter has been published in the *Journal of Experimental Psychology: Human Perception and Performance* (2024, Vol. 50, No. 4, 370–394).

In Chapter 2, we replicated the novel findings from Chapter 1, further clarifying that 2AFC and VAS tasks measure separate constructs and that more consistent native speech perception predicts better non-native speech perception. Additionally, we used electroencephalography to record the frequency following response (FFR), a neural measure that can reflect individual differences in speech sound processing. This is the first study to measure individual differences in native perception, non-native perception, and the FFR, and to examine relationships among these three factors. In particular, no other published work to date has explored the potential links between consistent speech sound processing at neural and behavioural levels. We did not find clear evidence that the consistency of neural sound encoding (as measured by the FFR) predicts the consistency of behavioural responses to speech sounds, but this work paves the way for future studies to clarify such relationships and to further investigate the rich construct of consistency. This chapter has been published in *Brain Research* (2024, article 149208).

Chapter 3 describes one of the first studies to investigate transcutaneous auricular vagus nerve stimulation (taVNS) in relation to language learning. taVNS is a relatively novel neurostimulation tool, and recent work has studied its potential benefits for the perception of Mandarin tones (e.g., Llanos et al., 2020). In Chapter 3, we studied the potential effects of taVNS on three as-yet unexplored factors: the perception of unfamiliar non-tonal speech sounds, the production of those same speech sounds, and language learning motivation. Our findings revealed that taVNS may not hold strong benefits for speech sound learning, and that more work is needed in the area of neurostimulation and language learning. This chapter has been published in *Frontiers in Language Sciences* (2024, Vol. 3, article 1403080).

On the whole, the current thesis employs a multidisciplinary and multimodal approach to advance knowledge in the fields of neuroscience, linguistics, psychology, and cognitive science. More specifically, we combine behavioural measures, electroencephalography, and neurostimulation to better understand how speech sounds are processed and acquired in adulthood.

## Contribution of Authors

### Chapter 1

Claire Honda, Meghan Clayards, and Shari Baum jointly conceptualized this study and established the methodology to be used. Claire Honda collected, analyzed, and visualized the data, and wrote the manuscript. Meghan Clayards and Shari Baum provided supervision throughout data collection and analysis, provided funding to remunerate participants, and edited the manuscript.

### Chapter 2

Claire Honda, Meghan Clayards, and Shari Baum jointly conceptualized this study and established the methodology to be used. Claire Honda collected, analyzed, and visualized the data, and wrote the manuscript. Meghan Clayards and Shari Baum provided supervision throughout data collection and analysis, provided funding to remunerate participants, and edited the manuscript.

### Chapter 3

Claire Honda, Neha Bhutani, Meghan Clayards, and Shari Baum jointly conceptualized this study and established the methodology to be used. Claire Honda collected, analyzed, and visualized the data, and wrote the manuscript. Neha Bhutani provided lab space and testing materials. Meghan Clayards and Shari Baum provided funding to remunerate participants. Neha Bhutani, Meghan Clayards, and Shari Baum provided supervision throughout data collection and analysis, and edited the manuscript.

## List of Figures

### Chapter 1



### Chapter 2

**Chapter 3**

# List of Tables

## Chapter 1



## Chapter 2



## Chapter 3

# General Introduction

*When we're brave enough to risk a conversation, we have the chance to rediscover what it means to be human.* — Margaret Wheatley

## Speech Perception

Communication is a core element of the human experience. By communicating—exchanging thoughts and feelings with others—we form rewarding and meaningful human connections that give us insight both into others and into ourselves. One of the most ubiquitous forms of communication is spoken conversation. Most of us engage in conversation with such frequency and ease that listening to speech can seem a deceptively simple task. And yet, speech perception is in fact a highly complex and specialized process. By way of illustration, consider the English language, which comprises approximately 44 different speech sounds (phonemes) that can be combined in nearly infinite ways to create words and sentences (Roach, 2009). These phonemes are often distinguished only by very subtle acoustic differences in timing or frequency; for instance, two phonemes might be perceived as different consonants based on a difference of just a few milliseconds, or as different vowels based on a difference of just a few dozen Hertz (Cruttenden, 2014). Not only this, but each phoneme's acoustic parameters depend on a multitude of factors including the speaker's gender, age, dialect, speaking rate, and emotional state, as well as the phoneme's position within a word and sentence (Fernandez & Cairns, 2010). Speech is therefore a remarkably variable and intricate signal, and in order to successfully perceive it, the listener must engage in a complex interplay of auditory and cognitive processing.

**Non-Native Speech Perception**

The complexity of speech perception becomes more evident in the context of acquiring a new language. Adult learners often struggle to distinguish novel language sounds, to the point that discrimination of non-native phonemes on behavioural tasks is sometimes no better than chance (e.g., Chang et al., 2017; Strange et al., 2011; Strange & Dittman, 1984). For example, it is well documented that native Japanese speakers have difficulty distinguishing between the English consonants /r/ and /l/ (e.g., Hattori & Iverson, 2010; Goto, 1971; Iverson et al., 2003; Miyawaki et al., 1975; Strange & Dittman, 1984). Similarly, native English speakers struggle to discriminate the Hindi consonants /ṭ/ and /t̪/ (e.g., Polka, 1991; Werker & Tees, 1984) and the French vowels /u/ and /y/ (e.g., Gottfried, 1984; Levy & Strange, 2007). These are but three among countless possible examples.

Various theoretical frameworks have been proposed to account for the difficulty that adults experience when learning new language sounds. Among these, three of the most influential are Flege's Speech Learning Model (SLM; Flege, 1995; Flege & Bohn, 2021), Best's Perceptual Assimilation Model (PAM; Best, 1994; Best et al., 2001), and Kuhl's Native Language Magnet Theory (NLM; Kuhl, 1993; Kuhl et al., 2008). While these models differ somewhat in their focus and their details, they share the general prediction that the difficulty involved in learning a given non-native phoneme will depend on the perceptual similarity between that phoneme and native phonemes. In the framework of the SLM, native and non-native phonemes exist in a shared phonological space, and non-native phonemes are harder to learn when they occupy a part of the phonological space that is already taken up by a native category (Flege, 1995). The SLM also predicts that as age of learning increases, it becomes more difficult for learners to distinguish non-native phonemes from native ones (Flege, 1995). The PAM, for its part, outlines different patterns whereby non-native phonemes can be perceived with relation to native phonemic categories (Best, 1994). It posits that two non-native phonemes are harder to tell apart when they are both perceived as belonging to a single native category; they are easier to tell apart when they are perceived as belonging to two different native categories; and they are easiest to tell apart when they are perceived as new phonemes that are distinct from native categories (Best, 1994). Finally, the NLM proposes that native phonemic categories act as "perceptual magnets" that attract non-native phonemes, making it hard to form a new non-native category when a phoneme resembles a native category (Kuhl, 1993).

These models have seen empirical support from many sources. For example, consistent with the SLM's prediction that non-native phonemes become harder to tell apart from native phonemes as age increases, research has shown that Korean adults are more likely than Korean children to perceive non-native (English) vowels as being similar to Korean vowel categories (Baker et al, 2002). Other work compared native English speakers' perception of Zulu and Tigrinya phonemes that differed in how much they resembled English phonemes (Best et al., 2001). In line with the PAM's predictions, discrimination of non-native phoneme pairs was best when the two phonemes within a pair were perceived as highly distinct from native phonemic categories; it was less good when the two phonemes were perceived as resembling two native phonemic categories; and it was worst when the two phonemes were perceived as resembling a single native category (Best et al., 2001). Such models also help to account for the previously mentioned findings about adult learners' difficulties with certain non-native phonemes. Specifically, native Japanese speakers tend to perceive both English /r/ and /l/ as the Japanese consonant /ɾ/ (Miyawaki et al., 1975), and native English speakers tend to perceive both Hindi /ṭ/ and /t̪/ as English /t/ (Werker & Tees, 1984). In these cases, the two non-native phonemes are therefore "attracted" by a single native phoneme (as in the NLM) and are both perceived as members of that native phonemic category (as in the PAM), with the result that these non-native phonemes are not readily discriminated.

Overall, then, theoretical models and empirical evidence both underline that non-native phonemic perception is generally a complex and challenging process for adults. The successful perception of non-native phonemes is difficult, but also vital, because it acts as a foundation for broader language learning and communication. This notion is evidenced by work showing that adults who can more successfully differentiate non-native phonemes are also more successful at learning words containing those same phonemes (Silbert et al., 2015). Indeed, phonemes are the basic units that make up spoken language; when non-native phonemes are inaccurately perceived, it follows that the perception of words and sentences containing those phonemes will also be compromised, potentially leading to breakdowns in communication. In order to support adults' successful communication in languages other than their native one(s), it is therefore of fundamental importance to study non-native phonetic perception and the ways in which it might be improved.

## Speech Perception and the Brain

Speech and language, like all other complex cognitive processes, are based in the brain. When speech sounds reach the ear, the cochlea converts them into neural signals that travel first to the cochlear nuclei and inferior colliculus in the brainstem, then to the medial geniculate nucleus in the thalamus, and finally to the auditory cortex and higher-order cortical areas that process linguistic information (Kandel, 2013). As with all brain activity, the neural signals associated with speech sound processing can be measured in a variety of ways, including using functional magnetic resonance imaging (fMRI) or electroencephalography (EEG).

One particularly interesting and rich measure of speech processing in the brain is the frequency following response (FFR). The FFR is a brainwave generated by the auditory system in response to a periodic sound such as a vowel (Skoe & Kraus, 2010). As suggested by its name, the FFR "follows" the frequency of the sound that elicited it. To illustrate this, picture a vowel such as /a/, with a fundamental frequency of 100 Hz. If this vowel is repeatedly presented to a participant while their neural activity is recorded with EEG, the resulting FFR will visually resemble the soundwave of the vowel and will have the same fundamental frequency of 100 Hz (Skoe & Kraus, 2010). In fact, the FFR follows the frequencies of the eliciting stimulus so well that if it is converted from a brainwave to a soundwave and is then played back to listeners, it can be heard as intelligible speech (Galbraith et al., 1995). While it was originally thought to be generated solely by the brainstem, the FFR has now been demonstrated to result from neural activity along the entire auditory pathway, from the cochlear nuclei all the way up to the cortex (Coffey et al., 2019). The FFR is thus an aggregate measure of auditory processing, and it provides an index of the robustness of neural sound encoding. The more faithfully and consistently a person's FFR represents the frequencies of the eliciting stimulus, the more robust that person's auditory processing (Krizman & Kraus, 2019).

Even newborn infants show robust FFRs in response to speech sounds (Ribas-Prats et al., 2019), and they also show greater neural activity in language-related cortical areas when listening to speech compared to nonspeech sounds (Peña et al., 2003), suggesting that the human brain is well-equipped to process speech from birth. Indeed, infants can effectively learn speech sounds in an unguided way through exposure (e.g., Maye et al., 2008). In contrast, adults commonly struggle to learn non-native speech sounds as mentioned above, and sound learning in adulthood can be assisted by explicit training (e.g., Iverson et al., 2005).

At the neural level, why might adults face challenges in learning new speech sounds? While the brain does remain plastic across the lifespan, it is shaped by experience, and plasticity becomes reduced with age (Burke & Barnes, 2006). As a native language is acquired in childhood, neural networks become specialized to process that language, which Kuhl (2004) frames as "native language neural commitment". Such neural commitment promotes the efficient processing of one's native language, but it can interfere with the acquisition of other languages that do not follow the same patterns as the native language (Kuhl, 2004). In favour of this hypothesis, adults show neural activity over a larger area and over a longer time period when processing non-native compared to native speech sounds, indicating neural inefficiency (Zhang et al., 2005). In general, non-native language processing seems to entail less efficient allocation of neural resources than native language processing (see Stowe & Sabourin, 2005, for a review). It is important to try to overcome these neural barriers so that people can learn new languages more optimally at any age.

Related to the idea of native language neural commitment, another proposition is that there are critical or sensitive periods for the establishment of certain neural systems, beyond which those systems can no longer be shaped (Werker & Hensch, 2015). While there is strong evidence for sensitive periods in some cases (e.g., the establishment of ocular dominance columns in the visual system; Wang et al., 2010), sensitive periods for language development are more controversial and complex (see Werker & Hensch, 2015, for a review). It does appear that there are certain time windows during which certain aspects of language learning most readily develop, but these windows are not necessarily fixed—and importantly, it may be possible to reopen them (Werker & Hensch, 2015).

One possible manner of reopening putative sensitive periods and of increasing the brain's plasticity in adulthood is using neurostimulation (Hogan et al., 2020; Werker & Hensch, 2015). Various neurostimulation techniques exist, including transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS), and vagus nerve stimulation (VNS; Frangos et al., 2015; Kricheldorff et al., 2022). These techniques are capable of increasing the plasticity of the adult auditory cortex (Boroda et al., 2020; Engineer et al., 2011; Engineer et al., 2015; Lorenz et al., 2010). Increased plasticity can in turn lead to enhanced neural processing—for example, one study found that when VNS was paired with speech sounds, the neural response to those sounds became faster, stronger, and less variable (Engineer et al., 2015).

VNS has traditionally been administered through surgically implanted electrodes, making its use limited to animal studies (e.g., Engineer et al., 2015) and clinical populations (e.g., Austelle et al., 2021). However, recent advances have resulted in the development of transcutaneous auricular vagus nerve stimulation (taVNS), a non-invasive version of VNS in which stimulation can be delivered quickly and conveniently through electrodes placed against the skin of the outer ear (e.g., Badran et al., 2018). Other neurostimulation techniques such as TMS require costly and cumbersome equipment, and both TMS and tDCS require targeting specific cortical regions; in comparison, taVNS is accessible, easy to administer, and can modulate the activity of multiple cortical and subcortical areas, making it a particularly promising technique (e.g., Frangos et al., 2015).

There is preliminary evidence that taVNS may improve speech sound learning in adults. In particular, there are reports that administering taVNS alongside Mandarin lexical tone training can improve perception of the trained tones (Llanos et al., 2020; McHaney et al., 2023; Pandža et al, 2020, Phillips et al., 2021). Although these findings are promising, taVNS has yet to be explored as a potential tool for improving the learning of speech sounds beyond Mandarin lexical tones. If taVNS can in fact increase neuroplasticity and assist adults in acquiring new phonemes, then this accessible neurostimulation technique could provide an exciting avenue for optimizing language learning.

## Individual Differences in Speech Perception

As described above, it is generally challenging for adults to perceive non-native phonemes. Nevertheless, some adults show much more successful performance on non-native speech perception tasks than others (e.g., Hanulíková et al., 2012; Hattori & Iverson, 2010; Lengeris & Hazan, 2010). For instance, identification or discrimination of non-native phonemes can range all the way from near-perfect to near-chance even within a sample of healthy young adults (Hanulíková et al., 2012; Lengeris & Hazan, 2010). Some factors have been found to relate to the success of non-native perception, including musical aptitude (Slevc & Miyake, 2006), native language background (Flege et al., 1997), age of non-native language acquisition (Stölten et al., 2013), and extent of non-native language experience (Flege & MacKay, 2004). However, even when attempting to control for these factors (for instance, by studying participants with the same native language background or age of non-native language acquisition), individual differences are still found (e.g., Hanulíková et al., 2012; Lengeris & Hazan, 2010); it is still unclear why such wide-ranging differences in non-native perception exist.

Although these differences in non-native phonetic perception are striking, it is perhaps even more surprising that healthy young adults also show differences in *native* phonetic perception. Such differences can be measured by presenting listeners with a continuum of native speech sounds and asking them to indicate what they heard at each step of the continuum. One common paradigm is the two-alternative forced choice (2AFC) task, in which listeners indicate what they heard by selecting one of two response options; and another more recent paradigm is the visual analog scaling (VAS) task, in which listeners indicate what they heard by selecting a point along a continuous line between two options. By plotting a listener's average response at each step of the continuum against the actual changes in the stimulus, it is possible to obtain an identification slope that shows how gradually or suddenly the listener's percept shifts in relation to changes in the stimulus. On these two tasks, adults differ in how steep or shallow their identification slopes are (e.g., Kapnoula et al., 2017; Kong & Edwards, 2016; Ou & Yu, 2022). That is, when presented with a continuum—for example, ranging from *den* to *ten*—some people perceive gradual, gradient changes from /d/ to /t/ and others perceive sudden, categorical changes (Kapnoula et al., 2017). Adults also differ in the consistency of their responses to such continua. It has been shown that some people consistently give the same response to a stimulus across trials of a VAS task, for example always labelling a given stimulus as /d/-like; and others perceive the stimulus inconsistently, labelling it as /d/-like on some trials and /t/-like on others (Fuhrmeister et al., 2023; Kapnoula et al., 2017).

As with differences in non-native perception, the origins and functions of differences in native perception remain largely uncertain. Interestingly, it is not clear whether listeners' identification slopes are related across 2AFC and VAS tasks. While VAS identification slopes seem to reflect the gradiency of perception (i.e., to what extent the listener perceives gradual, fine-tuned phonetic differences between speech sounds along the continuum), there is tentative evidence that 2AFC slopes do not relate to them—in which case 2AFC tasks may be tapping into a construct other than gradiency (Kapnoula et al., 2017). There is a need for systematic comparison of responses across the two tasks in order to better understand what these individual differences may reflect. Additional insight may be gained by investigating how these differences are linked to other linguistic and cognitive measures. For instance, individual differences in the gradiency of responses on VAS tasks may relate to differences in executive function, though this effect does not seem to be large or reliable (Kapnoula et al., 2017; Kapnoula & McMurray, 2021). For their part, individual differences in the consistency of native perception have not yet been significantly linked to other factors. Differences in gradiency and consistency may actually be related, with more gradient listeners tending to respond more consistently to native sounds as measured by a VAS task, but evidence so far is inconclusive (Kapnoula et al., 2017). It is necessary to explore these differences and their potential relationships in greater detail so that we can better characterize different listener profiles and understand what makes certain listeners particularly successful at perceiving speech sounds.

A question that naturally arises is whether individual differences in native speech perception might predict differences in non-native perception. The previously discussed theories of non-native phonetic perception (SLM, PAM, and NLM) do not directly address individual differences, but they do focus on how native categories influence non-native perception, and their logic can be extended to make predictions about individual variability in perception. Specifically, adults with more categorical perception of native phonemes might show worse non-native perception because their native categories act as stronger magnets and subsume non-native sounds, preventing discrimination between native and non-native categories. On the other hand, adults with more gradient native perception might be less likely to assimilate non-native sounds into their native categories and might better perceive the subtle acoustic differences needed to distinguish between new speech sounds. During the experimental design and data collection phase for the current thesis, no published studies had yet examined the potential links between these individual differences in native and non-native speech perception. One recent paper has since been published on the topic (Fuhrmeister et al., 2023). The authors measured the gradiency and consistency of participants' native perception using a VAS task, and related this to participants' accuracy at identifying and discriminating difficult non-native sounds. They did not find relationships between gradient native perception on the VAS task and successful non-native perception, but they did find that more *consistent* VAS responses were linked to better non-native discrimination (Fuhrmeister et al., 2023). These results suggest that the precise perception of native sounds may promote the accurate perception of non-native sounds; however, more work is needed to determine whether the findings can be generalized beyond this single study. On the whole, it is uncertain whether, and how, individual differences in native perception might predict differences in non-native perception.

Another understudied aspect of individual differences in speech perception is the neural correlates of these differences. Variation in brain structure seems to account in part for the previously discussed differences in perception: one study found that more gradient native perception on a VAS task was predicted by reduced surface area in part of the right frontal lobe, while more consistent native perception was predicted by reduced structural complexity bilaterally in part of the temporal lobe (Fuhrmeister & Myers, 2021). Other work has demonstrated that better discrimination of native and non-native sounds relates to reduced white matter volume in a right insulo/fronto-opercular region (Sebastián-Gallés et al., 2012) and that more successful learning of non-native sounds relates to greater white matter volume in left primary auditory cortex (Golestani et al., 2007).

While differences in native and non-native perception are therefore predicted to some extent by brain structure, it is less clear how they might be predicted by patterns of brain activity. Using EEG, one study has found that participants with less negative N1 responses (reflecting early cortical processing of sound) tend to perceive native speech more gradiently on a VAS task (Kapnoula & McMurray, 2021). Another has demonstrated that participants with better perception of a non-native contrast show larger mismatch negativity responses to changes in native and non-native sounds, indicating better auditory change detection (Díaz et al., 2008). However, these measures of cortical auditory processing are generally not well suited to the study of individual differences due to their variability; consequently, it is common to split participants based on their performance (e.g., better/worse non-native perception, as in Díaz et al., 2008) and to conduct group-level rather than individual-level analyses.

Unlike many other EEG measures, the FFR is exceptional because it differs across participants but shows robust test-retest reliability within participants, making it appropriate for use in individual-level analyses (Song et al., 2011). Indeed, individual differences in the FFR have successfully been linked to differences in performance on behavioural tasks such as speech-in-noise perception (Parbery-Clark et al., 2011) and pitch perception (Krishnan et al., 2012). An additional benefit of the FFR is that it can be recorded with as few as three electrodes (Skoe & Kraus, 2010), making its acquisition rapid and accessible compared to most other EEG measures that typically require dozens of electrodes. As alluded to earlier, the FFR can even be recorded in newborns and has been proposed as an informative supplement to existing infant hearing screenings (Ribas-Prats et al., 2019). Considering all of this evidence, we can conclude that the FFR is a reliable measure of individual differences in auditory processing and that it could eventually be harnessed as a straightforward and useful pre-screening tool for assessing language perception and learning ability.

The FFR may be at the source of some of the individual differences in native and non-native phonetic perception mentioned above. One relevant study has examined the FFR in relation to native phonetic perception on a 2AFC task, finding that participants with more faithful FFRs (i.e., with brainwaves that more closely resemble the soundwaves of the eliciting stimuli) tend to have less categorical 2AFC responses (Ou & Yu, 2022). This result demonstrates the value of studying the FFR as a neural correlate of individual differences in native speech perception. However, 2AFC tasks are not necessarily informative in isolation (as discussed further in Chapter 1), and the authors did not administer any other speech perception tasks (e.g., VAS or non-native perception). Other studies have found relationships between differences in the FFR and differences in non-native speech sound perception, though these relationships were not coherent across studies and emerged only for certain FFR measures and certain non-native sounds (Kachlicka et al., 2019; Omote et al., 2017; Saito et al., 2018). Overall, relationships among the FFR, native speech perception, and non-native speech perception have yet to be explored in depth, and no study has yet related the FFR to both native and non-native perception.

**The Present Work**

Overall, then, the current thesis set out to address three main questions. We first asked: What are the relationships among measures of individual differences in native and non-native speech perception? In particular, we were interested in the relationships between 2AFC and VAS identification slopes and between native gradiency and consistency, as well as whether these measures of native perception relate to the success of non-native perception. Chapter 1 investigated this by employing a variety of behavioural speech perception measures and exploring how the measures were related within individuals. Next, we investigated the neural activity patterns that might contribute to individual differences in speech perception. Specifically, do the individual differences identified in Chapter 1 relate to neural sound encoding as measured by the FFR? This question is the topic of Chapter 2, which employed a combination of behavioural tasks and electroencephalography to study the FFR as a possible neural source of individual differences in native and non-native speech perception. Finally, given the challenge of learning new non-native phonemes in adulthood, Chapter 3 considered how the learning process might be facilitated. Namely, we investigated whether taVNS—a relatively novel and accessible neurostimulation technique—can improve adults' learning of non-native phonemic categories.

The overarching goal of this thesis was therefore to uncover some of the mysteries surrounding how and why adults differ in their speech perception (as measured both behaviourally and neurally), as well as to explore how the acquisition of new speech sounds can be facilitated in adulthood. We hope that eventually, such work will lead to the implementation of measures and techniques to make language learning more accessible and successful for everyone. This topic is especially important in today's era of human migration and multiculturalism, where languages are keys that connect us to new people and opportunities.

## Chapter 1: Exploring Individual Differences in Native Phonetic Perception and Their Link to Non-Native Phonetic Perception

### Author Note

All data from these experiments are publicly available on the Open Science Framework (OSF) and can be accessed at https://osf.io/ez5qh/?view_only=8e4a1498e04f4ee0946752ee93b9ce71. The hypotheses, methods, and analyses for Experiment 2 were preregistered and are available at the same link.

The authors have no conflict of interest to declare.

This work was supported by a grant awarded by the Natural Sciences and Engineering Research Council of Canada (NSERC) to Shari R. Baum, a grant awarded by the Social Sciences and Humanities Research Council (SSHRC) to Meghan Clayards, and NSERC Canada Graduate Scholarship-Master's (CGS M) and Postgraduate Scholarship-Doctoral (PGS D) grants along with a McGill Faculty of Medicine Max Binz Fellowship awarded to Claire T. Honda.

Correspondence concerning this article should be addressed to Claire T. Honda, Integrated Program in Neuroscience, McGill University, 2001 McGill College Avenue, 8th floor, Montreal, QC, H3A 1G1, Canada. Email: claire.honda@mail.mcgill.ca.

### CRediT Authorship Contribution Statement

Claire T. Honda: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Meghan Clayards: Conceptualization, Methodology,

Supervision, Writing – review & editing. Shari R. Baum: Conceptualization, Methodology, Supervision, Writing – review & editing.

**Abstract**

Adults differ considerably in their perception of both native and non-native phonemes. For instance, when presented with continua of native phonemes on 2-alternative forced choice (2AFC) or visual analog scaling (VAS) tasks, some people show sudden changes in responses (i.e., steep identification slopes) and others show gradual changes (i.e., shallow identification slopes). Moreover, some adults are more successful than others at learning unfamiliar phonemes. The predictors of these individual differences and the relationships between them are poorly understood. It also remains unclear to what extent different tasks (2AFC vs. VAS) may reflect distinct individual differences in perception. In two experiments, we addressed these questions by examining the relationships between individual differences in performance on native and non-native phonetic perception tasks. We found that shallow 2AFC identification slopes were not related to shallow VAS identification slopes but were related to inconsistent VAS responses. Additionally, our results suggest that consistent native perception may play a role in promoting successful non-native perception. These findings help characterize the nature of individual differences in phonetic perception and contribute to our understanding of how to measure such differences. This work also has implications for encouraging successful acquisition of new languages in adulthood.

**Public Significance Statement:** Successfully perceiving speech sounds is a crucial skill for spoken communication; yet individuals show differences in how they perceive both native and non-native speech sounds. We studied the relationships between performance on different native and non-native speech perception tasks, finding that (a) different tasks measure different subtleties and (b) people with consistent perception of native speech sounds tend to be better at accurately perceiving non-native sounds. These findings have implications for understanding the nature of individual differences in speech perception and for helping adults to learn new languages successfully.

**Keywords:** phonetic perception, non-native perception, individual differences, gradient perception, consistency

## Introduction

It is well established that there are individual differences in speech perception. Even healthy young adults show differences in how they perceive speech in their native language. For example, some people have better perception of speech in noise compared to others (Surprenant & Watson, 2001). Similarly, some people show greater perceptual plasticity, i.e., an increased ability to successfully adapt their perception to changes in speaking rate or accent (Heffner & Myers, 2021). People also differ in the extent to which their speech perception is affected by different factors, such as coarticulation (Yu & Lee, 2014) or visual information about the speaker's mouth movement (as in the McGurk effect; Strand et al., 2014). Differences even in phonetic perception—an elemental building block of higher-level speech perception—have been documented for decades, such as differences in the categorization of stops (Hazan & Rosen, 1991) and in the discrimination of sibilants (Perkell et al., 2004).

While it is interesting to note these differences, current research is attempting to understand what underlies them, and in doing so to better understand speech perception. For example, some researchers have proposed that differences in basic auditory processing play a role (Cumming et al., 2015; Won et al., 2016). Others have found links between differences in executive function and in speech perception (Kapnoula et al., 2017; Kim et al., 2020). Such studies help us to understand the broader architecture of speech perception. One goal of the current paper is to better understand sources of individual differences as measured by two speech perception tasks – two-alternative forced choice (2AFC) and visual analog scaling (VAS) – by comparing different measures of performance on each one across a large sample of participants. In doing so, we test the hypothesis that the tasks each reflect distinct individual differences in speech sound perception.

In addition to the individual differences in native phonetic perception described above, there are differences in non-native phonetic perception. Adult learners of non-native phoneme contrasts show great variability in performance, with some successfully distinguishing contrasts and others having great difficulty even after receiving training and feedback (e.g., Bradlow et al., 1997; Hanulíková et al., 2012; Hattori & Iverson, 2010; Strange & Dittman, 1984). Non-native perception has been shown to depend in part on numerous factors, including native language background (e.g., Flege et al., 1997), musical ability (e.g., Slevc & Miyake, 2006), and auditory acuity measures such as temporal processing (Kempe et al., 2012) or formant and pitch discrimination (Kachlicka et al., 2019); however, the impact of these factors seems to depend on the particular non-native sounds being perceived, and accounts for only a portion of the variation in performance. As such, the predictors of successful non-native phonetic perception remain relatively poorly understood. The second goal of this paper is to test whether differences in native speech sound perception predict discrimination of difficult second language sound contrasts encountered for the first time.

**Individual Differences in Native Speech Perception**

One of the most ubiquitous methods for measuring phonetic perception is using 2-alternative forced choice (2AFC) tasks. In these tasks, participants are generally presented with a continuum of speech stimuli (e.g., ranging in small steps from *bet* to *bat*) and must classify each stimulus into one category or the other. When a participant's average response to each stimulus is plotted against actual changes in stimulus properties, the result yields an identification slope that can range from shallow to steep (indicating that responses are changing gradually/sharply with changes in stimuli).

There are differences between people in terms of how shallow or steep their identification slopes are. Shallower slopes on 2AFC tasks have previously been linked to various language impairments (Manis et al., 1997; Joanisse et al., 2000; Serniclaes et al., 2001; Werker & Tees, 1987) and to illiteracy (Serniclaes et al., 2005), and have accordingly been considered to reflect an unsuccessful and undesirable pattern of perception compared to steeper slopes. Shallow slopes have been thought to reflect poorly defined boundaries between phonemic categories, potentially due to enhanced discrimination within categories (Serniclaes et al., 2001), whereas steep slopes have been thought to reflect sharply defined boundaries between categories. The association between shallow slopes and language impairment has therefore led to the suggestion that sensitivity to within-category, sub-phonemic detail can be maladaptive. However, it is not clear whether shallow 2AFC slopes actually reflect fine-grained, within-category sensitivity. Instead, they might reflect an inconsistent ability to perceive or categorize sounds (Kapnoula et al., 2017; Serniclaes et al., 2001). Thus, it may be erroneous to relate within-category sensitivity to impairment (see Kapnoula et al., 2017 and Apfelbaum et al., 2022 for other examples of this point). In contrast to shallow slopes, steep slopes on 2AFC tasks are often assumed to indicate categorical perception, which has been proposed as an effective solution to the problem of how continuous cues in the acoustic signal are mapped onto discrete categories during perception (Liberman et al., 1957).

**Limits of a Categorical View of Perception**

Empirically, categorical perception refers to the observation that (1) when presented with a continuum that ranges in equal steps from one category to another, people tend to perceive a sharp distinction between categories (i.e., a steep identification slope as described above); and that (2) stimuli belonging to the same category are often discriminated more poorly than equivalently distant stimuli that cross a category boundary (Liberman et al., 1957). This finding has led to the theoretical view that our perceptual representations are warped based on our top-down knowledge of categories, facilitating processing (Goldstone & Hendrickson, 2010).

Despite the fact that the theory of categorical perception has been hugely popular and influential, there is also widespread evidence challenging it (see McMurray, 2022 for a review). Even from the early days of its proposal, categorical perception was not observed for all speech sounds (Fry et al., 1962) and there was evidence that task demands were at least partly responsible for the phenomenon (Pisoni & Tash, 1974; Hary & Massaro, 1982). Since then, work using behavioural, eye-tracking, and neurophysiological techniques has led to a growing consensus that auditory encoding and speech perception are in fact inherently gradient (McMurray et al., 2008; McMurray et al., 2002; Miller, 1994; Ou & Yu, 2022; Toscano et al., 2010). Gradient perception refers to an ability to distinguish gradual, fine-tuned phonetic differences rather than sudden phonemic ones as in categorical perception. For example, using category goodness ratings, Miller (1994) demonstrated that phonetic categories have a gradient and context-dependent structure; some stimuli are perceived as better exemplars of a category than others, and this perception can flexibly change when relevant contextual factors (e.g., speech rate, syllable structure, or lexical status) are altered. In a similar vein, McMurray and colleagues found that identification slopes became less steep (more gradient) when words were used instead of meaningless syllables, when pictures were used instead of letters, and when four alternatives were used instead of two (McMurray et al., 2008). Thus, the common finding of steep (or "categorical") slopes on 2AFC tasks being associated with successful perception may stem largely from task demands; after all, the task requires a categorical response, so it is natural for it to elicit categorical-looking response patterns in successful listeners. Given all of this evidence, it is useful to note that *categorization of speech sounds* is a necessary process to derive meaning from the speech signal and does not preclude gradient perception, while *categorical perception* (as a theoretical view involving perceptual warping) is not necessary to explain the patterns of responses that have been observed on tasks such as 2AFC.

**A Less Categorical Measurement: The VAS Task**

Unlike 2AFC tasks which elicit a categorical decision, visual analog scaling (VAS) tasks require the participant to indicate what they heard along a continuous line between two options (Massaro & Cohen, 1983). VAS tasks provide a valuable alternative to 2AFC tasks for a variety of reasons. For instance, VAS tasks appear to have superior psychometric properties to 2AFC tasks. Munson et al. (2017) found that fricative ratings on a 2AFC task differed in the extent to which they were influenced by particular acoustic cues, depending on whether the 2AFC ratings were interleaved with more continuous ratings (gender typicality of speech) or more categorical ratings (which category the adjacent vowel belonged to, among 5 options); in contrast, ratings of the same stimuli on a VAS task did not differ depending on these biasing conditions. VAS ratings therefore seem to be less influenced by concurrent tasks (Munson et al., 2017).

Critically, VAS tasks may be better suited to studying the phenomenon of gradient vs. categorical perception; they enable responses that are closely related to the acoustic characteristics of speech (Apfelbaum et al., 2022; Massaro & Cohen, 1983; Munson et al., 2012) and that correlate with continuous measures of production (Schellinger et al., 2017). Using a VAS task, Kong & Edwards (2016) found clear differences between participants' response patterns (some participants had more gradient-looking responses, others were more categorical), showing the task's potential as an alternative to the 2AFC format for studying individual differences in phonetic perception.

**Relationships Between Individual Differences in 2AFC and VAS**

Although they have been much studied, 2AFC slopes on their own are not very informative for reasons described below. However, by comparing data from both 2AFC and VAS tasks, it is possible to better understand the nature of the individual differences underlying different response patterns on these two tasks. Kapnoula et al. (2017) did precisely this, comparing participants' identification slopes on 2AFC and VAS tasks. They found that the slopes on the two tasks were not related within participants, suggesting that the tasks do not measure the same construct (Kapnoula et al., 2017). Note that they used different ways of estimating slopes for the two tasks, an issue we will return to later. Furthermore, Kapnoula et al. (2017) measured how consistently participants responded on the VAS task by calculating the difference between a given participant's actual response on each trial and their predicted response based on their VAS identification slope, and then calculating the standard deviation of these residuals for each participant. Interestingly, they found that shallower 2AFC slopes were marginally related to less consistent VAS responses. Their interpretation was that a shallow 2AFC slope may reflect inconsistent perception of speech sounds rather than actual gradiency of perception (Kapnoula et al., 2017).

To illustrate these findings and the limitations of 2AFC slopes, consider a listener with very gradient perception—that is, with fine-tuned sensitivity to within-category differences between sounds. When presented with a 2AFC task, such a listener might show very different identification slopes depending on their response strategy. One strategy would be to categorize the sounds consistently based on whichever response option they more closely resemble, which would result in a steep identification slope (sharp distinction between categories). Another strategy would be to respond probabilistically by matching the proportion of their two responses to the degree that the sound matches the two alternatives, which would result in a shallow identification slope (Clayards et al., 2008). On the 2AFC task, two very different identification slopes can thus arise from the same underlying perception of speech sounds. Furthermore, a shallow slope on the 2AFC task could arise due to two possibilities: the participant could have more signal-driven, gradient perception and be responding probabilistically as just described, or they could have more category-driven perception but be responding in a noisy and inconsistent way. These possibilities cannot be disambiguated without additional information from another task.

Now consider how the same listener with more gradient perception would respond on a VAS task. Unlike for the 2AFC task, there would be no ambiguity; the listener would show a shallow identification slope. Similarly, the VAS task can distinguish between whether the listener's perception is truly gradient—evidenced by a shallow slope—or in fact inconsistent—evidenced by dissimilar ratings for the same stimulus across trials. By comparing participants' slopes and consistency across the two tasks, it is therefore possible to determine whether 2AFC slopes reflect gradiency or consistency of perception, and whether a given participant's perception is more gradient or more categorical. In finding that 2AFC slopes were weakly related to VAS consistency but not to VAS slopes, the work by Kapnoula et al. (2017) provides preliminary evidence that 2AFC slopes may tap more into the construct of consistency whereas VAS slopes tap more into the construct of gradiency.

The relationship between shallow 2AFC slopes and inconsistency of perception provides a potential explanation for why the previously mentioned studies have linked shallow 2AFC slopes to language impairment. Thus, it is potentially problematic to use the term *gradient* when referring to shallow 2AFC slopes or to associate the concept of gradient/less-categorical perception with impairment (e.g., Manis et al., 1997; Werker & Tees, 1987) when the true issue may lie in inconsistent perception. For this reason, we will refer to identification slopes as being shallow or steep—terms that do not assume a direct association between slope and the construct of gradiency/categoricity—rather than gradient or categorical. Because these terms have unbiased interpretations and facilitate comparisons of results across tasks, we will often use them to refer to slopes derived both from 2AFC and from VAS tasks. This being said, VAS tasks naturally allow for a continuous/gradient form of responding that is more likely to reflect true gradiency compared to 2AFC responses (Apfelbaum et al., 2022), so we will occasionally follow previous work in referring to measures of gradiency when such measures have been derived from VAS tasks. Note, however, that some authors use gradiency to also refer to shallow 2AFC or 4AFC slopes (e.g., Ou et al., 2021; Ou & Yu, 2022).

**The Nature and Potential Functions of Gradiency**

Gradiency (as measured by VAS tasks) appears to be a relatively consistent property of the individual. It has been shown to be related across different testing sessions using the same stimuli (Kong & Edwards, 2016), across different contrasts (Fuhrmeister & Myers, 2021; Kapnoula & McMurray, 2021; but see Kapnoula et al., 2021 for contrasting evidence), and across native and non-native perception (Kong & Kang, 2022). Individual differences in gradiency may reflect anatomical differences in auditory processing architecture, since they relate to differences in cortical surface area (Fuhrmeister & Myers, 2021) and in how cues are neurally encoded and transformed along the auditory pathway (Kapnoula & McMurray, 2021; Ou & Yu, 2022).

Interestingly, various lines of evidence point to the idea that gradiency may not be an indicator of unsuccessful perception as previously thought. For instance, gradiency can reflect experience-related sensitivity to fine acoustic detail, with trained speech-language pathologists giving VAS ratings that are more closely related to acoustic characteristics of the signal compared to inexperienced listeners (Munson et al., 2012). In addition, more gradient VAS responses have been associated with an increased ability to integrate multiple acoustic cues in the speech signal (Kapnoula et al., 2017; Kapnoula & McMurray, 2021; Kim et al., 2020; Kong & Edwards, 2016; Kong & Kang, 2022). Gradiency thus relates to the ability to integrate multiple acoustic cues and to perceive fine-tuned changes in those cues, which appears to encourage perceptual flexibility in the face of ambiguous input (Clayards et al., 2008; Desmeules-Trudel & Zamuner, 2019).

In line with the notion that gradiency promotes perceptual flexibility, Kapnoula et al. (2021) found that listeners with shallower VAS slopes showed greater recovery from lexical garden paths during an eye-tracking task. For example, when presented with a stimulus such as *pumpernickel* in which the initial consonant had been manipulated to sound ambiguous between [p] and [b], such listeners were more likely to switch their gaze from a competitor item (*bumpercar*) to the appropriate target item compared to listeners with steeper VAS slopes (Kapnoula et al., 2021). In other words, by being sensitive to fine-grained acoustic details, the gradient listeners were more readily able to reconsider and flexibly adjust their initial interpretation of misleading stimuli. Further support comes from work that has demonstrated a relationship between inhibitory control and gradiency (Kapnoula et al., 2021). Greater inhibitory control appears to promote gradiency by enabling listeners to manage ambiguous input that activates competing phonemic representations, thus granting listeners greater perceptual flexibility (Kapnoula et al., 2021). The flexibility afforded by gradiency could have a range of benefits given that flexible perception is useful for adapting to variation in both native and non-native speech (Heffner & Myers, 2021)—successful listeners must constantly adapt to differences in the speech signal that arise from numerous factors such as speaking rate, coarticulation, speaker gender, and accent.

**Individual Differences in Non-Native Speech Perception**

As discussed above, adult learners of non-native phoneme contrasts show great variability in performance, and this variation is not fully accounted for by the factors that have been identified so far. At early learning stages, learners often start out with vastly different scores on tests of non-native perceptual ability; and even those with similar baseline scores often go on to show very different outcomes after non-native perceptual training (e.g., Bradlow et al., 1997; Golestani & Zatorre, 2009; Hanulíková et al., 2012).

Differences in native phonetic perception are one potential predictor of non-native perception. Individuals with better discrimination of native vowels have been shown to have better identification of non-native vowels on a ten-alternative forced-choice task (Lengeris & Hazan, 2010). Similarly, greater sensitivity to native contrasts on a gating task has been related to better discrimination of non-native Mandarin tones (Kalaivanan et al., 2023). Other work suggests that having clearly defined, compact representations of a native vowel in psychoacoustic space predicts greater sensitivity to a non-native vowel contrast (Kogan & Mora, 2022). There is also recent neurophysiological evidence that sensitivity to native contrasts is positively correlated with sensitivity to non-native contrasts (Norrman et al., 2022).

It is not surprising, then, that existing models of non-native phonetic learning emphasize the influence of native phonetic categories. The perceptual assimilation model (Best & Tyler, 2007), speech learning model (Flege, 1995), native language magnet model (Kuhl et al., 2008), and perceptual interference model (Iverson et al., 2003) all describe how a learner's difficulty with a given non-native phoneme will depend on the similarity between that phoneme and native phonemes. For example, one prediction that has received some support is that the difference between two non-native speech sounds is easier to distinguish when the non-native sounds are perceptually assimilated to two different native categories, compared to when they are assimilated to the same native category (Best & Tyler, 2007; Mayr & Escudero, 2010). These models address which phonemes are easier or harder for learners of a given language background overall, without directly addressing individual differences in success between learners. However, some studies have used these models as a framework to predict the success of non-native perception based on differences in assimilation patterns. Mayr and Escudero (2010) studied how native English speakers assimilated German vowels to native categories. They found variety in assimilation patterns, with some participants perceiving the German contrasts in terms of a single native category and others perceiving them in terms of two or more native categories. Importantly, these differences in assimilation were predictive of identification success: participants who assimilated the German contrasts to two distinct native categories showed better identification of those contrasts than participants who assimilated them to a single native category (Mayr & Escudero, 2010). Hattori and Iverson (2009) similarly observed individual differences in assimilation patterns for native Japanese speakers perceiving the English /ɹ/-/l/ contrast. While they did not find a relationship between assimilation patterns and identification success for the English contrast, they did find that identification success was predicted by differences in participants' representations of the third formant for /ɹ/ and /l/. The aforementioned models can thus provide some insight into links between native and non-native perception at the individual level. Moreover, by relating native categories to non-native sound learning, the models imply that differences in non-native perception should be predicted not only by assimilation patterns, but also by differences in the perception of native categories.

As an example, more gradient responses to native sounds on VAS tasks could indicate less of an influence of language-specific categories on perception, and thus yield an easier time learning new categories. Furthermore, gradiency may reflect fine-tuned and flexible perception as detailed above, which could conceivably assist with the discrimination of non-native phonemes. Conversely, steeper identification slopes on 2AFC tasks might predict better non-native perception, since an optimal strategy for a gradient listener on such tasks could be to clearly label each sound based on whichever category it best fits (as discussed above). Fuhrmeister et al. (2023) recently studied non-native discrimination ability and native gradiency as measured by a VAS task and were surprised not to find evidence for a relationship between the two. Instead they found that non-native discrimination related to the consistency of VAS responses, i.e., how similar participants' ratings were across trials for a given stimulus. However, they used a VAS task resembling a Likert scale, with only 7 discrete points (in contrast to the continuous scales used by other researchers such as Kapnoula et al., 2017 and Kong & Edwards, 2016). The presentation of discrete response options may have incited participants to treat the task more similarly to a 2AFC task, putting into question whether the task was truly measuring gradiency. Furthermore, Fuhrmeister et al. (2023) tested only consonants (no native or non-native vowels). Even though gradiency appears to be a relatively stable individual property that holds across different speech sounds as described above (Fuhrmeister & Myers, 2021; Kapnoula & McMurray, 2021; Kong & Edwards, 2016; Kong & Kang, 2022), certain sounds such as consonants are likely to elicit gradient responses to a lesser degree because listeners typically show greater sensitivity to within-category differences in vowels than in consonants (e.g., Fry et al., 1962; Schouten & Van Hessen, 1992). Perhaps a relationship did not emerge between native gradiency and non-native discrimination in their study because there was not a wide enough range of gradiency values due to the use of consonants alone, or not enough variability within the gradiency values due to the limited sensitivity of a 7-point scale. The relationships between native 2AFC and VAS performance and non-native discrimination thus remain to be clarified.

**The Current Study**

The current study had two primary aims. First, we wanted to clarify which individual differences are reflected in performance on 2AFC and VAS tasks. It is of interest to determine whether these two tasks measure the same construct—2AFC tasks are ubiquitous in psycholinguistic research, so it is important to understand what they may be tapping into and how they compare to other tasks. Some authors have concluded that 2AFC and VAS tasks do not measure the same construct, and that 2AFC responses relate to consistency rather than gradiency (e.g., Kapnoula et al., 2017). On the other hand, some authors have used the term gradiency when referring to 2AFC (Ou & Yu, 2022) or 4AFC (Ou et al., 2021) slopes, for example positing that such "gradiency" is in part due to how strongly one's subcortical and cortical representations of speech correlate with one another (Ou & Yu, 2022); this assumes that 2AFC slopes do measure the same construct as VAS slopes. Furthermore, more gradient responses on a VAS task have been related to greater use of secondary cues on a 2AFC task (Kapnoula et al., 2017; Kim et al., 2020). Steeper categorization of primary cues on a 2AFC task has additionally been linked to greater use of secondary cues on the same task (Clayards, 2018), and steeper slopes on a 4AFC task have similarly been linked to greater use of secondary cues in an eye-tracking task (Ou et al., 2021). Together, these findings seem to imply that steeper slopes on 2AFC tasks could be related to shallower slopes on VAS tasks. Such an inverse relationship might also be anticipated given that a gradient listener could show a steep 2AFC slope based on their response strategy, as outlined earlier. Developmental work by McMurray et al. (2018) has found that steeper identification functions and more gradient phonetic perception (as measured by eye-tracking) appear to develop in tandem during adolescence, further hinting at the possibility of an inverse relationship between 2AFC slopes and VAS slopes. However, it is also possible that the slopes are not related across the tasks if 2AFC responses relate more to inconsistency than gradiency, as tentatively proposed by Kapnoula et al. (2017).

Previous work that compared performance across the two tasks did not use identical continua of stimuli, instead presenting participants with VAS continua consisting of 35 stimuli and 2AFC continua consisting of only 14 stimuli (Kapnoula et al., 2017). This difference in the richness of continua across tasks could have contributed to the lack of relationship reported between 2AFC and VAS slopes; in order to more directly compare performance across the two tasks, exactly the same continua should be used for both. Kapnoula et al. (2017) also only found a marginal relationship between 2AFC slopes and VAS consistency (Kapnoula et al., 2017); a conceptual replication is needed in order to clarify whether this finding seems to be a spurious or a genuine one. Furthermore, different analysis techniques have been used across different tasks and across different studies, so it is unclear whether the results depend on the analysis techniques (more on this in the *Comparing Slope Estimate Methods* section). The relationship between tasks and the individual differences reflected by each task therefore requires further investigation, bringing us to our first hypothesis.

*Hypothesis 1*: 2AFC and VAS tasks provide different ways of measuring individual differences in speech sound perception, with VAS slopes reflecting gradiency and 2AFC slopes reflecting consistency. If this is the case, 2AFC slopes will not relate to VAS slopes but will relate to the consistency of VAS responses, with inconsistent VAS performance predicting shallower 2AFC slopes.

The second question we aimed to address was whether discrimination of difficult non-native contrasts could be predicted by differences in native phonetic perception as measured by VAS and 2AFC tasks. We predicted that shallower VAS slopes and steeper 2AFC slopes might both reflect the ability to make accurate and fine-tuned judgments about acoustic cues and might therefore relate to better non-native discrimination. If shallow VAS slopes do predict better non-native perception abilities, this would support the notion that gradiency, as measured by VAS tasks, may actually be adaptive and beneficial. This brings us to our second hypothesis.

*Hypothesis 2*: The ability to discriminate finely tuned differences in native speech sounds relates to the ability to accurately distinguish non-native speech sounds. If this is the case, steeper 2AFC slopes and shallower VAS slopes will relate to better non-native phonetic perception.

These hypotheses were tested in two experiments. In both experiments, we measured how English-speaking participants responded to identical continua of native speech sounds when the sounds were presented in a 2AFC and a VAS task. This enabled a direct comparison of responses across tasks. We also evaluated the participants' ability to discriminate unfamiliar non-native (German) phonemes to investigate potential predictors of good non-native perception. Finally, we collected measures of working memory and attention in order to account for variation in non-linguistic cognitive abilities. Other studies of native and non-native perception have not accounted for such factors (e.g., Fuhrmeister et al., 2023), and yet it is relevant to do so given that executive function has been found to modulate the gradiency of native perception (Kapnoula et al., 2017; Kapnoula & McMurray, 2021) and the success of second language learning outcomes (e.g., Kwakkel et al., 2021; Lee, 2016). These cognitive factors are also important to consider in light of prior evidence that the working memory demands of a task can affect participants' responses and thus bias the conclusions that we draw (Gerrits & Schouten, 2004).

Because we collected a large number of measures and because there were many possible comparisons and analysis techniques available, we treated the first experiment as exploratory. This allowed us to explore the data and to develop an analysis approach after data collection. The methods, exclusion criteria, and analyses established in Experiment 1 were then preregistered on the Open Science Framework (OSF; https://doi.org/10.17605/OSF.IO/9DKGQ) as Experiment 2. Experiment 2 allowed us to then test our hypotheses with a priori analysis decisions and a larger sample size, strengthening our conclusions. We also performed additional non-preregistered analyses on the data from both experiments that we had not considered in the preregistration. The Methods section below describes the preregistered analyses first, which consisted of canonical correlation and multivariate multiple regression to assess Hypothesis 1 and of multiple regression to assess Hypothesis 2. The non-preregistered analyses, outlined at the end of the methods, included additional canonical correlations and a principal component analysis.

## Methods

Aside from the sample size of participants recruited, the methods for Experiment 1 and Experiment 2 were identical. Data for Experiment 1 were collected from September to November 2020 and data for Experiment 2 were collected in July 2021.

### Participants

Participants were right-handed, aged 18-35, born and living in the United States or Canada, and had no history of head injury or of literacy, language, cognitive, or hearing impairments. All were monolingual speakers of English. Participants received monetary compensation ($12.50 USD) and signed an informed consent form. The entire study had a duration of approximately 1.25 hrs including breaks. The research protocol was approved by the Institutional Review Board of the Faculty of Medicine and Health Sciences of McGill University. All participants were recruited through the online platform Prolific.co and were required to have access to a computer to complete the study. While monolingual English speakers with computer access are unlikely to be a universally representative sample, such constraints were necessary to control for prior language experience and to present the experiment in a consistent way across participants. Given that a wider demographic range can be obtained when recruiting from online platforms such as Prolific compared to when recruiting university students, we believe that our results are relatively generalizable.

#### *Experiment 1 Sample Size*

56 participants (21 females) were recruited through the platform Prolific.co.

***Experiment 2 Sample Size***

Experiment 2 was designed to replicate the results of Experiment 1 using a larger sample size. An appropriate sample size was estimated through a triangulation of approaches. As an initial step, we reviewed the sample size in comparable studies, most notably that of Kapnoula et al. (2017), which is closest to the current study and included a sample of 120 participants, leading to some marginal and some significant effects. As a second approach, we relied on Harrell's (2015) rule of thumb applied to our design, which includes 6 predictors; multiplying the 6 predictors by 15 participants per predictor yields a sample size estimate of 90 participants minimum. Finally, we computed a power analysis based on multiple regression with 6 predictors, which reflects our regression models testing Hypothesis 1 and Hypothesis 2 (in fact Hypothesis 1 involved multivariate multiple regression with more than one response variable, but each model within the multivariate model had a single response variable and 6 predictors as in the power analysis, and conducting such an analysis on a multivariate design would be overly complex). This power analysis (with power = 0.95, alpha = 0.05, and number of predictors = 6) using the power.f2.test function from the pwr package (Champely, 2020) in R revealed that a sample size of 120 is required in order to reliably detect effects of a similar size ($r \geq 0.4$) to those reported in related studies (e.g., Clayards, 2018; Grimaldi et al., 2014; Kong & Edwards, 2016). Therefore, we settled on a sample size of 120. In order to arrive at a final sample size of around 120 after taking into account participant exclusion based on language experience and data quality issues, we recruited 139 participants (97 females) through the platform Prolific.co.

### Questionnaires

Information about demographics, language history and proficiency, and musical experience was collected through a questionnaire adapted from the *Language History Questionnaire* (LHQ 2.0; Li et al., 2013) and the *Montreal Music History Questionnaire* (MMHQ; Coffey et al., 2011).

### Tasks

Participants completed five tasks: two measuring native phonetic perception, one measuring non-native perception, one measuring sustained attention, and one measuring working memory (all described further below). Participants completed these tasks online at home using the Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020), with their own headphones. In order to standardize online sound presentation and ensure an acceptable listening environment, participants completed a headphone screening before the other tasks (Woods et al., 2017).

#### *Native Phonetic Perception Tasks*

Participants completed two native phonetic perception tasks. The tasks involved listening to minimal pairs that varied in different phonological contrasts (*bet-bat* and *dear-tear*; stimuli from Clayards 2018, publicly available at https://osf.io/369my/). These two pairs were selected because they enabled us to test perception of both a vowel and a consonant contrast, and they have successfully been used in the past to study individual differences in phonetic perception (Clayards, 2018). The minimal pairs were manipulated so that each one varied systematically in two acoustic cues relevant to the contrast (formant frequency and vowel duration for *bet-bat*, voice onset time and onset fundamental frequency for *dear-tear*). Each cue varied in 5 steps, and each version of the first cue was paired with each version of the second cue, leading to 25 stimuli per pair. This results in some ambiguous and some clear stimuli (stimuli whose cue values are both at the extremes—i.e., step 1 or step 5—sound clear and unambiguous; stimuli with more intermediate cue values sound more ambiguous). Details of stimulus properties are listed in Table 1, and further details of stimulus construction can be found in Clayards (2018). The same stimuli were used in both native phonetic perception tasks.

In the 2AFC task, participants indicated via mouse click which of two words they heard on each trial (e.g., *bet* or *bat*; side of the screen counterbalanced across participants). In the VAS task, participants were shown a slider on the computer screen with a word at each end (ends of the slider counterbalanced across participants). Participants indicated where along the continuous scale they perceived the stimulus to be (values were coded from 0 at one end to 100 at the other end, but were not displayed to participants during the task). Each stimulus from each minimal pair was presented 5 times in each task, for a total of 250 stimuli per task. Stimuli were blocked so that all 25 stimuli per pair appeared in a random order before any stimulus was repeated. *Bet-bat* and *dear-tear* trials were mixed in each block. All participants completed the VAS task first to avoid biasing responses based on the more categorical demands of the 2AFC task.

***Non-Native Phonetic Perception Task***

In the non-native perception task, participants differentiated German vowels and consonants (øː vs. œ, yː vs. ʏ, ʃ vs. ç in the International Phonetic Alphabet) which are known to be perceptually challenging speech sounds for native English speakers (Mayr & Escudero, 2010). German words containing these phonemes were presented in a 3-interval oddity (3-I oddity) task, in which participants heard three stimuli in a row and indicated which one (if any) was different. 3-I oddity tasks are useful for studying non-native phonetic perception since they do not require the participant to explicitly know the nature of the differences between unfamiliar stimuli (Strange & Shafer, 2008). They also have an advantage over similar tasks such as AXB, in that they are more intuitive for participants and their level of chance performance is lower (25% instead of 50%), allowing for greater variability in scores (Grimaldi et al., 2014). The complete set of German minimal pairs used in the task is found in Table 2.

In order to construct the stimuli for the task, three native German speakers were recorded producing each German word 5 times. The 1st and 5th productions were then discarded to leave 3 productions of each word per speaker. Sound files were edited to leave 20 ms before and after each production, and maximum amplitudes were normalized across speakers using GoldWave version 6.15 (GoldWave Inc., 2015). Each trial contained three words, one from each speaker, with an interstimulus interval of 500 ms. Participants indicated which word sounded different by clicking “1”, “2”, or “3” on a computer screen, or clicking “None” if all three words sounded the same. Half of the trials were switch trials where one of the words was the other member of the minimal pair, and the other half were catch trials where all three words were the same. For example, for the minimal pair "selig" and "seelisch" (/ˈzeːlɪç/ and /ˈzeːlɪʃ/), participants might hear "selig, seelisch, selig" on a switch trial and "selig, selig, selig" on a catch trial. There were 12 trials (6 switch and 6 catch trials) per minimal pair and 14 minimal pairs, for a total of 168 trials. Speaker order, odd speaker out, and odd minimal pair member were balanced across trials, and trial order was randomized. Before implementing the task, piloting with 6 participants was conducted in order to check for floor or ceiling effects. Piloting revealed overall accuracy rates of 39-65% (keeping in mind that chance performance is 25%), falling within the range of previous studies (e.g., Rauber et al., 2005; Silveira, 2011).

**Table 1**

*Stimulus properties for the native perception tasks*

| **bet-bat** | | | **dear-tear** | |
|---|---|---|---|---|
| Formant frequencies of spectral steps (Hz) | | Duration steps (ms) | Voice onset time steps (ms) | Onset F0 steps (Hz) |
| F1 | F2 | | | |
| 625 | 1677 | 100 | 10 | 185 |
| 647 | 1610 | 140 | 20 | 195 |
| 663 | 1560 | 180 | 30 | 205 |
| 682 | 1546 | 220 | 40 | 215 |
| 740 | 1556 | 260 | 50 | 225 |

**Table 2**

*German minimal pairs used in the 3-I oddity task*

| Consonant contrast | | Vowel contrast 1 | | Vowel contrast 2 | |
|---|---|---|---|---|---|
| Palatal fricative (ç) | Postalveolar fricative (ʃ) | Tense high front rounded vowel (yː) | Lax high front rounded vowel (ʏ) | Tense mid front rounded vowel (øː) | Lax mid front rounded vowel (œ) |
| Fichte /fɪçtə/ | fischte /fɪʃtə/ | Brühl /bʁy:l/ | brüll /bʁʏl/ | blöke /blø:kə/ | Blöcke /blœkə/ |
| Kirche /kɪəçə/ | Kirsche /kɪəʃə/ | Düne /dy:nə/ | dünne /dʏnə/ | gewöhne /gəvø:nə/ | gewönne /gəvœnə/ |
| Löchern /løçɪən/ | löschern /løʃɪən/ | fühlen /fy:lən/ | füllen /fʏlən/ | Höhle /hø:lə/ | Hölle /hœlə/ |
| selig /zelɪç/ | seelisch /zelɪʃ/ | Hüte /hy:tə/ | Hütte /hʏtə/ | Söhne /zø:nə/ | Sönne /zœnə/ |
| Wicht /vɪçt/ | wischt /vɪʃt/ | Wüste /vy:stə/ | wüsste /vʏstə/ | | |

***Cognitive Tasks***

Finally, participants completed a version of the Continuous Performance Test (CPT; Conners et al., 2003) and a working memory task in order to assess whether any observed relationships between performance on the other tasks might be driven by individual differences in non-linguistic cognitive factors rather than in perception.

In the AX-CPT, participants were presented with a string of letters. They had to press a particular key whenever they saw the letter X preceded by the letter A (this was the case for 70% of trials) and press a different key in any other case (with keys counterbalanced across participants). There were 140 AX trials (A followed by X), 20 AY trials (A followed by a consonant other than X), 20 BX trials (B followed by X), and 20 BY trials (B followed by a consonant other than X), for a total of 200 trials.

A backwards digit span task was used to assess working memory (Wechsler, 2008). In this task, participants heard recorded series of numbers (presented with a 1s interstimulus interval) and were then asked to type them out in the reverse order. The number of digits to be recalled increased every 3 trials, starting with 2 digits and increasing to a maximum of 10 digits. The task was terminated whenever the participant incorrectly answered all 3 trials of a given difficulty level.

***Reliability of Measures***

Given our focus on individual differences, one important consideration is whether the measures being used here are reliable within participants. To address this, we calculated the split-half reliability of each of our measures, adjusted with the Spearman-Brown correction. These reliability values are displayed in Supplemental Table S.1, revealing good reliability of all measures apart from the VAS slopes (this is simply due to bad fitting when not enough data is provided; see Supplemental Materials for further details).

Test-retest reliability is another informative measure of reliability. Common measures of test-retest reliability include Cronbach's alpha, test-retest correlation coefficients, and intraclass correlations which exist in ten different forms depending on the data structure and the type of reliability being calculated (Koo & Li, 2016). It is important to interpret reliability values according to the particular research context, and so we refer to benchmarks from the field of psychology: Cronbach's alpha and test-retest correlation values of 0.7-0.79, 0.80-0.89, and > 0.90 indicate fair, good, and excellent reliability respectively; and intraclass correlation values of 0.4-0.59, 0.6-0.74, and > 0.75 indicate fair, good, and excellent reliability respectively (Cicchetti, 1994). Previous studies have found fair to excellent test-retest reliability for various perceptual measures related to the ones used here, including auditory discrimination (Christopherson & Humes, 1992: Cronbach's alpha = 0.795; Saito & Tierney, 2022: intraclass correlation = 0.625; Wang & Humes, 2008: test-retest correlations > 0.90), sensitivity to the McGurk effect (Strand et al., 2014: test-retest correlation = 0.77), magnitude of the Ganong effect (Giovannone & Theodore, 2023: test-retest correlation = 0.72), use of a VAS scale (Brietzke et al., 2021: intraclass correlation = 0.50), consonant identification (Geller et al., 2021: intraclass correlation = 0.80), and weighting of acoustic cues (Idemaru et al., 2012: test-retest correlation = 0.69; Souza et al., 2018: no difference in cue weightings across two sessions, as determined by Wilcoxon signed-ranks analyses). Importantly, individuals' 2AFC identification slopes for stimuli varying in voice onset time (VOT) and fundamental frequency (F0)—two of the same acoustic cues varying in our stimuli—have shown good to excellent reliability across sessions (Schertz et al., 2015: test-retest correlations = 0.90 for VOT and 0.84 for F0), and individuals' gradiency of speech perception on a VAS task has shown fair to good reliability across ratings of the same stimulus (Munson et al., 2021: intraclass correlation > 0.5 for 89% of listeners, average = 0.73). Furthermore, fair to good test-retest reliability has been shown for the backwards digit span task (Fox-Fuller et al., 2022: intraclass correlation = 0.66; Müller et al., 2012: intraclass correlation = 0.64; Wechsler, 2008: r = 0.71; Woods et al., 2011: r = 0.81), while fair to excellent test-retest reliability has been observed for the AX-CPT task (Barch et al., 2009: intraclass correlation

= 0.81, Cooper et al., 2017: intraclass correlation = 0.70; Halperin et al., 1991: test-retest correlations = 0.65-0.74; Kraus et al., 2020: intraclass correlation = 0.72).

Based on both split-half and test-retest reliability, we can therefore conclude that our measures are reliable and appropriate for use in the context of individual differences studies such as the present one.

**Transparency and Openness**

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All stimuli, tasks, questionnaires, program code, and analysis methods developed by others have been cited in-text and included in the References section. The research materials (tasks and questionnaires) described above are available upon request. Stimuli from the native phonetic perception tasks are publicly available, on the OSF page (https://osf.io/369my/) for Clayards (2018). The raw data for both experiments, along with the code needed to process and analyze it, is publicly available on the OSF (https://osf.io/ez5qh/?view_only=8e4a1498e04f4ee0946752ee93b9ce71). The design, hypotheses, and analysis plan of Experiment 2 were preregistered based on Experiment 1 and are available on the same OSF page.

## Analysis and Results

Here we include tables and figures displaying results of primary interest. Additional tables and figures (for example, of model validation) can be found in the Supplemental Materials and in the R Markdown document on the OSF page for this project.

## Data Exclusion

Participants who reported having phonetic training or being exposed to German were excluded (two participants in Experiment 1, 19 participants in Experiment 2), as this could affect performance on the non-native perception task. Participants were also excluded on a task-by-task basis depending on performance-based criteria. Criteria are outlined in the OSF preregistration (https://doi.org/10.17605/OSF.IO/9DKGQ). The total number of participants included in a given analysis is reported at the bottom of the figure (in the case of canonical correlation) or table (in the case of regression) displaying the output of that analysis.

## Preparatory Data Analysis

Before conducting primary analyses, various preliminary analyses were carried out to obtain variables of interest.

### *Native Phonetic Perception Tasks*

Slopes from the 2AFC task were calculated by fitting two mixed-effects logistic regression models to participant responses (one for bet/bat, one for dear/tear). Responses were coded as 0 for bet/dear and 1 for bat/tear. The fixed effects for each model were the first acoustic cue (which varied in 5 steps) and the second acoustic cue (which also varied in 5 steps) for the contrast in question, both of which were coded as continuous numeric variables and centered. The grouping factor was participant. The following correlated random effects were included in each model: by-participant random intercepts, and by-participant random slopes for the first acoustic cue and the second acoustic cue. The by-participant random slopes coefficients for each acoustic cue were extracted as the four variables of interest, since they quantify how much each participant differs from the group average (i.e., from the fixed effect coefficient) in their use of a given cue when categorizing stimuli (Clayards, 2018; Kong & Edwards, 2015). Larger random slopes coefficients (steeper slopes) for a given cue indicate greater use of that cue when categorizing stimuli. This analysis was carried out in R (R Development Core Team, 2020), using the lme4 package (Bates et al., 2015). The R syntax for the models described above was: glmer(2AFC response ~ Acoustic cue 1 step + Acoustic cue 2 step + (Acoustic cue 1 step + Acoustic cue 2 step | Participant), family = "binomial", control = glmerControl(optimizer = "bobyqa")).

Slopes from the VAS task were calculated by fitting the rotated logistic developed by Kapnoula et al. (2017) to participants' responses. The rotated logistic is conceptually similar to the 2AFC logistic regression coefficients mentioned above, but it models gradiency independently of acoustic cue use (since our stimuli vary in two acoustic cues). It is based on a four-parameter logistic function with estimates for minimum and maximum asymptotes, slope, and crossover point, but with one additional parameter: $\theta$, which represents the angle of the crossover point. The coordinate space is rotated to be orthogonal to this angle, with the result that the slope parameter provides a single measure of gradiency which is independent of the two acoustic cues constituting the space. These analyses were conducted in MatLab (version 2015a, The MathWorks Inc., USA). For each participant and each minimal pair, the average of the 5 responses to each of the 25 different stimuli in the VAS task was calculated, and the equation for the rotated logistic was fit to these averages. This resulted in two slope measures per participant: one for bet-bat responses, and one for dear-tear responses. Larger slope values from the rotated logistic function reflect shallower slopes and therefore more gradient responses.

To calculate differences in the consistency of participants' acoustic cue encoding, the rotated logistic was fitted to each participant's unaveraged responses. For each trial, the difference between the participant's actual VAS response and the response predicted by the rotated logistic was calculated. The standard deviation of these residuals was then averaged per minimal pair to provide two estimates of consistency per participant: one for bet-bat responses, and one for dear-tear responses. Greater standard deviation of residuals reflects less consistent responses. This is the same method used by Kapnoula et al. (2017) to calculate consistency, and closely resembles the method used by Fuhrmeister et al. (2023) who also calculated residuals from a logistic function fit to participants' VAS responses (but theirs was a regular rather than a rotated logistic function, since their continua varied only along one acoustic dimension).

***Non-Native Perception Task***

To quantify differences in non-native phonetic perception, the non-parametric sensitivity index *A* (a corrected version of *A'*; Zhang & Mueller, 2005) was calculated across performance on the fricative contrast and the vowel contrasts from the 3-I oddity task. This score is based on hits (correctly selecting the odd item in a switch trial) and false alarms (incorrectly selecting an odd item in a catch trial). An *A* score of 1.0 indicates perfect discrimination, while a score of 0.5 indicates null discrimination. The calculation was done by implementing Zhang & Mueller's (2005) equation in R.

***Cognitive Tasks***

As a measure of sustained attention, a bin score was calculated from each participant's AX-CPT responses (Hughes et al., 2014). Unlike traditional reaction time (RT) difference measures, bin scores take into account both RT and accuracy, making them more reliable and suitable for use in individual differences studies (Draheim et al., 2019).

In preparation for bin scoring, trials were labeled by type (AX/AY/BX/BY) and also labeled as nonswitch (AX) or switch (AY/BX/BY). Only rows corresponding to the second letter of each trial were kept (i.e., X/Y, not A/B), and reaction times (RTs) were cleaned: for each participant, RTs < 200 ms were replaced with that participant's mean RT value, and RTs > 3 SD above their mean RT were replaced with a cutoff value of 3 SD above the mean. To calculate a participant's bin score, their mean RT on non-switch (AX) trials was subtracted from their RT for each switch trial (AY/BX/BY trials). The resulting RT differences were placed into ten bins which were assigned values ranging from 1 (smallest RT differences) to 10 (largest RT differences). Inaccurate responses were placed in a "bad" bin with a value of 20 to provide a penalty for low accuracy. Finally, the bin values for all of the participant's trials were summed to produce a final bin score. Lower bin scores indicate better attention due to smaller RT differences and/or higher accuracy.

From the backwards digit span test, the highest number of digits successfully recalled was taken as a measure of working memory.

**Descriptive Overview**

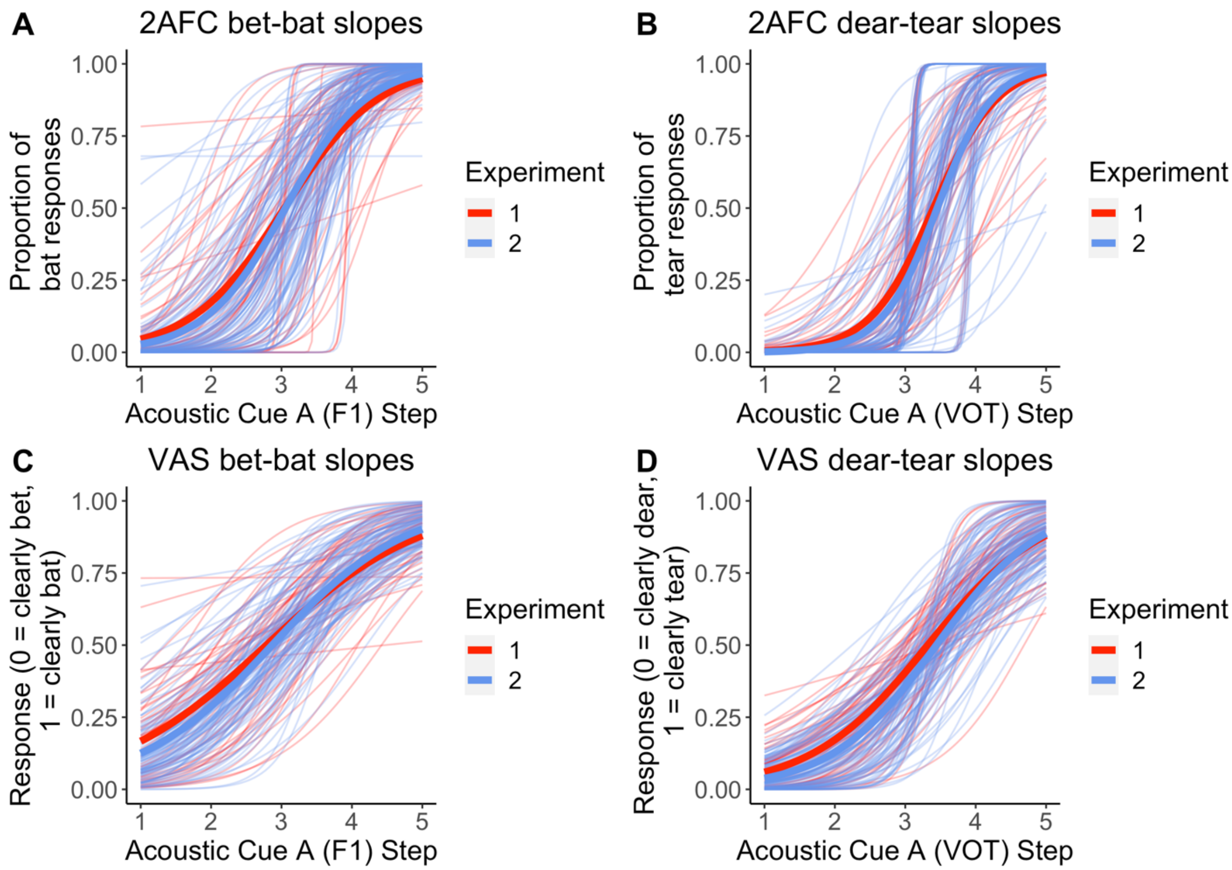
Performance on the two native language perception tasks is shown in Figure 1, and representative individual results are shown in Figure 2. When averaged across all participants, overall response patterns were similar across Experiment 1 and Experiment 2 (compare thick red and blue lines in Figure 1). However, significant individual variability was observed across both tasks and both experiments, with participants showing identification slopes ranging from very shallow to very steep (see thin lines in Figure 1 and example participants in Figure 2). Steeper slopes are more evident on the 2AFC task (no doubt due to its categorical nature) than on the VAS task. Participants also differed in the consistency of their responses, that is, in how closely their response to each stimulus fell around their predicted identification slope (Figure 2).

Violin plots of scores on the non-native perception task and the cognitive tasks are displayed in Figure 3. Overall performance (mean and standard deviation on each task) was very similar for Experiment 1 and Experiment 2, and similar variability in scores was also observed across both experiments as shown by the overlap between red and blue plots. As anticipated, the non-native discrimination task was generally challenging, with mean accuracy falling at 53-54% for both experiments; and at the individual level some participants had particular difficulty discriminating the non-native sounds (accuracy around 25%, at chance), while others were quite successful (accuracy of 75% and above; Figure 3A). Participants also showed a range of scores on the attention and memory tasks (Figure 3B and 3C).

Prior to the main analyses, for the sake of ease of interpretation and exploration of the data, pairwise correlations were computed and visualized between the variables of interest for each hypothesis. These pairwise correlations are provided in Supplemental Figures S.1 to S.4.

**Figure 1**

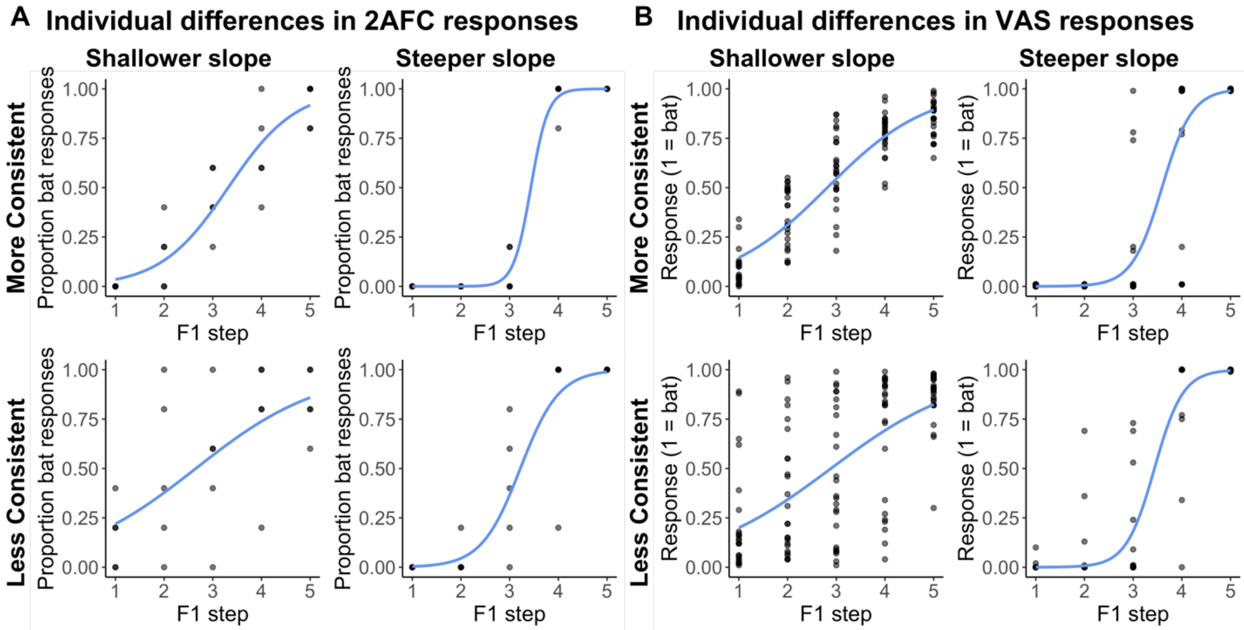
*Group and individual responses on the native perception tasks (2AFC and VAS), for both experiments*




*Note*. (A) 2AFC bet-bat responses by cue A, (B) 2AFC dear-tear responses by cue A, (C) VAS bet-bat responses by cue A, and (D) VAS dear-tear responses by cue A. Thin lines are logistic curves fit to each individual participant for each step of acoustic cue A, and thick lines are logistic curves fit to the whole dataset. VAS responses varied continuously from 0-100, but were transformed to range from 0-1 for the purposes of fitting logistic curves to the data for these plots (regular logistic regression was used here for visualization purposes, rather than the rotated logistic function fit to the VAS data as described in the *Preparatory Data Analysis* section).

**Figure 2**

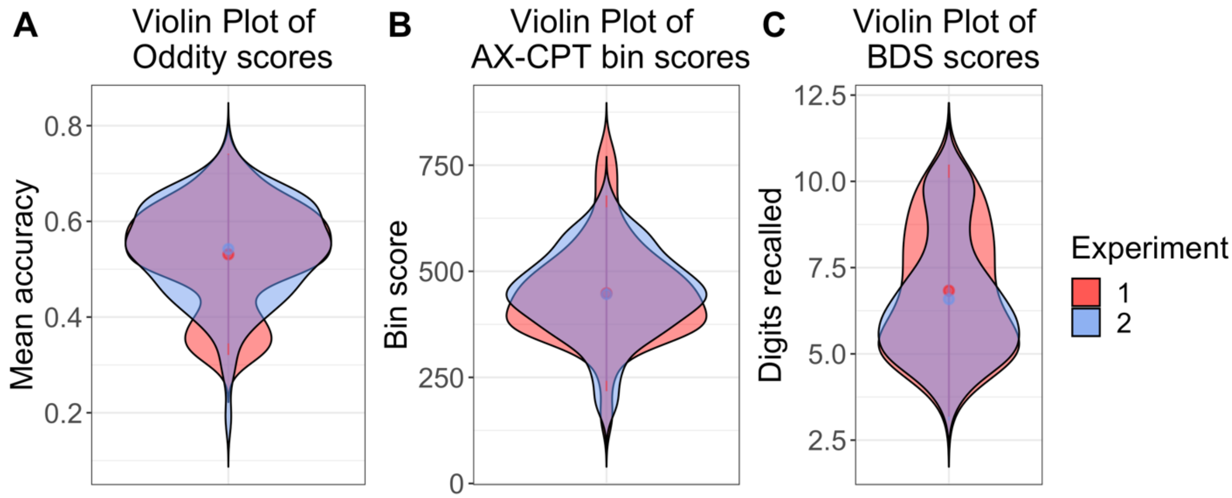
*Examples of individual variability on the native perception tasks (2AFC and VAS)*




*Note.* For the 2AFC task (A), each dot is the participant's average response across the five presentations of a given stimulus. For the VAS task (B), each dot is a participant's response on a given trial. Lines are logistic curves fit to responses; dots clustered closely around the fitted curve indicate more consistent responses. VAS responses varied continuously from 0-100, but were transformed to range from 0-1 for the purposes of fitting logistic curves to the data for these plots (regular logistic regression was used here for visualization purposes, rather than the rotated logistic function fit to the VAS data as described in the *Preparatory Data Analysis* section). Top left of each plot: shallow and consistent, top right: steep and consistent, bottom left: shallow and inconsistent, bottom right: steep and inconsistent. The participants in each panel are chosen as representative examples of variability on the task and are not the same across both tasks.

**Figure 3**

*Distributions of performance on non-native and control tasks*




*Note.* (A) Oddity task, (B) AX-CPT task, and (C) Backwards Digit Span task, for both experiments. Mean and standard deviation are indicated by the dot and vertical line within each plot. Purple indicates overlap between the two experiments.

## Hypothesis 1 – Canonical Correlation

### *Analysis*

Our first hypothesis was that 2AFC slopes would not relate to VAS slopes but would relate to the consistency of VAS responses, with inconsistent VAS performance predicting shallower 2AFC slopes. Since we had 4 2AFC slope measures, 2 VAS slope measures and 2 VAS consistency measures, we tested our hypothesis by first running canonical correlation analyses, which test the strength of relationships between two sets of variables. Canonical correlation is a dimensionality reduction technique similar to principal component analysis (PCA), but while PCA aims to determine the dimensions that account for the most variance *within* a set of variables, canonical correlation analyses aim to determine the dimensions that account for the most covariance *between* two sets of variables. Canonical correlation analyses output canonical correlation coefficients, which measure the strength of the association between pairs of canonical variates (each pair of canonical variates is called a canonical dimension, so the canonical correlation coefficients can also be thought of as representing the strength of each canonical dimension). A canonical variate is an orthogonal, linear combination of the variables within a set—the variables are weighted so as to maximize the correlation between the canonical variate derived from that set of variables and the canonical variate derived from the other set of variables of interest (i.e., to maximize the correlation coefficient for a given canonical dimension). Canonical variates are latent variables and can be considered analogous to the factors derived from factor analysis. The number of canonical variate pairs or canonical dimensions is equal to the number of variables in the smallest set; in this case, there are two canonical dimensions for each canonical correlation. A significant correlation along one or both dimensions suggests a relationship between the two sets of variables. Statistical significance of the canonical correlation coefficients for each dimension was evaluated using Wilks' lambda. More information on canonical correlation analysis can be found in Sherry & Henson (2005) and UCLA: Statistical Consulting Group (n.d.).

Canonical correlation 1 was between the four 2AFC random slopes coefficients and the two VAS slope measures. Hypothesis 1 predicted that these sets of variables would not be related. Canonical correlation 2 was between the four 2AFC random slopes coefficients and the two VAS consistency measures. Hypothesis 1 predicted that these sets of variables would be related. These analyses were conducted in R using the packages CCA (Canonical Correlation Analysis; González & Déjean, 2021) and CCP (Significance Tests for Canonical

Correlation Analysis; Menzel, 2012). When interpreting effect size of the results, we follow the guidelines established by Gignac & Szodorai (2016) for individual differences research (small: $r = 0.1$; medium: $r = 0.2$; large: $r = 0.3$) and those established by Plonsky & Oswald (2014) for second language research (small: $r = 0.25$; medium: $r = 0.4$; large: $r = 0.6$). As such, a correlation < 0.1 is considered small and > 0.6 is considered large, while intermediate values are referred to by a combination of the two guidelines (e.g., 0.4 is considered medium by Plonsky & Oswald and large by Gignac & Szodorai, so we consider such a value to reflect a medium-large effect size).

***Results – Experiment 1***

Canonical correlation 1 revealed that the relationship between 2AFC coefficients and VAS slope measures was large and significant along the first canonical dimension ($r_c = 0.690$, $p < 0.001$), and medium-large but did not reach significance along the second canonical dimension ($r_c = 0.439$, $p = 0.037$). The significant correlation along the first dimension suggests that, contrary to our hypothesis, there does appear to be a relationship between the 2AFC coefficients and VAS slopes. A scatterplot of this significant correlation is displayed in Figure 4A. Note that due to the complex patterns of loadings of the original variables onto the first canonical dimension, this positive canonical correlation coefficient does not indicate that slopes are positively related across the 2AFC and VAS tasks; rather, the relationship between slopes across tasks appears to depend on the contrast and acoustic cue. Furthermore, the result should be treated with caution as the effect is smaller and no longer significant with increased statistical power, as described in the results of Experiment 2 below.

Canonical correlation 2 revealed that the relationship between 2AFC coefficients and VAS consistency measures was large and significant along the first canonical dimension ($r_c$ = 0.617, $p$ < 0.001) and medium-large but did not reach significance along the second canonical dimension ($r_c$ = 0.410, $p$ = 0.064). Thus, in line with our hypothesis, there does appear to be a relationship between the 2AFC coefficients and VAS consistency measures. A scatterplot of the significant correlation along the first canonical dimension is displayed in Figure 4B.

For both correlations, Supplemental Table S.2 displays the canonical correlation coefficients and their significance, and Supplemental Table S.3 displays canonical coefficients showing loadings of the original variables onto each canonical dimension. The interpretation of canonical coefficients is analogous to the interpretation of regression coefficients.

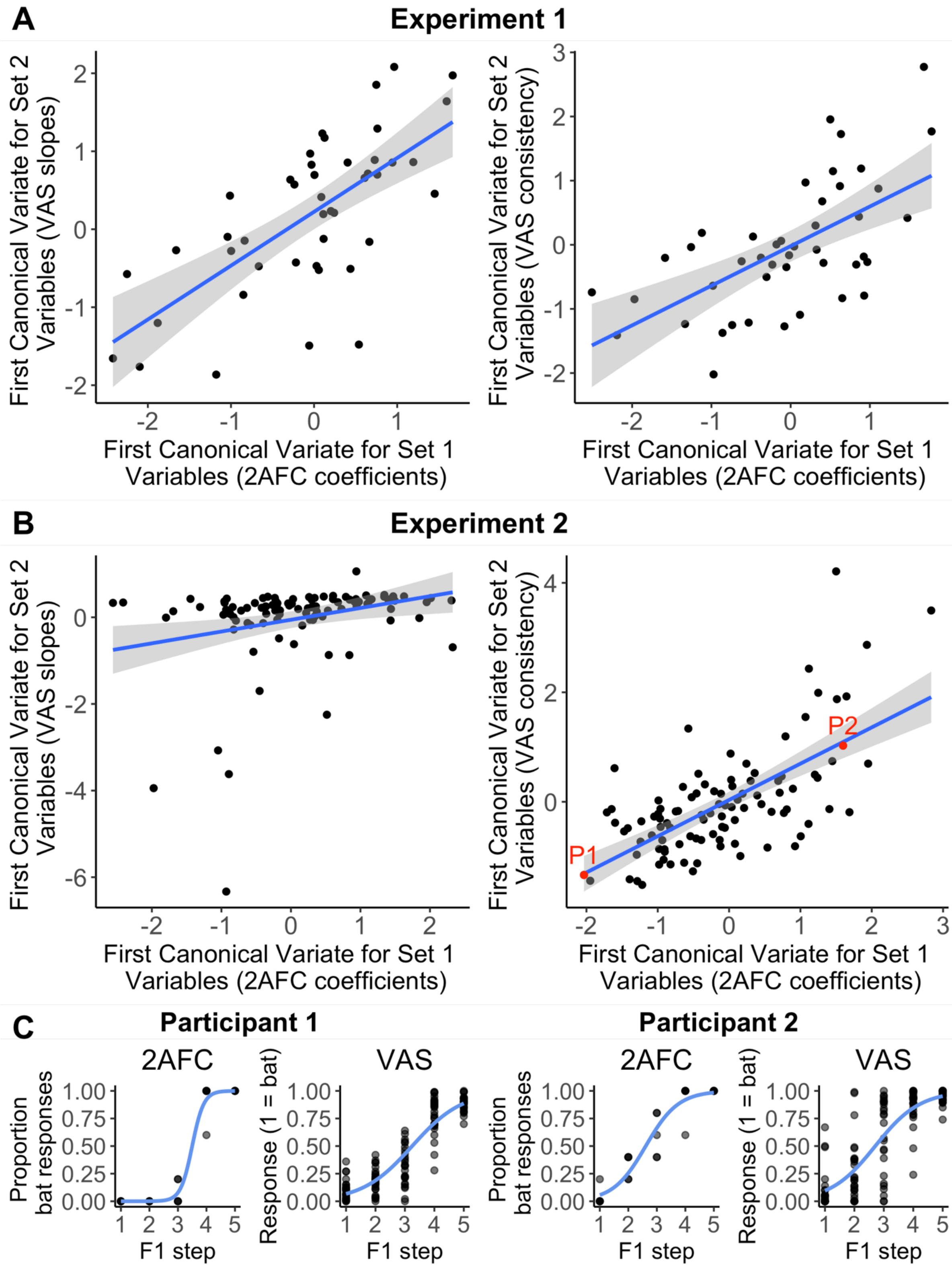
***Results – Experiment 2***

Canonical correlation 1 revealed that the relationship between 2AFC coefficients and VAS slope measures was small-medium and statistically insignificant along the first canonical dimension ($r_c$ = 0.272, $p$ = 0.423) and small and statistically insignificant along the second canonical dimension ($r_c$ = 0.090, $p$ = 0.855). Notice how the effect size is smaller than in Experiment 1. See Supplemental Table S.2 for canonical correlation coefficients and their significance, and Supplemental Table S.3 for canonical coefficients showing loadings of the original variables onto each canonical dimension. The lack of significant correlation is in line with our hypothesis that 2AFC coefficients and VAS slopes are not related; perhaps the significant correlation found in Experiment 1 was due to an insufficient sample size. A scatterplot of the non-significant correlation along the first canonical dimension is displayed in Figure 4C.

Canonical correlation 2 revealed that the relationship between 2AFC coefficients and VAS consistency measures was large and significant along the first canonical dimension ($r_c$ = 0.663, $p < 0.001$), and medium-large and significant along the second canonical dimension ($r_c$ = 0.398, $p < 0.001$). This effect size is similar to what was found for the same analysis in Experiment 1, and stands in comparison to the small effect size observed for correlation 1 between 2AFC coefficients and VAS slopes. See Supplemental Table S.2 for canonical correlation coefficients and their significance. Thus, across both Experiment 1 and Experiment 2, we find support for our hypothesis that 2AFC coefficients and VAS consistency measures are related. A scatterplot of the significant correlation along the first canonical dimension is displayed in Figure 4D, and two participants from opposite ends of the correlation are highlighted as examples. The response patterns of these two participants on the 2AFC and VAS tasks are shown in Figure 4E. Participant 1 illustrates how people with *steeper* 2AFC slopes tend to have *more consistent* VAS responses. Participant 2 illustrates how people with *shallower* 2AFC slopes tend to have *less consistent* VAS responses. Participant 1 has a steep 2AFC slope and shallow VAS slope whereas Participant 2 has similar slopes across both tasks, demonstrating how slopes on the two tasks do not necessarily relate within participants.

**Figure 4**

*Relationships between variables of interest for hypothesis 1*



*Note*. (A) Experiment 1: Scatterplots of the correlation between the first pair of canonical variates, for the canonical correlation between 2AFC coefficients and VAS slopes (left) and the canonical correlation between 2AFC coefficients and VAS consistency (right). $n = 44$. (B) Experiment 2: Scatterplots of the correlation between the first pair of canonical variates, for the canonical correlation between 2AFC coefficients and VAS slopes (left) and the canonical correlation between 2AFC coefficients and VAS consistency (right). Each dot represents a participant, and blue lines are lines of best fit with 95% CIs. Two representative participants from different ends of the correlation are highlighted in red. $n = 100$. (C) Experiment 2: Response patterns on the 2AFC and VAS tasks, for the two representative participants highlighted in (B). For the 2AFC task, each dot is the participant's average response across the five presentations of a given stimulus. For the VAS task, each dot is a response on a given trial. Lines are logistic curves fit to responses; dots clustered closely around the fitted curve indicate more consistent responses. Participant 1 has a steep 2AFC slope, shallow VAS slope, and consistent VAS responses. Participant 2 has shallow 2AFC and VAS slopes, and inconsistent VAS responses. These two participants illustrate how slopes across tasks are not necessarily related within participants, and how steeper 2AFC slopes are associated with more consistent VAS responses. Note that VAS responses varied continuously from 0-100, but were transformed to range from 0-1 for the purposes of fitting logistic curves to the data for these plots (regular logistic regression was used here for visualization purposes, rather than the rotated logistic function fit to the VAS data as described in the *Preparatory Data Analysis* section).

## Hypothesis 1 – Multivariate Multiple Regression

### *Analysis*

Following up on the correlations, we conducted a multivariate multiple regression analysis. This enabled us to include all four 2AFC coefficients as the response and all four VAS measures of interest as predictors, as well as attention and memory measures as additional control predictors. In doing so, we were able to determine whether any relationships found through canonical correlation would still hold after controlling for these additional predictors.

Using the lm() function, the model equation in R was: cbind(2AFC bet-bat acoustic cue A slope, 2AFC bet-bat acoustic cue B slope, 2AFC dear-tear acoustic cue A slope, 2AFC dear-tear acoustic cue B slope) ~ VAS bet-bat slope + VAS dear-tear slope + VAS bet-bat consistency + VAS dear-tear consistency + AX-CPT bin score + Digit span level. We then used multivariate tests (Type II MANOVA) to evaluate the significance of each predictor across the four models, while accounting for the covariances between coefficients. This was done with the Anova() function from the car package in R (Fox & Weisberg, 2019).

***Results***

Output from the multivariate multiple regression model includes regression tables from four separate regression models, fit with each 2AFC coefficient as the response; this output is shown in Supplemental Tables S.4 and S.5 for Experiment 1 and Experiment 2 respectively. Model validation plots (quantile-quantile plots of residuals, plots of fitted values against residuals, and plots of Cook's distance per participant) can be found in Supplemental Figures S.5 to S.7.

Multivariate tests (Type II MANOVA) were used to evaluate the significance of each predictor across the four models while taking into account the covariances between coefficients (Table 3). These analyses revealed that, in line with our hypothesis and with the canonical correlation results, the VAS consistency measures significantly predicted 2AFC coefficients after accounting for other predictors. The AX-CPT and backwards digit span predictors were not significant. These findings held across both experiments. For Experiment 1, the VAS slope measures significantly predicted the 2AFC coefficients (contrary to our hypothesis); however, with the increased power obtained in Experiment 2, this relationship disappeared.

**Table 3**

*Summary of the multivariate multiple regression model predicting 2AFC coefficients, for each experiment*

| **Experiment 1** | | | | | |
|---|---|---|---|---|---|
| Predictor | Pillai's trace | *F* | Num *df* | Den *df* | *p* |
| VAS bet-bat slope | 0.591 | 10.854 | 4 | 30 | <0.001 |
| VAS dear-tear slope | 0.305 | 3.291 | 4 | 30 | 0.024 |
| VAS bet-bat consistency | 0.318 | 3.496 | 4 | 30 | 0.019 |
| VAS dear-tear consistency | 0.411 | 5.239 | 4 | 30 | 0.003 |
| AX-CPT bin score | 0.234 | 2.293 | 4 | 30 | 0.082 |
| Backwards digit span | 0.076 | 0.614 | 4 | 30 | 0.656 |
| *n = 40* | | | | | |
| **Experiment 2** | | | | | |
| Predictor | Pillai's trace | *F* | Num *df* | Den *df* | *p* |
| VAS bet-bat slope | 0.049 | 1.058 | 4 | 83 | 0.383 |
| VAS dear-tear slope | 0.370 | 0.796 | 4 | 83 | 0.531 |
| VAS bet-bat consistency | 0.201 | 5.207 | 4 | 83 | <0.001 |
| VAS dear-tear consistency | 0.206 | 5.391 | 4 | 83 | <0.001 |
| AX-CPT bin score | 0.039 | 0.852 | 4 | 83 | 0.497 |
| Backwards digit span | 0.078 | 1.761 | 4 | 83 | 0.145 |
| *n = 93* | | | | | |

*Note.* Model equation: cbind(2AFC bet-bat acoustic cue A slope, 2AFC bet-bat acoustic cue B slope, 2AFC dear-tear acoustic cue A slope, 2AFC dear-tear acoustic cue B slope) ~ VAS bet-bat slope + VAS dear-tear slope + VAS bet-bat consistency + VAS dear-tear consistency + AX-CPT bin score + Backwards digit span.

## Hypothesis 2 – Multiple Regression

### *Analysis*

Hypothesis 2 involved predicting non-native perception from all native perception and control measures, which would have resulted in a model with ten predictors. To reduce the number of predictors and thus reduce overfitting while increasing power, dimensionality of the native perception measures was reduced using PCA, as implemented by the prcomp() function in R. The same procedure was followed for both experiments: one PCA was run on the four 2AFC coefficients and another was run on the four VAS variables (two slope and two consistency measures). The first two components from each PCA were then extracted for analysis. Correlations between the original variables and the extracted principal components for both experiments are displayed in Table 4. Across the two experiments, all four 2AFC variables were correlated in the same direction with the first component suggesting that this component reflected 2AFC slopes in general, and bet-bat acoustic cue B was strongly positively correlated with the second component. For the VAS measures across the two experiments, slopes and consistency were correlated in opposite directions with the first component while bet-bat and dear-tear measures were correlated in opposite directions with the second component, suggesting that the first component distinguishes between slope and consistency while the second one distinguishes between the two contrasts.

In order to test hypothesis 2, a multiple regression model was fit. The response was oddity *A* scores, and the predictors were the first two principal components derived from the PCA of the 2AFC coefficients, the first two principal components derived from the PCA of the VAS measures, and the two control predictors. Because visualization of the distribution of oddity *A* scores for both experiments revealed some negative skew, the scores were exponentially transformed; models were then fit predicting the scores both with and without the transformation, and the model with the best performance is reported. Using the lm() function, the model equation in R was: Oddity A score ~ 2AFC principal component 1 + 2AFC principal component 2 + VAS principal component 1 + VAS principal component 2 + AX-CPT bin score + Digit span level. Hypothesis 2 posited that oddity scores would be predicted by the 2AFC and VAS measures even after accounting for the control predictors.

***Results***

Hypothesis 2 was not supported; the anticipated predictors did not significantly predict oddity scores. The multiple regression model for Experiment 1 is summarized in Table 5, and model validation plots can be found in Supplemental Figures S.8 to S.10. For Experiment 1, the first principal component derived from the 2AFC measures was a significant predictor; however, with the increased power obtained in Experiment 2, this relationship disappeared. For Experiment 2, none of the predictors was significant (the first principal component derived from the VAS measures showed the largest coefficient but did not reach significance, $\hat{\beta} = 0.042$, $p = 0.076$; see Supplemental Table S.6).

For all of the regression models run for these two experiments, we checked for influential participants as indicated by Cook's distance. In the case of the multiple regression model for Experiment 2, one participant was found to have higher influence than the others (see Supplemental Figure S.11); upon further examination, this participant interpreted the VAS task differently, responding primarily at the endpoints of the slider rather than along its entire range. A model was run excluding this high-influence participant, since this individual did not appear to be representative of the behaviour of our sample. This additional model is summarized in Table 5, along with model validation plots in Supplemental Figures S.8 to S.10. The first principal component derived from the VAS measures—primarily reflecting VAS consistency—was a significant predictor ($\hat{\beta}$ = 0.096, *p* = 0.002). The relationship between non-native perception and VAS consistency (averaged across both contrasts) is displayed in Figure 5, revealing how more consistent VAS responses were associated with better non-native discrimination. The original model including the influential participant can be found in Supplemental Table S.6, along with model validation plots in Supplemental Figure S.11.

**Table 4**

*Correlations between the original 2AFC variables and the first two principal components extracted from them (left), and between the original VAS variables and the first two principal components extracted from them (right)*

**Experiment 1**

| | PC1 | PC2 | | PC1 | PC2 |
|---|---|---|---|---|---|
| 2AFC bet-bat acoustic cue A | 0.466 | -0.223 | VAS bet-bat slope | 0.421 | -0.076 |
| 2AFC bet-bat acoustic cue B | 0.290 | 0.952 | VAS dear-tear slope | 0.222 | 0.818 |
| 2AFC dear-tear acoustic cue A | 0.592 | -0.112 | VAS bet-bat consistency | -0.593 | -0.287 |
| 2AFC dear-tear acoustic cue B | 0.590 | -0.180 | VAS dear-tear consistency | -0.650 | 0.492 |
| *Percent variance explained* | 62% | 22% | *Percent variance explained* | 36% | 30% |

**Experiment 2**

| | PC1 | PC2 | | PC1 | PC2 |
|---|---|---|---|---|---|
| 2AFC bet-bat acoustic cue A | -0.360 | 0.537 | VAS bet-bat slope | 0.399 | -0.359 |
| 2AFC bet-bat acoustic cue B | -0.169 | 0.766 | VAS dear-tear slope | 0.257 | 0.773 |
| 2AFC dear-tear acoustic cue A | -0.647 | -0.266 | VAS bet-bat consistency | -0.625 | -0.321 |
| 2AFC dear-tear acoustic cue B | -0.651 | -0.232 | VAS dear-tear consistency | -0.620 | 0.414 |
| *Percent variance explained* | 52% | 29% | *Percent variance explained* | 41% | 29% |

**Table 5**

*Summary of the multiple regression model predicting oddity A scores, for each experiment*

**Experiment 1**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 1.731 | 0.047 | 37.146 | <.001 |
| 2AFC principal comp. 1 | 0.071 | 0.033 | 2.138 | 0.040 |
| 2AFC principal comp. 2 | -0.038 | 0.050 | -0.754 | 0.456 |
| VAS principal comp. 1 | 0.033 | 0.042 | 0.782 | 0.440 |
| VAS principal comp. 2 | -0.034 | 0.044 | -0.779 | 0.441 |
| AX-CPT bin score | -0.022 | 0.052 | -0.436 | 0.666 |
| Backwards digit span | 0.030 | 0.050 | 0.607 | 0.548 |
| *Multiple $R^2$ = 0.233, Adjusted $R^2$ = 0.094, Residual SE = 0.292 (df = 33), n = 40* | | | | |

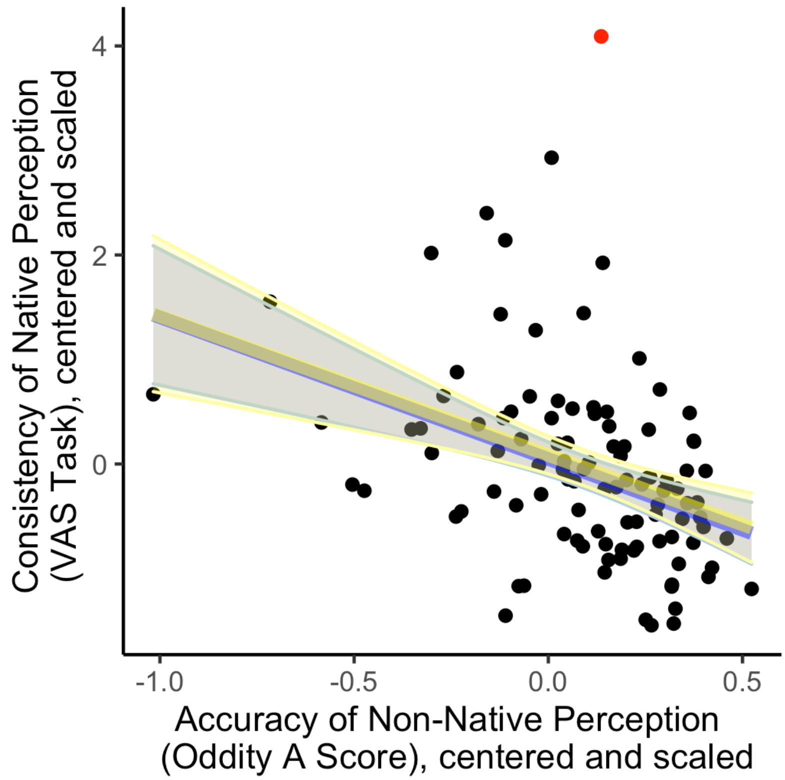
**Experiment 2**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 0.079 | 0.025 | 3.153 | 0.002 |
| 2AFC principal comp. 1 | -0.010 | 0.021 | -0.452 | 0.653 |
| 2AFC principal comp. 2 | -0.002 | 0.023 | -0.073 | 0.942 |
| VAS principal comp. 1 | 0.096 | 0.030 | 3.257 | 0.002 |
| VAS principal comp. 2 | -0.021 | 0.028 | -0.737 | 0.463 |
| AX-CPT bin score | -0.053 | 0.028 | -1.896 | 0.061 |
| Backwards digit span | -0.056 | 0.029 | -1.925 | 0.058 |
| *Multiple $R^2$ = 0.211, Adjusted $R^2$ = 0.155, Residual SE = 0.232 (df = 84), n = 91* | | | | |

*Note.* Model equation: Oddity A score (exponentially transformed for Experiment 1 but not for experiment 2, based on comparisons of model performance) ~ 2AFC principal component 1 + 2AFC principal component 2 + VAS principal component 1 + VAS principal component 2 + AX-CPT bin score + Backwards digit span.

**Figure 5**

*Relationship between non-native discrimination and native VAS consistency*




*Note.* Data are from Experiment 2. Higher values indicate less consistency and better non-native perception. Each dot represents a participant, with the outlier excluded from analyses in red. In blue is the line of best fit with 95% CI when the outlier is excluded, and in yellow is the line of best fit with 95% CI when the outlier is included.

## Non-Preregistered Analyses

In addition to the analyses that were preregistered on the OSF, a variety of additional analyses were run. The details of all of these analyses can be found in the R Markdown document on the OSF. Together with the preregistered analyses, these analyses provided a more in-depth understanding of how individual variability is structured across the two tasks.

### ***Comparing Slope Estimate Methods***

In our analyses above, as in Kapnoula et al. (2017), we found that the 2AFC slopes and the VAS slopes were not correlated across individuals, which seems to indicate that they are not measuring the same aspect of performance. However, as discussed in the introduction, they are using different methods to measure slope, and thus they might not be directly comparable. We therefore extended the work of Kapnoula et al. (2017) by fitting their rotated logistic function to the 2AFC data (as was done in Ou et al., 2021) as well as to the VAS data. This enabled a more direct comparison of slopes across the two tasks. Because we had more than one slope variable per task, we compared slope estimates across tasks using canonical correlation. As mentioned in the *Analysis* section for hypothesis 1, canonical correlation evaluates the strength of relationships between two sets of variables and outputs canonical correlation coefficients representing the strength of canonical dimensions. Canonical dimensions are combinations of the original sets of variables, weighted in such a way as to maximize the correlation between sets. Canonical correlation revealed that 2AFC slopes were significantly related across the two calculation methods (mixed-effects logistic regression vs. rotated logistic function), with a large effect size for Experiment 1 ($r_c = 0.63$, $p < 0.001$ for the first canonical dimension) and medium-large effect size for Experiment 2 ($r_c = 0.38$, $p < 0.005$ for the first canonical dimension). Further canonical correlations were then used to determine whether the new 2AFC rotated logistic slopes related to VAS slopes and consistency in similar ways to the original 2AFC mixed-effects regression slopes. These analyses revealed that the relationship between 2AFC rotated logistic slopes and VAS slopes was small to small-medium and did not reach significance, in line with the results from our preregistered analyses (first canonical dimension: $r_c = 0.23$, $p = 0.92$ for Experiment 1, and $r_c = 0.08$, $p = 0.59$ for Experiment 2). This means that the different ways of measuring slope in the two tasks cannot account for the lack of evidence for a relationship between them. We note that, unlike the mixed-effect regression slopes, the 2AFC rotated logistic slopes showed a small and statistically insignificant relationship to VAS consistency (first canonical dimension: $r_c = 0.05$, $p = 0.88$ for Experiment 1, and $r_c = 0.18$, $p = 0.33$ for Experiment 2). However, they do pattern together in the PCA analysis discussed in the *Dimensionality Reduction* section below.

***Relating Slopes and Consistency Within 2AFC and VAS Tasks***

In our preregistered analyses we compared the predictability of slopes versus consistency measures and in Figure 2 we illustrated examples of participants with all four combinations of high and low consistency and steep and shallow slopes. However, we don't know to what extent these measures are independent of each other. It could be the case that gradient perception facilitates highly consistent responses through providing a detailed and accurate phonetic representation. On the other hand, it could be that those who tend to use just the endpoints of the continuum are the most consistent. These two possibilities would give very different interpretations to the consistency measure. To better understand response consistency, canonical correlations were used to examine the relationship between slopes and consistency *within* each task. These correlations revealed that the relationship between 2AFC slopes (as calculated by the rotated logistic) and 2AFC consistency was large and significant for Experiment 1 (first canonical dimension: $r_c = 0.68$, $p < 0.001$) and medium-large and significant for Experiment 2 (first canonical dimension: $r_c$

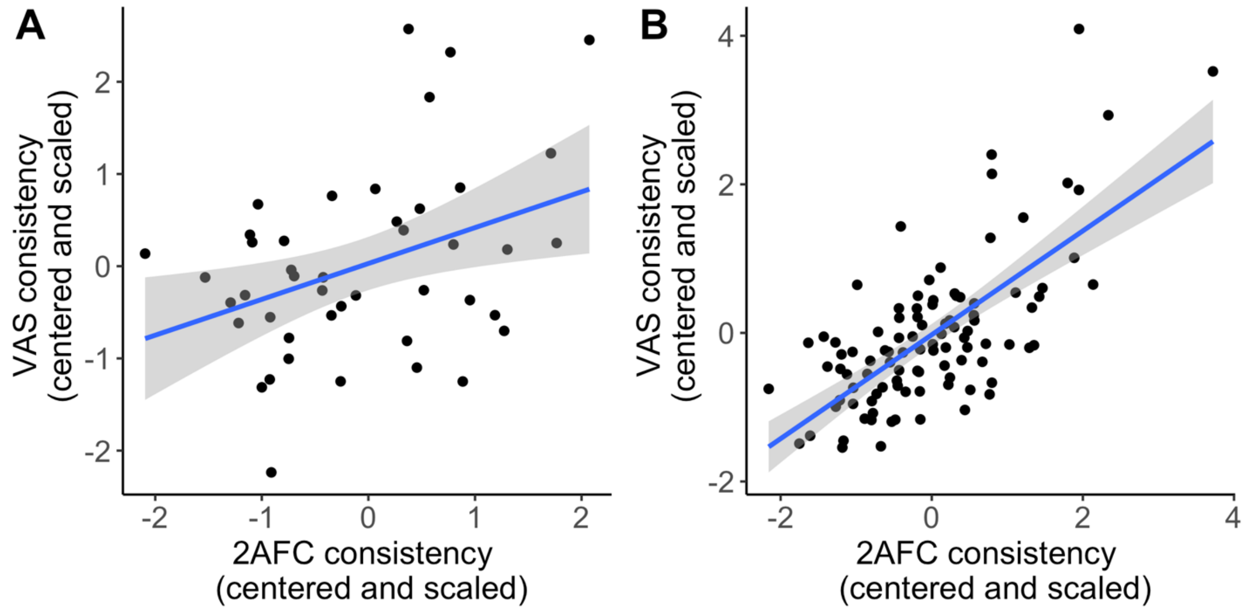
= 0.38, $p < 0.001$). Similarly, the relationship between VAS slopes and VAS consistency was medium-large and significant for Experiment 1 (first canonical dimension: $r_c = 0.44$, $p = 0.017$) and large and significant for Experiment 2 (first canonical dimension: $r_c = 0.57$, $p < 0.001$). Specifically, steeper 2AFC slopes were associated with more consistent 2AFC responses, and shallower VAS slopes were associated with more consistent VAS responses. This is an important observation that we will return to in the discussion.

***Relating Consistency Across Tasks***

Since we now had consistency measures for both tasks, we also examined the relationship between consistency across tasks. Using canonical correlation, we related 2AFC consistency to VAS consistency in order to determine whether some individuals generally show more consistent phonetic perception than others. This analysis showed that the relationship between consistency on the two tasks was large and significant along the first canonical dimension ($r_c = 0.58$, $p < 0.001$ for Experiment 1, and $r_c = 0.70$, $p < 0.001$ for Experiment 2) and medium-large and significant along the second canonical dimension ($r_c = 0.34$, $p = 0.018$ for Experiment 1, and $r_c = 0.41$, $p < 0.001$ for Experiment 2), suggesting a robust relationship. Figure 6 displays the significant relationship between 2AFC and VAS consistency for both experiments.

**Figure 6**

*Relationship between 2AFC consistency and VAS consistency*




*Note*. Data are presented from Experiment 1 (A) and Experiment 2 (B). Consistency is averaged across the two contrasts presented in the experiments, and higher values indicate less consistency. Each dot represents a participant, and the blue line is a line of best fit with 95% CI.

***Dimensionality Reduction of 2AFC and VAS Variables***

The above analyses suggest that shallow VAS slopes, steep 2AFC slopes (measured by mixed-effect logistic regression), and consistent responses all pattern together across individuals. Our final analysis confirmed this overall picture by putting all 12 variables (two VAS slopes, two VAS consistency measures, four 2AFC mixed-effects regression slopes, two 2AFC rotated logistic slopes, and two 2AFC consistency measures) into a PCA analysis to see how well they could be reduced to a smaller set of dimensions. Correlations between the original variables and the first five principal components derived from them are displayed in Supplemental Table S.7, and biplots are displayed in Supplemental Figure S.12. We found that the first principal component was made up primarily of the four consistency measures and the slope measures from the mixed-effect logistic regression of the 2AFC task (with opposite signs from the consistency measures). This confirms that differences in consistency (Figure 6) and their relationship to categorization steepness (right side of Figure 4) capture the greatest amount of variability between individuals. The second principal component shows a similar pattern, with the mixed-effect slopes for the 2AFC task patterning opposite to all of the rotated logistic values (including slope this time as well as consistency for both tasks). The second component also reflects a distinction between the two contrasts, as the bet-bat and dear-tear measures have different signs. Thus, the PCA analysis confirms the patterns observed in the previous canonical correlation analyses and provides a coherent picture of the structure of individual variability in these tasks.

## Discussion

The objectives of the current studies were twofold. First, we aimed to clarify whether responses on 2AFC and VAS tasks reflect distinct individual differences in native speech sound perception, with VAS slopes relating to gradiency and 2AFC slopes relating to consistency. We compared participants' responses to identical continua of stimuli on a 2AFC and a VAS task and found that there was no evidence for a relationship between 2AFC identification slopes and VAS identification slopes, but there was a relationship between 2AFC identification slopes and the consistency of VAS responses. Thus, for the first time the findings clearly show that the two tasks measure separate constructs: 2AFC slopes tap into the consistency of perception, while VAS slopes tap into the gradiency of perception.

Second, we aimed to determine whether discrimination of difficult non-native contrasts could be predicted by differences in native phonetic perception as measured by 2AFC and VAS tasks. While we did not find evidence for a relationship between gradiency and non-native perception, we found preliminary evidence that consistent native perception may play a role in discriminating unfamiliar language sounds.

**Identification Slopes on 2AFC and VAS Tasks Reflect Different Constructs**

Recall that there is ambiguity as to what 2AFC slopes represent, since it is unclear whether a participant with a shallow 2AFC slope (1) has underlyingly gradient perception and is responding probabilistically across trials, or (2) is responding inconsistently across trials. VAS tasks can disambiguate the constructs of gradiency and consistency, and so by comparing VAS performance to 2AFC performance we can determine how the tasks are related and which individual differences each one seems to be measuring.

Based on a marginal relationship between 2AFC slopes and consistency of VAS responses, Kapnoula et al. (2017) proposed that 2AFC and VAS tasks assess different aspects of speech perception. We hypothesized that the two tasks do indeed measure distinct constructs—with 2AFC slopes largely reflecting consistency of perception and VAS slopes largely reflecting gradiency of perception—and that this result might emerge more clearly with some methodological modifications and a large sample size. Instead of presenting continua with different numbers of steps on the two tasks as in Kapnoula et al. (2017), we used exactly the same stimuli in both tasks to facilitate comparison of performance across tasks. Our stimuli included both vowels and consonants rather than consonants alone, increasing the generalizability of the results. We also derived 2AFC and

VAS slopes both using different analysis methods (by-participant random slopes from mixed-effects logistic regression vs. slopes from a rotated logistic function developed by Kapnoula et al. (2017)) and using the same rotated logistic function across tasks, which enabled a more direct comparison of slopes than in previous studies.

It was important to conceptually replicate Kapnoula et al. (2017)'s work in order to advance the field by determining which individual differences are measured by different tasks. The relationship that they reported between 2AFC slopes and VAS consistency could have been spurious, especially given that it was marginal; if the two measures were in fact not related, this would leave us without an understanding of what 2AFC slopes are truly measuring (not consistency or gradiency, but some other construct). On the other hand, if a clearer relationship did emerge between 2AFC slopes and VAS consistency after the implementation of a few methodological changes, this would imply differences in what each task is measuring and would have repercussions for speech perception researchers in terms of which tasks and measures to employ.

In Experiment 1, we found that 2AFC slopes related to VAS consistency as hypothesized, but unexpectedly they also related to VAS slopes. This finding may have been a spurious one due to limited power, because when re-running the analyses with a larger sample size in Experiment 2, we found evidence for a relationship between 2AFC slopes and VAS consistency but not between 2AFC slopes and VAS slopes, in line with our hypothesis. Importantly, in both studies, the relationships between 2AFC slopes and VAS consistency were statistically significant (not only marginal as had previously been found), showing replicability of this finding. These relationships also held across both studies after taking into account individual differences in attention and working memory.

It could be argued that the lack of evidence for a relationship between 2AFC and VAS slopes in Experiment 2 was due in part to the different methods used to calculate slopes on each task (regular logistic mixed-effects regression for the 2AFC task vs. rotated logistic function for the VAS task). In order to provide a more direct comparison of slopes across tasks, our non-preregistered analyses involved fitting the rotated logistic function from Kapnoula et al. (2017) to both 2AFC and VAS data. This approach enabled us to derive slope and consistency measures in the same way for both tasks, yielding insight into the relationships between slopes and consistency across tasks. Even when calculated using the same rotated logistic method, there was no evidence for a relationship between slopes on the 2AFC and VAS tasks. This finding is also striking given that participants were responding to identical stimuli in both tasks. These analyses provide further evidence that 2AFC and VAS slopes reflect different constructs, strengthening the findings from our preregistered analyses.

The fact that all participants completed the VAS task prior to the 2AFC task (following the procedure described by Kapnoula et al., 2017) could potentially be viewed as a limitation due to the possibility that participants adapted to the stimuli from one task to the next. In their work which measured lexical effects on speech perception over two sessions, Giovannone & Theodore (2023) found that participants showed a weakened Ganong effect from the first to the second session, suggesting increased reliance on acoustic-phonetic information and decreased reliance on lexical information over time (though this was not the case for other tasks such as phoneme restoration). If listeners do indeed tend to increase their reliance on acoustic-phonetic information the more they are exposed to stimuli, this could potentially affect performance on our native perception tasks. However, this possibility would be more of a concern if the 2AFC task had been completed before the VAS task; in that case, increased acoustic-phonetic reliance during the second task could have resulted in more gradient response tendencies and therefore shallower VAS slopes (although this effect might have been counteracted by the categorical 2AFC task which could bias participants to mainly respond at the VAS slider's endpoints instead of along its whole length). In the case of the 2AFC task being completed second, greater gradiency/reliance on acoustic-phonetic information should not affect responses because we have found 2AFC slopes to be reflective of response consistency rather than of gradiency. Thus, we maintain that the choice to always present the VAS task before the 2AFC task was a theoretically and methodologically sound one.

**Individual Differences in the Consistency of Native Perception**

The present work clarifies that shallow 2AFC slopes appear to reflect inconsistent rather than gradient perception. This finding is in line with recent electrophysiological work that related participants' 2AFC slopes to measures of their subcortical and cortical auditory encoding (Ou & Yu, 2022). The researchers found that participants with less faithful subcortical encoding of speech had shallower 2AFC slopes, supporting the notion that 2AFC slopes reflect inconsistency in perception (Ou & Yu, 2022). Additionally, our results shed light on why previous studies have suggested an association between shallow 2AFC slopes and language impairment—such an impairment appears to be accompanied by inconsistency or imprecision in perception. This conclusion is further supported by work showing that children with developmental dyslexia, who are known to have shallower 2AFC slopes (Manis et al., 1997; Joanisse, et al., 2000; Serniclaes et al., 2001), also have inconsistent or atypical neural representation of speech (Destoky et al., 2020; Keshavarzi et al., 2022; Power et al., 2016).

Interestingly, our non-preregistered analyses revealed that participants' response consistency values (as extracted from the rotated logistic function from Kapnoula et al. 2017) were related across the 2AFC and VAS tasks. Response consistency also patterned together across the vowel and consonant contrasts in the PCA analysis. These findings suggest that consistency may be a stable and task-independent property of the individual. While previous work has demonstrated individual differences in consistency on VAS tasks (Kapnoula et al., 2017), we provide evidence that these differences seem to hold across tasks. That is, some listeners appear to be more consistent than others in how they map perceived speech sounds to response options, regardless of the specific format of the response options. An interesting topic for future study could be how and why consistency may reflect optimal perception, as well as the extent to which differences in consistency of perception generalize to other tasks (e.g., speech-in-noise perception, assimilation of non-native sounds to native categories) and other modalities (e.g., ratings of colour stimuli).

An important question for future research is at which level the consistency measured by phonetic perception tasks arises (i.e., whether it is somewhere along the perceptual pathway and/or during higher-level decision-making processes). The work by Ou & Yu (2022) suggests that early subcortical auditory encoding of sound is a source of consistency, but they also found that steep slopes on a 2AFC task were further related to a difference in the representation between cortical and subcortical encoding. This seems to indicate that steep slopes require accurate gradient encoding and consistent transformation into categories. This suggests that perhaps consistency at higher levels of perception and cognition may play an additional role in predicting individual differences in responses, for example through attention or memory. It would also be of interest to investigate whether atypical and typically-developing populations show similar or different sources of inconsistency.

**Consistency and Gradiency as Distinct Yet Related Constructs**

Separately from consistency, gradiency of perception seems to be its own construct that is best measured by VAS tasks and that may be adaptive (rather than suggestive of an impairment) in various situations as described further below. As discussed by Kapnoula et al. (2017), gradiency and consistency may be orthogonal, and VAS tasks are useful precisely because they enable researchers to calculate a separate measure of each construct. A conceptual distinction between consistency and gradiency makes sense given recent evidence that measures of the two constructs (as extracted from a VAS task) have separate structural correlates in the brain (Fuhrmeister & Myers, 2021). Thus, there appears to be a difference between distinguishing gradual changes along a continuum (gradiency) and having highly reliable mapping between stimulus and response (consistency). This being said, our results do suggest that the constructs relate to one another. We found that listeners with more consistent responses tended to have steeper slopes on the 2AFC task and shallower slopes on the VAS task. This outcome probably reflects an optimal pattern of perception whereby the listener shows categorical responses when presented with a categorical task and gradient options when presented with a gradient task. The most successful listeners therefore appear to be the ones who are consistently able to map their percept to a response option, which promotes precise and optimal responding across tasks.

Although the mechanisms of both consistency and gradiency remain to be elucidated, based on existing work we can speculate that they may have partially distinct and partially overlapping underpinnings which could explain our findings. Behavioural response consistency may arise at least to some extent from neural response consistency, which can be quantified as the similarity of the evoked neural response across repeated presentations of a sound (Krizman & Kraus, 2019; Ou & Yu, 2022). Differences in gradiency may also arise partly from this same neural response consistency, with more similar neural responses promoting more gradient perception by facilitating the faithful encoding of subtle differences between stimuli; but gradiency may additionally result from the transformations that the neural response undergoes as it travels up the auditory pathway from the brainstem to higher-level cortex, with greater transformation leading to less gradient perception (Ou & Yu, 2022)—this process is referred to as perceptual warping by Kapnoula et al. (2021). Under this possibility, gradiency and consistency share some basic mechanisms and would relate to each other as found in the present work; yet two listeners with equally consistent neural and behavioural responses could still differ in gradiency based on how their neural responses were transformed along the auditory pathway. Nevertheless, this explanation remains purely theoretical, and future work with neural measures will be needed to determine the precise origins of both constructs and to untangle the nature of the relationship between them.

**Beyond Categorical Perception and Categorical Tasks**

The current findings add to the growing conviction that psycholinguistics should move beyond a purely categorical view of phonetic perception (e.g., Holt & Lotto, 2010; Kapnoula et al., 2017; McMurray, 2022; McMurray et al., 2002; Schouten et al., 2003). We support the view that gradiency can be a beneficial (not suboptimal) strategy during perception (Clayards et al., 2008; Desmeules-Trudel & Zamuner, 2019; Kapnoula et al., 2021). In fact, both categorical and gradient modes of perception are likely to be useful in their own way: the ability to fit sounds into one category or another appears to be an important part of processing sounds efficiently (e.g., Shen & Froud, 2016), and the ability to distinguish within-category differences seems to promote flexibility during perception (e.g., Kapnoula et al., 2021). A given listener's sensitivity to between- versus within-category differences in speech sounds likely depends on idiosyncrasies of their perceptual systems (Kapnoula & McMurray, 2021) and varies according to the particular context (e.g., more between-category sensitivity when listening to predictable native input that easily fits one's preestablished categories, and more within-category sensitivity when listening to accented or non-native speech that requires perceptual flexibility). In other words, listeners may use different strategies—of which gradiency is one—to arrive at the common goal of deriving concrete representations from the continuous speech signal. Beyond its role in encouraging flexible phonetic perception, gradiency is also no doubt important for the perception of various social factors related to a given speaker, such as emotion (Cowen et al., 2019), geographic dialect (Plichta & Preston, 2005), and perceived masculinity/femininity (Munson, 2007). Considering that all of these social factors exist along continua, it is logical that perceiving them in an accurate and nuanced way would require gradient acoustic representations. The potential sources and functions of gradiency, as well as the context-dependent ways in which it may be combined with other strategies during the perception of speech and of speakers, are pertinent questions to continue exploring with future research.

Our findings have important methodological implications in psycholinguistics and related fields. Notably, researchers should select tasks carefully based on the constructs that they wish to measure, while taking into account the limitations and demands of different tasks. When looking to study gradiency, VAS tasks (with their continuous gradient of response options) are a much more appropriate choice than 2AFC tasks (see Apfelbaum et al., 2022 for further discussion of VAS tasks and their utility). The development and adoption of tasks that encourage more fine-tuned and gradated responses, such as VAS and magnitude estimation tasks (Sprouse, 2007), seem to be an important step in advancing psycholinguistic research by revealing nuances of human perception and cognition that may not otherwise be captured by tasks with limited response options.

**Predictors of Non-Native Perception**

We hypothesized that shallow VAS slopes and steep 2AFC slopes might both be indicative of the ability to make accurate and fine-tuned judgments about acoustic cues and might therefore relate to better non-native discrimination. This hypothesis was not supported. With our preregistered analyses, we found that performance on the non-native perception task was not robustly predicted by any of the native perception measures across our two studies. However, additional analysis excluding an influential participant in Experiment 2 revealed a potential relationship between VAS consistency and non-native perception. This relationship held even after accounting for non-linguistic cognitive factors (attention and working memory). This finding implies that in order to successfully distinguish new speech sounds, an important underlying factor is not so much the exact nature of native speech sound representations (categorical/gradient), but rather the similarity of these representations across time. While not anticipated, such a link between consistent native perception and accurate non-native perception is reasonable when considered in the context of the latest literature on non-native perception.

Very recent work by Fuhrmeister et al. (2023) is in line with our findings. Similarly to us, the authors hypothesized that more gradient VAS slopes on a native phonetic perception task would relate to better discrimination of non-native phonemes; and yet they found that more *consistent* VAS responses related to better non-native discrimination. In addition, preliminary work by Kapnoula & Samuel (2023) has revealed the same pattern of results: better non-native perception was predicted by more consistent VAS responses rather than by more gradient VAS slopes. Across our experiments and other recent research, the same picture is therefore emerging: in order to discriminate non-native speech sounds, it appears to be helpful to have a strong link between a stimulus and one's response to it. As mentioned above, the level at which such consistency emerges remains to be clarified. Consistency in auditory brainstem responses relates to preliteracy skills (Bonacina et al., 2021; White-Schwoch et al., 2015) and to phonetic discrimination

(Tecoulesco et al., 2020), so it is possible that consistency begins playing a role at the level of early neural encoding and is an important element of native and non-native language acquisition.

Further insight comes from studies that have asked participants to listen to native sounds and assimilate them to non-native categories. Such studies have shown that the ability to consistently map a given non-native phoneme to a particular native category is related to having greater non-native perceptual proficiency (i.e., patterns of acoustic cue weighting during non-native perception that more closely resemble those of native speakers; Kang & Schertz, 2021), a larger non-native vocabulary (Bundgaard-Nielsen et al., 2011), and more extensive experience with the non-native language (Levy, 2009). It therefore seems that consistency of phonetic perception can predict various outcomes of non-native language learning success.

While we are unaware of any existing theories which might explain the precise nature of the relationship between native perceptual consistency and non-native perceptual success, the category precision hypothesis of the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021) addresses a similar relationship in the context of speech production rather than perception. According to this hypothesis, the more consistent and precise a person's native categories are (in this case, consistency being defined as low acoustic variability across multiple productions of a phoneme), the better the person will be at distinguishing new non-native sounds and establishing categories for them. Based on our findings, it is conceivable that a similar hypothesis might apply in the realm of perception, where listeners with more consistent and precise native perception can more readily perceive differences between non-native sounds.

Although our preregistered analyses revealed a somewhat surprising lack of evidence for a relationship between native and non-native perception, similar findings have been reported in the past. For instance, other work has found that gradiency of native perception on a VAS task did not relate to non-native discrimination ability (Fuhrmeister et al., 2023) or to scores on a standardized non-native proficiency task (Kong & Kang, 2022). It may be that native gradiency does not relate strongly to non-native outcomes due to differences in some of the processing strategies involved. This possibility is supported by work showing that native and non-native listeners rely on different strategies—namely, lexical knowledge vs. acoustic cues—during word segmentation (Mattys et al., 2010). It has also been found that native speakers show gradient integration of phonetic information (as measured by eye-tracking) during word recognition, whereas non-native speakers show a categorical pattern (Desmeules-Trudel, 2018). An additional possibility is that greater sensitivity to native speech sounds does promote better non-native perception, but that this relationship emerges later in life. In line with this, Kalaivanan et al. (2023) recently found that for older adults, native perceptual sensitivity (as measured by a gating task) was a robust predictor of non-native discrimination; but for younger adults, general intelligence was a stronger predictor. Perhaps younger adults (like the participants in the present experiments) rely more on fluid cognitive factors including attention and memory, while older adults rely more on crystallized factors including their knowledge of native phonemes (Spreng & Turner, 2019). Future work with older populations could clarify this possibility.

The lack of evidence for a strong relationship observed between our native and non-native perception measures could also be due in part to differences in the tasks used to derive the measures. As an example, on the native perception tasks the stimuli had been manipulated to form a continuum, and each trial involved the presentation of one stimulus; on the non-native perception task the stimuli were not manipulated, and each trial involved the presentation of three stimuli. The fact that we do see some relationships between native and non-native performance despite the differences in tasks suggests that the relationship may be even more robust when more similar measures are used. Future work could compare native and non-native performance more directly, for example by training non-native perception in advance so that participants can respond to non-native sounds on 2AFC and VAS tasks, or by measuring both native and non-native perception using oddity tasks.

It is also worth pointing out that some of the individual variability observed on our native and non-native perception tasks could have arisen from differences in participants' sociolinguistic knowledge and/or labelling strategies. For instance, the *bet-bat* contrast that we tested here is known to participate in ongoing sound change processes such as the Northern Cities Vowel Shift (McCarthy, 2011) and the California Vowel Shift (D'Onofrio et al., 2019). Given that our participants were recruited from across North America, their varied sociolinguistic knowledge could have contributed to some of the differences in performance observed on the native perception tasks. In the future, it would be interesting to measure sociolinguistic factors and relate them to the kinds of individual differences observed here. Additionally, performance on the non-native perception task could have been influenced by whether participants treated the speech sounds as entirely unfamiliar or as better/worse exemplars of native sounds. As an example, a participant that perceived German /ç/ as a new and unfamiliar sound may have been more successful on our task compared to one that perceived /ç/ as a bad exemplar of English /ʃ/ and consequently assimilated /ç/ and /ʃ/ to the same category. In accordance with this possibility, Mayr & Escudero (2010) have shown that participants who assimilated German contrasts to a single English category (rather than to two distinct categories) had more difficulty identifying those contrasts. By incorporating other tasks where non-native sounds must be labelled or rated for category goodness, future studies could uncover more nuances of the factors relating to individual differences in non-native perception. Note that we do not view these possible sources of variability in native and non-native perception as limitations; while they may have contributed to the variation in performance that we observed, they do not invalidate the relationships we found.

The present work and work by Fuhrmeister et al. (2023) suggest that non-native perception is predicted by the consistency of native perception. If this finding continues to be replicated, it could provide an exciting avenue for further exploration. For instance, perhaps native perception tasks could be administered as brief pre-screenings in language learning settings as a means of identifying people who would benefit from greater support during the learning process. In any case, an important topic that remains to be addressed is why healthy young adults show such variability in their ability to discriminate non-native phonemes. Work in this area is particularly relevant given that successful phonetic perception appears to be a precursor for language learning more generally, predicting outcomes such as non-native vocabulary learning and reading comprehension (Jakoby et al., 2011; Silbert et al., 2015).

## Conclusion

In summary, we demonstrated that identification slopes on 2AFC and VAS tasks do not measure the same individual differences in phonetic perception. While shallow VAS slopes seem to reflect gradient perception that involves fine-tuned sensitivity to within-category differences, shallow 2AFC slopes seem to reflect inconsistent perception. This is important given that 2AFC tasks have been extensively employed in previous work and that their slopes have been thought to support the theory of categorical perception, when in fact the slopes were not necessarily measuring what researchers intended. This work points to the necessity of accounting for task demands during research and of revising theoretical views in light of new evidence. We join other researchers in recommending the use of VAS tasks rather than commonly used 2AFC tasks in psycholinguistic research, and in encouraging views of phonetic perception that account for within-category sensitivity (Apfelbaum et al., 2022; Kapnoula et al., 2017; McMurray et al., 2002; Munson et al., 2017).

Our analyses also pointed to the construct of consistency as a fruitful subject for future investigation. We found that consistency of responses was related across the 2AFC and VAS tasks, suggesting that it may be a stable property of the individual. We further found that consistent responses were associated with steeper 2AFC slopes and shallower VAS slopes. This pattern seems to indicate that people who can consistently associate a given stimulus with a response show the most optimal pattern of perception across tasks (categorical responses on the 2AFC task, gradient responses on the VAS task).

Finally, we found preliminary evidence that successful non-native phonetic perception may be predicted by the consistency of VAS responses. In the future, this could lead to the development of personalized methods of assisting adult language learners based on their individual perceptual and cognitive profiles. The potential benefits of personalized approaches to learning become evident when considering the notable individual differences in performance that are observed across various phonetic perception and cognitive tasks, both here and in other work (e.g., Golestani & Zatorre, 2009; Hanulíková et al., 2012; Hattori & Iverson, 2010; Lee, 2016; Linck & Weiss, 2015). Furthermore, optimizing language learning in adults is particularly relevant in today's highly diverse and interconnected world, in which learning new languages has become key for many people's social integration and advancement.

## References


Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x.

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728-3745. https://doi.org/10.1121/10.0015201

Barch, D. M., Berman, M. G., Engle, R., Jones, J. H., Jonides, J., MacDonald III, A., ... & Sponheim, S. R. (2009). CNTRICS final task selection: working memory. *Schizophrenia bulletin*, *35*(1), 136-152.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,67*(1). https://doi.org/10.18637/jss.v067.i01.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13-34). Amsterdam: John Benjamins.

Bonacina, S., Huang, S., White-Schwoch, T., Krizman, J., Nicol, T., & Kraus, N. (2021). Rhythm, reading, and sound processing in the brain in preschool children. *npj Science of Learning*, *6*(1), 20. https://doi.org/10.1038/s41539-021-00097-5.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Brietzke, C., Vinícius, Í., Franco-Alvarenga, P. E., Canestri, R., Goethel, M. F., Santos, L. E. R., ... & Pires, F. O. (2021). Proof-of-concept and test-retest reliability study of psychological and physiological variables of the mental fatigue paradigm. *International Journal of Environmental Research and Public Health*, *18*(18), 9532.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in* *second language acquisition*, *33*(3), 433-461. https://doi.org/10.1017/S0272263111000040.

Champely, S. (2020). pwr: Basic Functions for Power Analysis. R package version 1.3-0. https://CRAN.R-project.org/package=pwr.

Christopherson, L. A., & Humes, L. E. (1992). Some psychometric properties of the Test of Basic Auditory Capabilities (TBAC). *Journal of Speech, Language, and Hearing Research*, *35*(4), 929-935.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological assessment*, *6*(4), 284.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *J. Acoust. Soc. Am.*, *144*(3), EL172-EL177. https://doi.org/10.1121/1.5052025.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804-809. https://doi.org/10.1016/j.cognition.2008.04.004.

Coffey, E.B.J., Herholz, S.C., Scala, S. and Zatorre, R.J., 2011, June. Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research. In *The Neurosciences and Music IV: Learning and Memory, Conference. Edinburgh, UK*.

Conners, C. K., Epstein, J. N., Angold, A., & Klaric, J. (2003). Continuous performance test performance in a normative epidemiological sample. *Journal of abnormal child psychology*, *31*, 555-562. https://doi.org/10.1023/A:1025457300409.

Cooper, S. R., Gonthier, C., Barch, D. M., & Braver, T. S. (2017). The role of psychometrics in individual differences research in cognition: A case study of the AX-CPT. *Frontiers in psychology*, *8*, 1482.

Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2019). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*, *74*(6), 698.

Cumming, R., Wilson, A., & Goswami, U. (2015). Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: A new look at a perceptual hypothesis. *Frontiers in Psychology*, *6*, 972. https://doi.org/10.3389/fpsyg.2015.00972.

Desmeules-Trudel, F. (2018). *Spoken word recognition in native and second language Canadian French: Phonetic detail and representation of vowel nasalization* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

Desmeules-Trudel, F., & Zamuner, T. S. (2019). Gradient and categorical patterns of spoken-word recognition and processing of phonetic details. *Attention, Perception, & Psychophysics*, *81*, 1654-1672. https://doi.org/10.3758/s13414-019-01693-9.

Destoky, F., Bertels, J., Niesen, M., Wens, V., Vander Ghinst, M., Leybaert, J., ... & Bourguignon, M. (2020). Cortical tracking of speech in noise accounts for reading strategies in children. *PLoS biology*, *18*(8), e3000840. https://doi.org/10.1371/journal.pbio.3000840.

D'Onofrio, A., Pratt, T., & Van Hofwegen, J. (2019). Compression in the California vowel shift: Tracking generational sound change in California's Central Valley. *Language Variation and Change*, *31*(2), 193-217.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological bulletin*, *145*(5), 508. https://doi.org/10.1037/bul0000192.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), Second language speech learning: Theoretical and empirical progress (pp. 3–83). Cambridge University Press.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470. https://doi.org/10.1006/jpho.1997.0052.

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Fox-Fuller, J. T., Ngo, J., Pluim, C. F., Kaplan, R. I., Kim, D. H., Anzai, J. A., ... & Quiroz, Y. T. (2022). Initial investigation of test-retest reliability of home-to-home teleneuropsychological assessment in healthy, English-speaking adults. *The Clinical Neuropsychologist*, *36*(8), 2153-2167.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and speech*, *5*(4), 171-189. https://doi.org/10.1177/002383096200500404.

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language*, *215*, 104919. https://doi.org/10.1016/j.bandl.2021.104919.

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001213.

Geller, J., Holmes, A., Schwalje, A., Berger, J. I., Gander, P. E., Choi, I., & McMurray, B. (2021). Validation of the Iowa test of consonant perception. *The Journal of the Acoustical Society of America*, *150*(3), 2131-2153.

Gerrits, E., & Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Perception & psychophysics*, *66*, 363-376. https://doi.org/10.3758/BF03194885.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and individual differences*, *102*, 74-78.

Giovannone, N., & Theodore, R. M. (2023). Do individual differences in lexical reliance reflect states or traits?. *Cognition*, *232*, 105320.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69-78. https://doi.org/10.1002/wcs.26.

GoldWave Inc. (2015). Goldwave (Version 6.15) [Computer software]. Retreived from: www.goldwave.com.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and language*, *109*(2-3), 55-67. https://doi:10.1016/j.bandl.2008.01.005.

González, I., & Déjean, S. (2021). CCA: Canonical Correlation Analysis. R package version 1.2.1. https://CRAN.R-project.org/package=CCA.

Grimaldi, M., Sisinni, B., Gili Fivela, B., Invitto, S., Resta, D., Alku, P., & Brattico, E. (2014). Assimilation of L2 vowels to L1 phonemes governs L2 learning in adulthood: a behavioral and ERP study. *Frontiers in human neuroscience*, *8*, 279. https://doi.org/10.3389/fnhum.2014.00279.

Halperin, J. M., Sharma, V., Greenblatt, E., & Schwartz, S. T. (1991). Assessment of the Continuous Performance Test: Reliability and validity in a nonreferred sample. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*(4), 603.

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning*, *62*, 79-109.

Harrell, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer. https://doi.org/10.1007/978-3-319-19425-7.

Hary, J. M., & Massaro, D. W. (1982). Categorical results do not imply categorical perception. *Perception & Psychophysics*, *32*(5), 409-418.

Hattori, K., & Iverson, P. (2009). English/r/-/l/category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, *125*(1), 469-479. https://doi.org/10.1121/1.3021295.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English/r/-/l/perception and production by adult Japanese speakers. In *Second Language Studies: Acquisition, Learning, Education and Technology*.

Hazan, V., & Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Perception & Psychophysics*, *49*(2), 187-200.

Heffner, C. C., & Myers, E. B. (2021). Individual differences in phonetic plasticity across native and nonnative contexts. *Journal of Speech, Language, and Hearing Research*, *64*(10), 3720-3733. https://doi.org/10.1044/2021_JSLHR-21-00004.

Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, & Psychophysics*, *72*(5), 1218-1227. https://doi.org/10.3758/APP.72.5.1218.

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior research methods*, *46*(3), 702-721. https://doi.org/10.3758/s13428-013-0411-5.

Idemaru, K., Holt, L. L., & Seltman, H. (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. *The Journal of the Acoustical Society of America*, *132*(6), 3950-3964.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57. https://doi.org/10.1016/S0010-0277(02)00198-1.

Jakoby, H., Goldstein, A., & Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology*, *48*(11), 1517-1531. https://doi.org/10.1111/j.1469-8986.2011.01227.x.

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of experimental child psychology*, *77*(1), 30-60. https://doi:10.1006/jecp.1999.2553.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and language*, *192*, 15-24. https://doi.org/10.1016/j.bandl.2019.02.004.

Kalaivanan, K., Wong, P. C., Wong, F. C., & Chan, A. H. (2023). Native Language Perceptual Sensitivity Predicts Nonnative Speech Perception Differently in Younger and Older

Singaporean Bilinguals. *Journal of Speech, Language, and Hearing Research*, *66*(3), 987-1017. https://doi.org/10.1044/2022_JSLHR-22-00199.

Kang, Y., & Schertz, J. (2021). The influence of perceived L2 sound categories in on-line adaptation and implications for loanword phonology. *Natural Language & Linguistic Theory*, *39*, 555-578. https://doi.org/10.1007/s11049-020-09477-9.

Kapnoula, E. C., Edwards, J., & McMurray, B. (2021). Gradient activation of speech categories facilitates listeners' recovery from lexical garden paths, but not perception of speech-in-noise. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(4), 578. https://doi.org/10.1037/xhp0000900.

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, *223*, 105031. https://doi.org/10.1016/j.bandl.2021.105031.

Kapnoula, E. C., & Samuel, A. G. (2023, May 31-June 2). *Examining the links between L1 phoneme categorization and non-native phonetic learning* [Poster presentation]. XVI International Symposium of Psycholinguistics, Vitoria-Gasteiz, Spain.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594. https://doi.org/10.1037/xhp0000410.

Kempe, V., Thoresen, J. C., Kirk, N. W., Schaeffler, F., & Brooks, P. J. (2012). Individual differences in the discrimination of novel speech sounds: effects of sex, temporal processing, musical and cognitive abilities. *PloS one*, *7*(11), e48623. https://doi.org/10.1371/journal.pone.0048623.

Keshavarzi, M., Mandke, K., Macfarlane, A., Parvez, L., Gabrielczyk, F., Wilson, A., ... & Goswami, U. (2022). Decoding of speech information using EEG in children with dyslexia: Less accurate low-frequency representations of speech, not "Noisy" representations. *Brain and Language*, *235*, 105198. https://doi.org/10.1016/j.bandl.2022.105198.

Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *Journal of Phonetics*, *81*, 100984. https://doi.org/10.1016/j.wocn.2020.100984.

Kogan, V. V., & Mora, J. C. (2022). The effects of individual differences in native perception on discrimination of a novel non-native contrast. *Laboratory Phonology*, *24*(1). https://doi.org/10.16995/labphon.6431.

Kong, E. J., & Edwards, J. (2015, August). Individual differences in L2 learners' perceptual cue weighting patterns. *Proceedings of the International Congress of Phonetic Sciences*.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40-57. https://doi.org/10.1016/j.wocn.2016.08.006.

Kong, E. J., & Kang, S. (2022). Individual differences in categorical judgment of L2 stops: A link to proficiency and acoustic cue-weighting. *Language and Speech*, 00238309221108647. https://doi.org/10.1177/00238309221108647.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163.

Kraus, M. S., Gold, J. M., Barch, D. M., Walker, T. M., Chun, C. A., Buchanan, R. W., ... & Keefe, R. S. (2020). The characteristics of cognitive neuroscience tests in a schizophrenia cognition clinical trial: Psychometric properties and correlations with standard measures. *Schizophrenia Research: Cognition*, *19*, 100161.

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing research*, *382*, 107779.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000. https://doi.org/10.1098/rstb.2007.2154.

Kwakkel, H., Droop, M., Verhoeven, L., & Segers, E. (2021). The impact of lexical skills and executive functioning on L1 and L2 phonological awareness in bilingual kindergarten. *Learning and Individual Differences*, *88*, 102009. https://doi.org/10.1016/j.lindif.2021.102009.

Lee, S. P. (2016). Computer-detected attention affects foreign language listening but not reading performance. *Perceptual and Motor Skills*, *123*(1), 33-45. https://doi.org/10.1177/0031512516657337

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757-3768. https://doi.org/10.1121/1.3506351.

Levy, E. S. (2009). Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *The Journal of the Acoustical Society of America*, *125*(2), 1138-1152. https://doi.org/10.1121/1.3050256.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2013). Language history questionnaire (lhq 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17,* 673-680. https://doi.org/10.1017/S1366728913000606

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358.

Linck, J.A. and Weiss, D.J. (2015) Can working memory and inhibitory control predict second language learning in the classroom? *Sage Open Journal*. https://doi.org/10.1177/2158244015607352.

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia?. *Journal of experimental child psychology*, *66*(2), 211-235.

Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech communication*, *2*(1), 15-35. https://doi.org/10.1016/0167-6393(83)90061-4.

The MathWorks Inc., “MATLAB.” Natick, Massachusetts, 2015a.

Mattys, S. L., Carroll, L. M., Li, C. K., & Chan, S. L. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech communication*, *52*(11-12), 887-899. https://doi.org/10.1016/j.specom.2010.01.005.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279-297. https://doi.org/10.1017/S1366728909990022.

McCarthy, C. (2011). The Northern Cities Shift in Chicago. *Journal of English Linguistics*, *39*(2), 166-187.

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819-3842. https://doi.org/10.1121/10.0016614.

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance, 34*(6), 1609. https://doi.org/10.1037/a0011747.

McMurray, B., Danelz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. *Developmental psychology*, *54*(8), 1472.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33-B42. https://doi.org/10.1016/S0010-0277(02)00157-9.

Menzel, U. (2012). CCP: Significance Tests for Canonical Correlation Analysis (CCA). R package version 1.1. https://CRAN.R-project.org/package=CCP.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*(1-3), 271-285.

Müller, U., Kerns, K. A., & Konkin, K. (2012). Test–retest reliability and practice effects of executive function tasks in preschool children. *The Clinical Neuropsychologist*, *26*(2), 271-287.

Munson, B. (2007). The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Language and speech*, *50*(1), 125-142.

Munson, B., Johnson, J. M., & Edwards, J. (2012). The role of experience in the perception of phonetic detail in children's speech: a comparison between speech-language pathologists and clinically untrained listeners. *American Journal of Speech-Language Pathology*, *21*(2), 124–139. https://doi.org/ 10.1044/1058-0360(2011/11-0009).

Munson, B., Logerquist, M. K., Kim, H., Martell, A., & Edwards, J. (2021). Does early phonetic differentiation predict later phonetic development? Evidence from a longitudinal study of /ɹ/ development in preschool children. *Journal of Speech, Language, and Hearing Research*, *64*(7), 2417-2437.

Munson, B., Schellinger, S. K., & Edwards, J. (2017). Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales. *Clinical Linguistics & Phonetics*, *31*(1), 56-79. https://doi.org/10.1080/02699206.2016.1233292.

Norrman, G., Bylund, E., & Thierry, G. (2022). Irreversible specialization for speech perception in early international adoptees. *Cerebral Cortex*, *32*(17), 3777-3785. https://doi.org/10.1093/cercor/bhab447.

Ou, J., & Yu, A. C. (2022). Neural correlates of individual differences in speech categorisation: evidence from subcortical, cortical, and behavioural measures. *Language, Cognition and Neuroscience*, *37*(3), 269-284. https://doi.org/10.1080/23273798.2021.1980594.

Ou, J., Yu, A. C., & Xiang, M. (2021). Individual Differences in Categorization Gradience As Predicted by Online Processing of Phonetic Cues During Spoken Word Recognition: Evidence From Eye Movements. *Cognitive Science*, *45*(3), e12948. https://doi.org/10.1111/cogs.12948.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E., & Guenther, F. H. (2004). The distinctness of speakers' /s/—/ʃ/contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1259–69. https://doi.org/10.1121/1.1787524.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & psychophysics*, *15*(2), 285-290.

Plichta, B., & Preston, D. R. (2005). The /ay/s have it: The perception of /ay/ as a north-south stereotype in United States English. *Acta Linguistica Hafniensia*, *37*(1), 107-130.

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language learning*, *64*(4), 878-912.

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*, 1-10. http://dx.doi.org/10.1016/j.bandl.2016.06.006.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rauber, A. S., Escudero, P., Bion, R. A., & Baptista, B. O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In *Ninth European Conference on Speech Communication and Technology*.

Saito, K., & Tierney, A. (2022). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 1-25.

Schellinger, S. K., Munson, B., & Edwards, J. (2017). Gradient perception of children's productions of /s/ and /θ/: A comparative study of rating methods. *Clinical Linguistics & Phonetics*, *31*(1), 80-103. https://doi.org/10.1080/02699206.2016.1205665.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of phonetics*, *52*, 183-204.

Schouten, B., Gerrits, E., & Van Hessen, A. (2003). The end of categorical perception as we know it. *Speech communication*, *41*(1), 71-80. https://doi.org/10.1016/S0167-6393(02)00094-8.

Schouten, M. E. H., & Van Hessen, A. J. (1992). Modeling phoneme perception. I: Categorical perception. *The Journal of the Acoustical Society of America*, *92*(4), 1841-1855.

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2001/032)

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, *98*(2), B35-B44. https://doi.org/10.1016/j.cognition.2005.03.002.

Shen, G., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *The Journal of the Acoustical Society of America*, *140*(6), 4396-4403. https://doi.org/10.1017/S136672891800038X.

Sherry, A., & Henson, R. K. (2005). Conducting and interpreting canonical correlation analysis in personality research: A user-friendly primer. *Journal of personality assessment*, *84*(1), 37-48.

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*, 99-119. https://doi.org/10.1016/j.wocn.2015.03.001.

Silveira, R. (2011). Pronunciation instruction and syllabic-pattern discrimination. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, *27*, 21-36. https://doi.org/10.1590/S0102-44502011000100002.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter?. *Psychological science*, *17*(8), 675-681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

Souza, P., Wright, R., Gallun, F., & Reinhart, P. (2018). Reliability and repeatability of the speech cue profile. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2126-2137.

Spreng, R. N., & Turner, G. R. (2019). The shifting architecture of cognition and brain function in older adulthood. *Perspectives on Psychological Science*, *14*(4), 523-542. https://doi.org/10.1177/1745691619827511.

Sprouse, J. (2007). Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics*, *1*, 123-134. https://doi.org/10.5964/bioling.8597.

Strand, J., Cooperman, A., Rowe, J., & Simenstad, A. (2014). Individual differences in susceptibility to the McGurk effect: Links with lipreading and detecting audiovisual incongruity. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2322-2331. https://doi.org/10.1044/2014_JSLHR-H-14-0059.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & psychophysics*, *36*(2), 131-145. https://doi.org/10.3758/BF03202673.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp.153-191). Amsterdam: John Benjamins.

Surprenant, A. M., & Watson, C. S. (2001). Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *The Journal of the Acoustical Society of America*, *110*(4), 2085-2095. https://doi.org/10.1121/1.1404973.

Tecoulesco, L., Skoe, E., & Naigles, L. R. (2020). Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. *Brain and Language*, *208*, 104810. https://doi.org/10.1016/j.bandl.2020.104810.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science*, *21*(10), 1532-1540. https://doi.org/10.1177/0956797610384142.

UCLA: Statistical Consulting Group. *Canonical Correlation Analysis: R Data Analysis Examples.* UCLA Statistical Methods and Data Analytics. https://stats.oarc.ucla.edu/r/dae/canonical-correlation-analysis/.

Wang, X., & Humes, L. E. (2008). Classification and cue weighting of multidimensional stimuli with speech-like cues for young normal-hearing and elderly hearing-impaired listeners. *Ear and hearing*, *29*(5), 725.

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*(498), 1.

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *41*(1), 48. https://doi.org/doi: 10.1037/h0084150.

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., ... & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS biology*, *13*(7), e1002196. https://doi.org/10.1371/journal.pbio.1002196.

Won, J. H., Tremblay, K., Clinard, C. G., Wright, R. A., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *139*(1), 1-11. https://doi.org/10.1121/1.4931909.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., ... & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of clinical and experimental neuropsychology*, *33*(1), 101-111.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064-2072. https://doi.org/10.3758/s13414-017-1361-2.

Yu, A. C., & Lee, H. (2014). The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *The Journal of the Acoustical Society of America*, *136*(1), 382-388. https://doi.org/10.1121/1.4883380.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*, 203-212. https://doi.org/10.1007/s11336-003-1119-8.

## Supplemental Materials

As mentioned in the main article, the raw data for both experiments, along with the code needed to process and analyze it, is publicly available on the OSF (https://osf.io/ez5qh/?view_only=8e4a1498e04f4ee0946752ee93b9ce71). The design, hypotheses, and analysis plan of Experiment 2 were preregistered based on Experiment 1, and are available on the same OSF page. Stimuli from the native phonetic perception tasks are publicly available on the OSF page for Clayards (2018). Beyond these comprehensive materials available on the OSF, here we provide details of a subset of relevant analyses.

### Reliability of Measures

To assess internal reliability of each of our measures, split-half reliability analyses with the Spearman-Brown correction were conducted. Each participant's trials were split randomly in half according to odd/even trial numbers, the measure of interest was calculated from each half, and the resulting corrected correlations between measures derived from each half are displayed in Table S.1. Reliability is generally acceptably high (> 0.70; Nunnally, 1978) across our measures, in particular for the oddity non-native perception, 2AFC native perception, and VAS consistency measures. Note that we find lower split-half reliability for our attention measure (bin scores from the AX-CPT task) than for other measures. This is to be expected given research that has shown reaction time-based measures to lack reliability in the context of individual difference studies (Draheim et al., 2019); however, we chose bin scores because they are more reliable than a reaction time measure alone since they also incorporate accuracy into their calculation. Also note that split-half reliability is low for the VAS slopes because splitting the data in half resulted in too few data points going into the rotated logistic calculation (the fits are much more variable when not enough data is provided; across experiments and contrasts, average standard deviation of the slopes calculated from the full datasets was 108, compared to an average standard deviation of 324 for slopes calculated from half datasets). However, the VAS rotated logistic slope measure has been shown to have a medium-large correlation across sessions (r = 0.45; E. Kapnoula, personal communication, June 22, 2023), similarly to the other gradiency measure derived by Kong & Edwards (2016). In addition, VAS rotated logistic slopes show good construct validity and reliability as a measure of gradiency because they relate to other measures of gradient speech processing including eye-tracking in a visual word paradigm task and neural activity as measured by EEG (Apfelbaum et al., 2022).

**Table S.1**

*Split-half reliability of each measure, with Spearman-Brown correction*

| Measure | Experiment 1 reliability | Experiment 2 reliability |
|---|---|---|
| Oddity A scores | 0.67 | 0.97 |
| VAS bet-bat slopes | 0.27 | 0.22 |
| VAS dear-tear slopes | 0.16 | 0.48 |
| VAS bet-bat consistency | 0.90 | 0.82 |
| VAS dear-tear consistency | 0.81 | 0.84 |
| 2AFC bet-bat step A coefficients | 0.92 | 0.87 |
| 2AFC dear-tear step A coefficients | 0.87 | 0.75 |
| 2AFC bet-bat step B coefficients | 0.85 | 0.68 |
| 2AFC dear-tear step B coefficients | 0.48 | 0.82 |
| AX-CPT bin scores | 0.62 | 0.56 |

**Pairwise Pearson Correlations Between Variables**

In order to explore relationships between our variables prior to running our primary analyses, we computed and visualized the pairwise Pearson correlations between the variables of interest for hypothesis 1 and hypothesis 2. These correlations are displayed in Figures S.1 and S.2 for Experiment 1 and Figures S.3 and S.4 for Experiment 2.

**Figure S.1**

*Experiment 1: Correlation plot showing Pearson correlation values between each pair of variables involved in Hypothesis 1. Circle size indicates correlation strength and circle colour indicates correlation strength and sign. Asterisks denote statistically significant correlations (*** = p < .001, ** = p < .01, * = p < .05).*

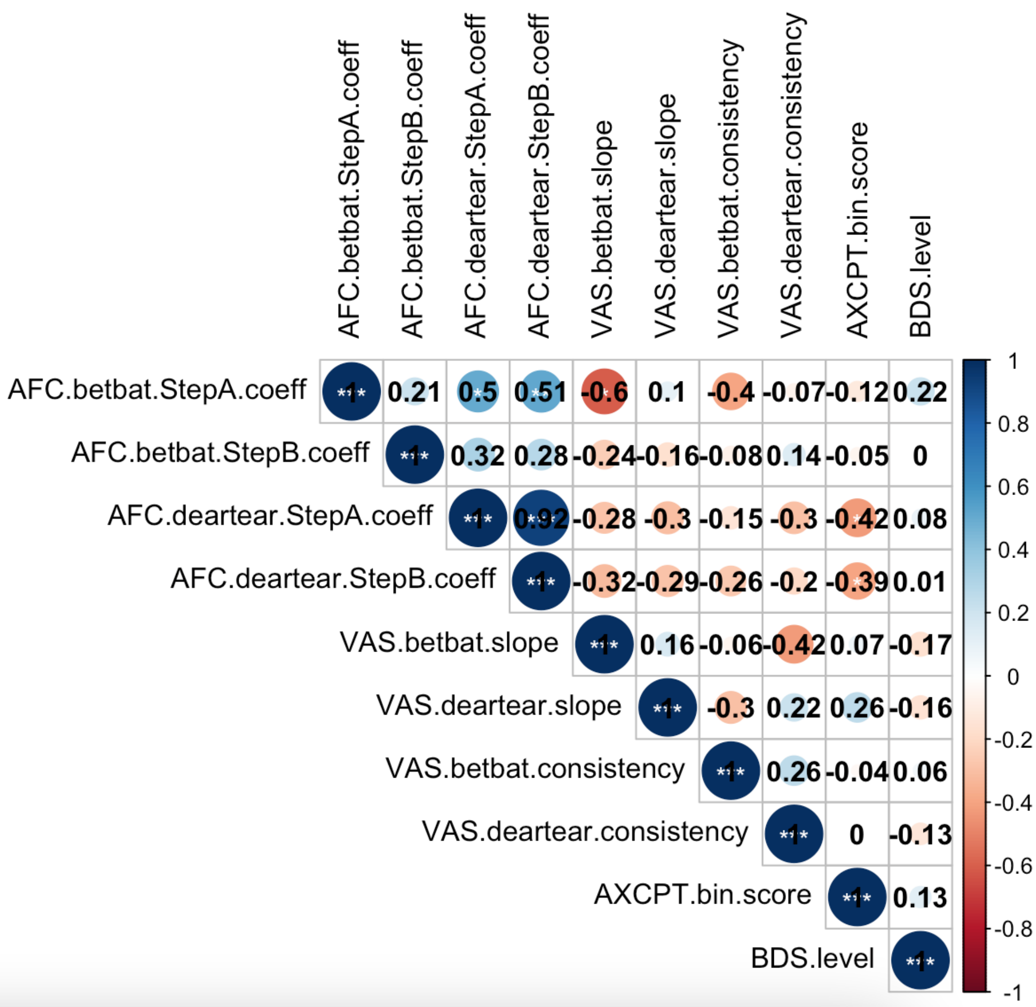


**Figure S.2**

*Experiment 1: Correlation plot showing Pearson correlation values between each pair of variables involved in Hypothesis 2. Circle size indicates correlation strength and circle colour indicates correlation strength and sign. Asterisks denote statistically significant correlations (*** = p < .001, ** = p < .01, * = p < .05).*

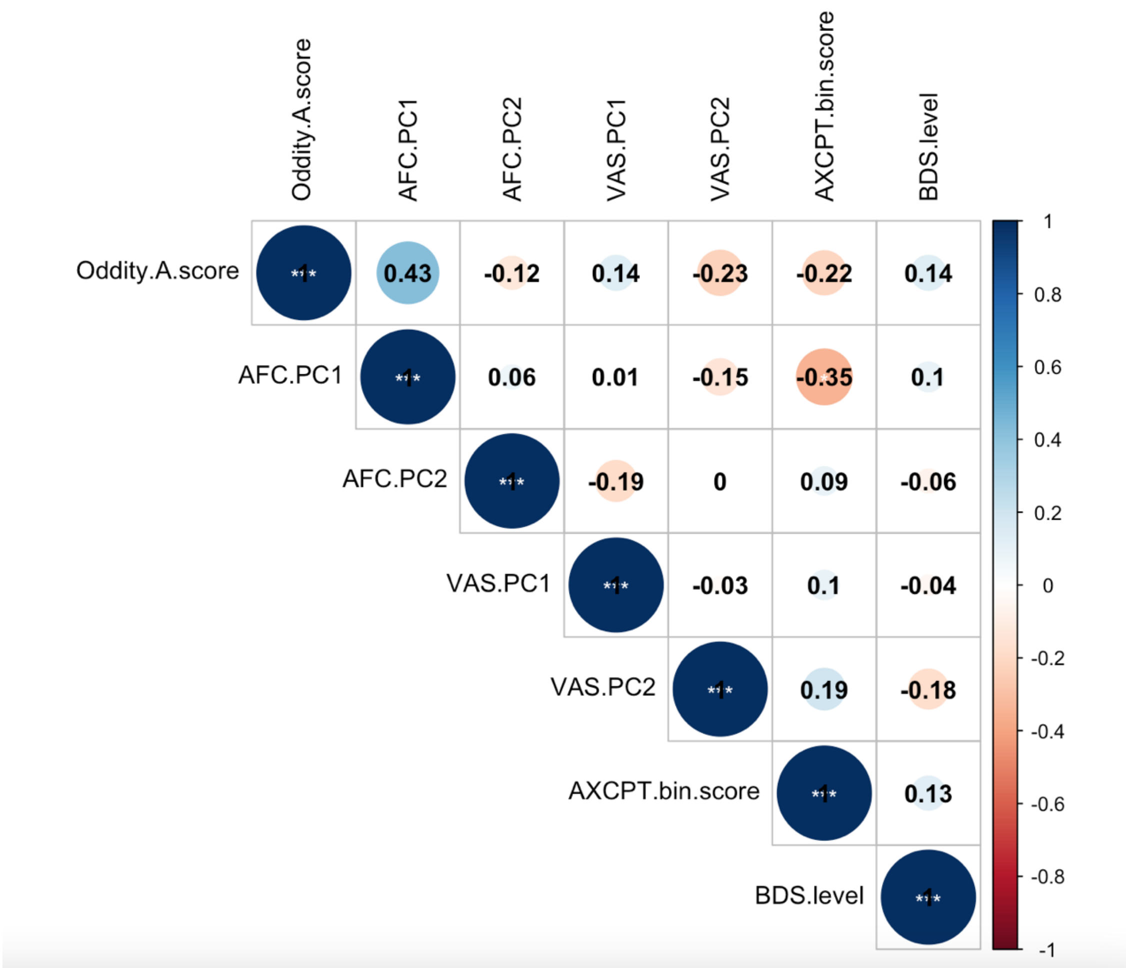


**Figure S.3**

*Experiment 2: Correlation plot showing Pearson correlation values between each pair of variables involved in Hypothesis 1. Circle size indicates correlation strength and circle colour indicates correlation strength and sign. Asterisks denote statistically significant correlations (*** = p < .001, ** = p < .01, * = p < .05).*

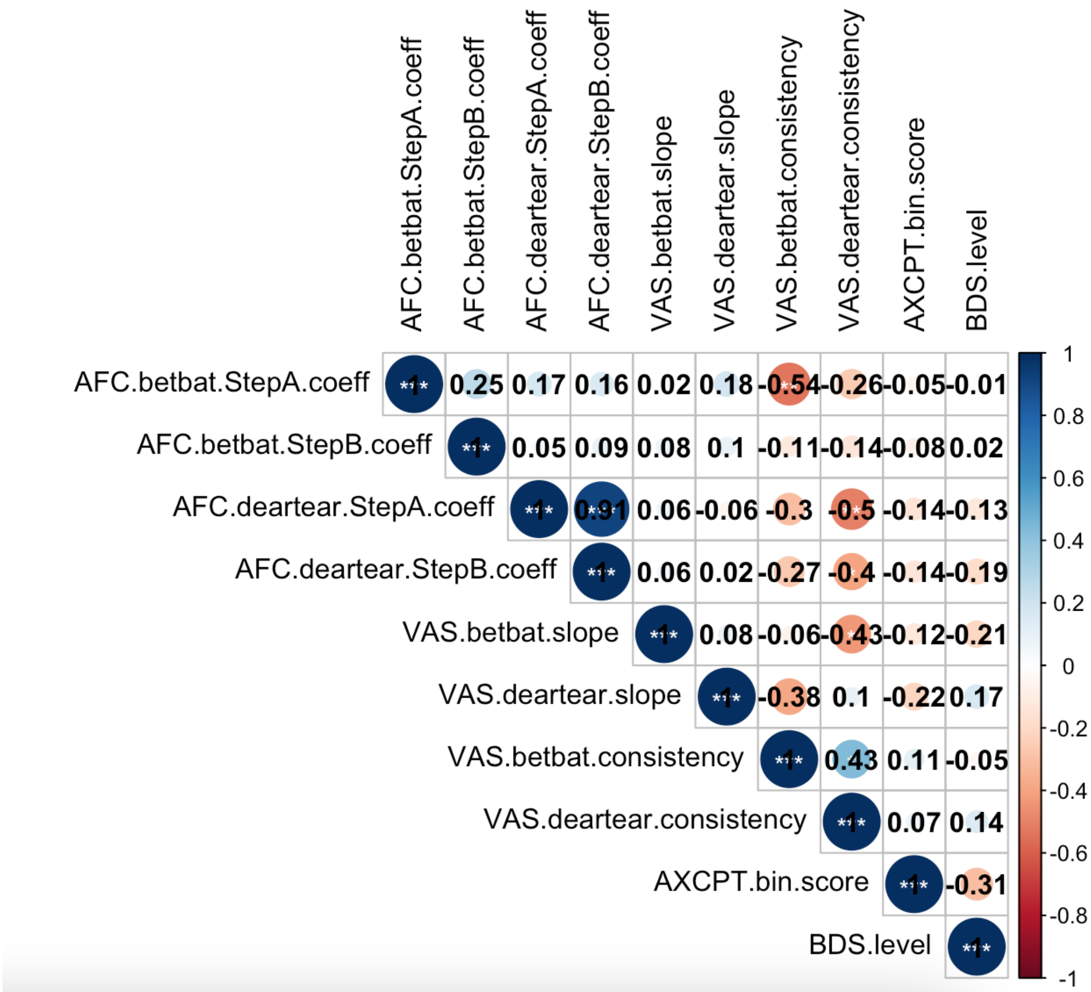
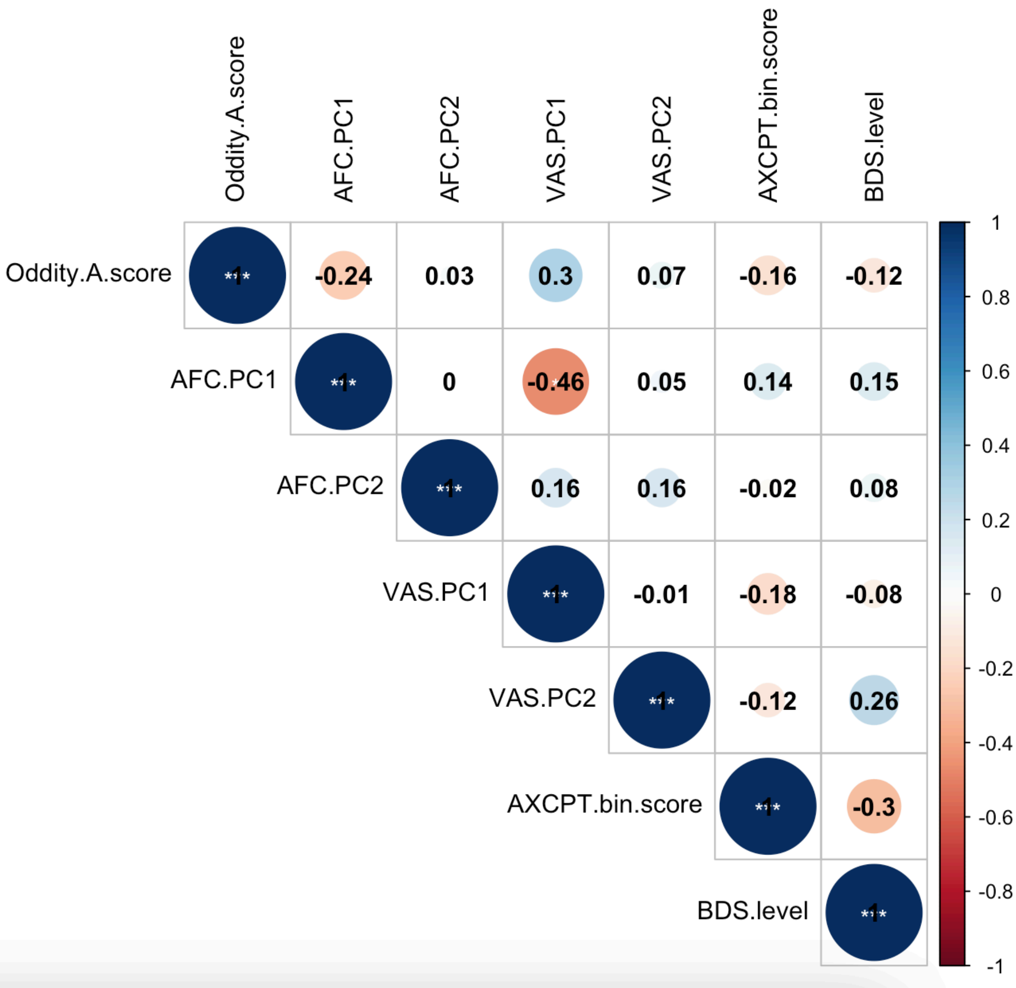


**Figure S.4**

*Experiment 2: Correlation plot showing Pearson correlation values between each pair of variables involved in Hypothesis 2. Circle size indicates correlation strength and circle colour indicates correlation strength and sign. Asterisks denote statistically significant correlations (*** = p < .001, ** = p < .01, * = p < .05).*



### Hypothesis 1 - Canonical Correlation

As reported in the *Hypothesis 1 - Canonical Correlation* section in the main article, canonical correlation analyses were run to test hypothesis 1. Table S.1 displays the canonical correlation coefficients derived from these analyses, for both experiments and both correlations run. In line with our hypothesis, there was a significant correlation between 2AFC coefficients and VAS consistency across both experiments. Table S.2 displays the canonical coefficients for both experiments and both correlations run, i.e., the loadings of the original variables onto the canonical dimensions.

**Table S.2**

*Canonical correlation coefficients and their significance, for both experiments and both correlations run*

| **Experiment 1, Correlation 1 (2AFC coefficients and VAS slopes)** | | | | | |
|---|---|---|---|---|---|
| Dimension | Canonical Cor. | *F* | *df*1 | *df*2 | *p* |
| 1 | 0.690 | 5.117 | 8 | 76 | < 0.001 |
| 2 | 0.439 | 3.108 | 3 | 39 | 0.037 |
| **Experiment 1, Correlation 2 (2AFC coefficients and VAS consistency)** | | | | | |
| Dimension | Canonical Cor. | *F* | *df*1 | *df*2 | *p* |
| 1 | 0.617 | 3.743 | 8 | 76 | < 0.001 |
| 2 | 0.410 | 2.629 | 3 | 39 | 0.064 |
| **Experiment 2, Correlation 1 (2AFC coefficients and VAS slopes)** | | | | | |
| Dimension | Canonical Cor. | *F* | *df*1 | *df*2 | *p* |
| 1 | 0.272 | 1.020 | 8 | 188 | 0.423 |
| 2 | 0.090 | 0.258 | 3 | 95 | 0.855 |
| **Experiment 2, Correlation 2 (2AFC coefficients and VAS consistency)** | | | | | |
| Dimension | Canonical Cor. | *F* | *df*1 | *df*2 | *p* |
| 1 | 0.663 | 10.700 | 8 | 188 | < 0.001 |
| 2 | 0.398 | 5.954 | 3 | 95 | < 0.001 |

**Table S.3**

*Canonical coefficients for each variable involved in the canonical correlation analyses, for Experiment 1 (left) and Experiment 2 (right)*

| **Experiment 1, Correlation 1 (2AFC coefficients and VAS slopes)** | | | **Experiment 2, Correlation 1 (2AFC coefficients and VAS slopes)** | | |
|---|---|---|---|---|---|
| | Canonical Dimension | | | Canonical Dimension | |
| | 1 | 2 | | 1 | 2 |
| 2AFC variables | | | 2AFC variables | | |
| bet-bat acoustic cue A | -1.115 | 0.324 | bet-bat acoustic cue A | 0.735 | -0.265 |
| bet-bat acoustic cue B | -0.120 | -0.335 | bet-bat acoustic cue B | 0.237 | 0.930 |
| dear-tear acoustic cue A | 0.477 | -0.686 | dear-tear acoustic cue A | -1.778 | 0.821 |
| dear-tear acoustic cue B | -0.090 | -0.321 | dear-tear acoustic cue B | 1.346 | -0.338 |
| VAS slope variables | | | VAS slope variables | | |
| bet-bat slope | 1.305 | 0.431 | bet-bat slope | -0.073 | 0.965 |
| dear-tear slope | -0.466 | 0.876 | dear-tear slope | 0.957 | -0.002 |
| **Experiment 1, Correlation 2 (2AFC coefficients and VAS consistency)** | | | **Experiment 2, Correlation 2 (2AFC coefficients and VAS consistency)** | | |
| | Canonical Dimension | | | Canonical Dimension | |
| | 1 | 2 | | 1 | 2 |
| 2AFC variables | | | 2AFC variables | | |
| bet-bat acoustic cue A | -0.738 | 0.516 | bet-bat acoustic cue A | -0.636 | 0.889 |
| bet-bat acoustic cue B | -0.356 | -0.429 | bet-bat acoustic cue B | 0.004 | -0.194 |
| dear-tear acoustic cue A | 1.999 | 0.936 | dear-tear acoustic cue A | -0.871 | -1.612 |
| dear-tear acoustic cue B | -1.517 | -0.268 | dear-tear acoustic cue B | 0.213 | 0.971 |
| VAS consistency variables | | | VAS consistency variables | | |
| bet-bat consistency | 1.022 | -0.390 | bet-bat consistency | 0.555 | -1.002 |
| dear-tear consistency | -0.694 | -0.781 | dear-tear consistency | 0.604 | 0.951 |

*Note*. These canonical coefficients are interpreted similarly to linear regression coefficients. For example, for Experiment 1 Correlation 1, a one-unit increase in bet-bat acoustic cue A corresponds to a decrease of -1.115 in the first canonical variate for the set of 2AFC variables, when holding other variables constant.

**Hypothesis 1 - Multivariate Multiple Regression**

As reported in the *Hypothesis 1 - Multivariate Multiple Regression* section in the main article, the canonical correlation analyses were followed up with multivariate multiple regression. Table S.3 displays the regression output from the four separate regression models fit within each multivariate model (one with each of the four 2AFC coefficients as the response) for Experiment 1, and Table S.4 displays the same regression output for Experiment 2. Model validation plots for both experiments are displayed in Figure S.1 (QQ-plots of residuals), Figure S.2 (plots of fitted values against residuals), and Figure S.3 (plots of Cook's distance per participant). The residuals appear to be relatively normally distributed based on the Q-Q plots and to have roughly constant variance based on the fitted-residual plots. While some participants have higher influence than others based on Cook's distance, none seem to be problematically influential (one common cut-off is Cook's distance >1, which all participants fall well below).

**Table S.4**

*Regression output for the four separate models fit as part of the multivariate multiple regression model for Experiment 1*

| **Response: 2AFC bet-bat acoustic cue A** | | | | |
|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
| (Intercept) | 0.136 | 0.100 | 1.363 | 0.182 |
| VAS bet-bat slope | -1.072 | 0.157 | -6.826 | < 0.001 |
| VAS dear-tear slope | 0.285 | 0.113 | 2.522 | 0.016 |
| VAS bet-bat consistency | -0.318 | 0.133 | -2.387 | 0.023 |
| VAS dear-tear consistency | -0.392 | 0.127 | -3.100 | 0.004 |
| AX-CPT bin score | -0.156 | 0.100 | -1.564 | 0.127 |
| Backwards digit span | 0.116 | 0.104 | 1.116 | 0.273 |
| *Multiple $R^2$ = 0.690, Adjusted $R^2$ = 0.633, Residual SE = 0.602 (df = 33), n = 40* | | | | |
| **Response: 2AFC bet-bat acoustic cue B** | | | | |
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
| (Intercept) | 0.030 | 0.177 | 0.172 | 0.865 |
| VAS bet-bat slope | -0.197 | 0.279 | -0.706 | 0.485 |
| VAS dear-tear slope | -0.253 | 0.200 | -1.263 | 0.216 |
| VAS bet-bat consistency | -0.263 | 0.237 | -1.113 | 0.274 |
| VAS dear-tear consistency | 0.197 | 0.225 | 0.878 | 0.387 |
| AX-CPT bin score | 0.019 | 0.177 | 0.110 | 0.913 |
| Backwards digit span | -0.032 | 0.185 | -0.172 | 0.864 |
| *Multiple $R^2$ = 0.117, Adjusted $R^2$ = -0.044, Residual SE = 1.069 (df = 33), n = 40* | | | | |
| **Response: 2AFC dear-tear acoustic cue A** | | | | |
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
| (Intercept) | 0.071 | 0.129 | 0.550 | 0.586 |
| VAS bet-bat slope | -0.579 | 0.204 | -2.830 | 0.008 |
| VAS dear-tear slope | -0.057 | 0.147 | -0.386 | 0.702 |
| VAS bet-bat consistency | -0.101 | 0.173 | -0.583 | 0.564 |
| VAS dear-tear consistency | -0.445 | 0.165 | -2.700 | 0.011 |
| AX-CPT bin score | -0.360 | 0.130 | -2.778 | 0.009 |
| Backwards digit span | -0.010 | 0.136 | -0.075 | 0.940 |
| *Multiple $R^2$ = 0.450, Adjusted $R^2$ = 0.350, Residual SE = 0.784 (df = 33), n = 40* | | | | |
| **Response: 2AFC dear-tear acoustic cue B** | | | | |
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
| (Intercept) | 0.041 | 0.131 | 0.316 | 0.754 |
| VAS bet-bat slope | -0.532 | 0.206 | -2.577 | 0.015 |
| VAS dear-tear slope | -0.154 | 0.148 | -1.043 | 0.304 |
| VAS bet-bat consistency | -0.316 | 0.175 | -1.805 | 0.080 |
| VAS dear-tear consistency | -0.272 | 0.166 | -1.634 | 0.112 |
| AX-CPT bin score | -0.303 | 0.131 | -2.315 | 0.027 |
| Backwards digit span | -0.066 | 0.137 | -0.484 | 0.632 |
| *Multiple $R^2$ = 0.431, Adjusted $R^2$ = 0.327, Residual SE = 0.790 (df = 33), n = 40* | | | | |

*Note.* Model equation: cbind(2AFC bet-bat acoustic cue A slope, 2AFC bet-bat acoustic cue B slope, 2AFC dear-tear acoustic cue A slope, 2AFC dear-tear acoustic cue B slope) ~ VAS bet-bat slope + VAS dear-tear slope + VAS bet-bat consistency + VAS dear-tear consistency + AX-CPT bin score + Backwards digit span.

**Table S.5**

*Regression output for the four separate models fit as part of the multivariate multiple regression model for Experiment 2*

| **Response: 2AFC bet-bat acoustic cue A** | | | | |
|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* |
| (Intercept) | 0.165 | 0.083 | 1.991 | 0.050 |
| VAS bet-bat slope | -0.032 | 0.091 | -0.349 | 0.728 |
| VAS dear-tear slope | -0.009 | 0.098 | -0.092 | 0.927 |
| VAS bet-bat consistency | -0.513 | 0.115 | -4.454 | < 0.001 |
| VAS dear-tear consistency | -0.040 | 0.121 | -0.333 | 0.740 |
| AX-CPT bin score | -0.006 | 0.093 | -0.067 | 0.947 |
| Backwards digit span | -0.035 | 0.097 | -0.359 | 0.720 |
| *Multiple $R^2$ = 0.291, Adjusted $R^2$ = 0.241, Residual SE = 0.788 (df = 86), n = 93* | | | | |
| **Response: 2AFC bet-bat acoustic cue B** | | | | |
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* |
| (Intercept) | 0.059 | 0.107 | 0.551 | 0.583 |
| VAS bet-bat slope | 0.005 | 0.117 | 0.041 | 0.968 |
| VAS dear-tear slope | 0.096 | 0.126 | 0.762 | 0.448 |
| VAS bet-bat consistency | -0.011 | 0.148 | -0.073 | 0.942 |
| VAS dear-tear consistency | -0.151 | 0.156 | -0.969 | 0.335 |
| AX-CPT bin score | -0.042 | 0.120 | -0.347 | 0.729 |
| Backwards digit span | 0.011 | 0.124 | 0.092 | 0.927 |
| *Multiple $R^2$ = 0.033, Adjusted $R^2$ = -0.034, Residual SE = 1.015 (df = 86), n = 93* | | | | |
| **Response: 2AFC dear-tear acoustic cue A** | | | | |
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* |
| (Intercept) | 0.035 | 0.084 | 0.422 | 0.674 |
| VAS bet-bat slope | -0.190 | 0.092 | -2.056 | 0.043 |
| VAS dear-tear slope | -0.033 | 0.099 | -0.328 | 0.744 |
| VAS bet-bat consistency | -0.094 | 0.117 | -0.803 | 0.424 |
| VAS dear-tear consistency | -0.529 | 0.123 | -4.312 | < 0.001 |
| AX-CPT bin score | -0.168 | 0.094 | -1.782 | 0.078 |
| Backwards digit span | -0.151 | 0.098 | -1.538 | 0.128 |
| *Multiple $R^2$ = 0.319, Adjusted $R^2$ = 0.271, Residual SE = 0.800 (df = 86), n = 93* | | | | |
| **Response: 2AFC dear-tear acoustic cue B** | | | | |
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* |
| (Intercept) | 0.058 | 0.086 | 0.676 | 0.501 |
| VAS bet-bat slope | -0.172 | 0.094 | -1.824 | 0.072 |
| VAS dear-tear slope | 0.046 | 0.102 | 0.449 | 0.654 |
| VAS bet-bat consistency | -0.073 | 0.119 | -0.610 | 0.543 |
| VAS dear-tear consistency | -0.407 | 0.125 | -3.248 | 0.002 |
| AX-CPT bin score | -0.175 | 0.096 | -1.817 | 0.073 |
| Backwards digit span | -0.233 | 0.100 | -2.328 | 0.022 |
| *Multiple $R^2$ = 0.243, Adjusted $R^2$ = 0.190, Residual SE = 0.818 (df = 86), n = 93* | | | | |

*Note.* Model equation: cbind(2AFC bet-bat acoustic cue A slope, 2AFC bet-bat acoustic cue B slope, 2AFC dear-tear acoustic cue A slope, 2AFC dear-tear acoustic cue B slope) ~ VAS bet-bat slope + VAS dear-tear slope + VAS bet-bat consistency + VAS dear-tear consistency + AX-CPT bin score + Backwards digit span.

**Figure S.5**

*Q-Q plots of residuals for the multivariate multiple regression model predicting 2AFC coefficients for each contrast and acoustic cue*

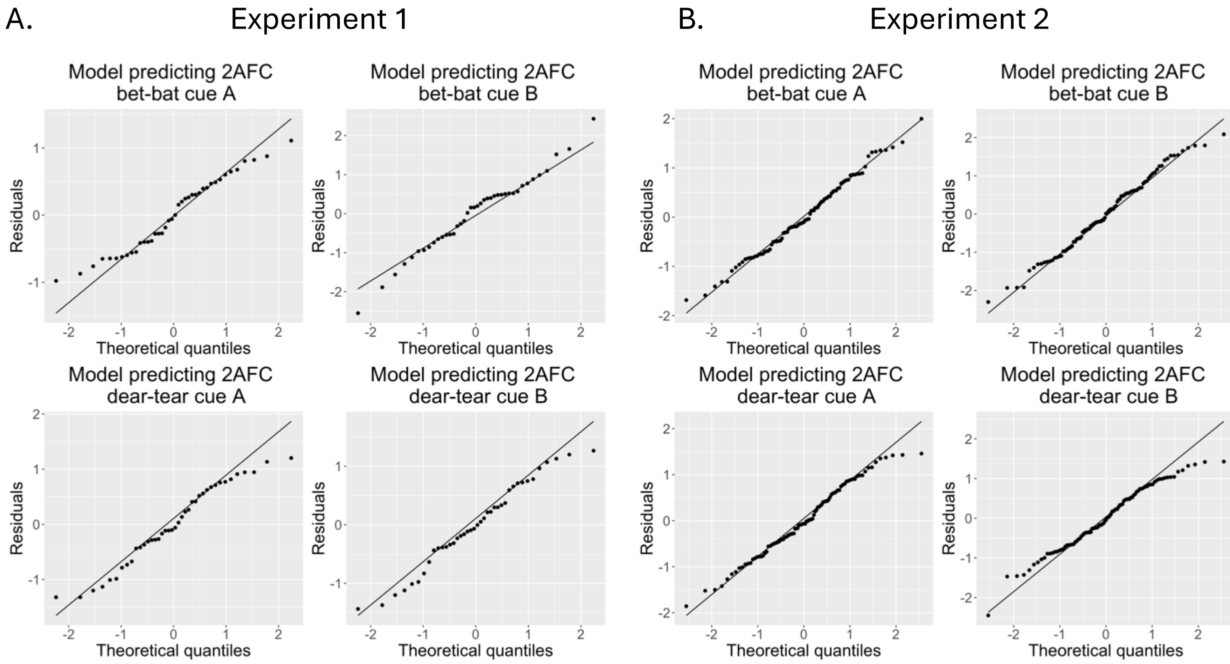



**Figure S.6**

*Fitted-residual plots for the multivariate multiple regression model predicting 2AFC coefficients for each contrast and acoustic cue*

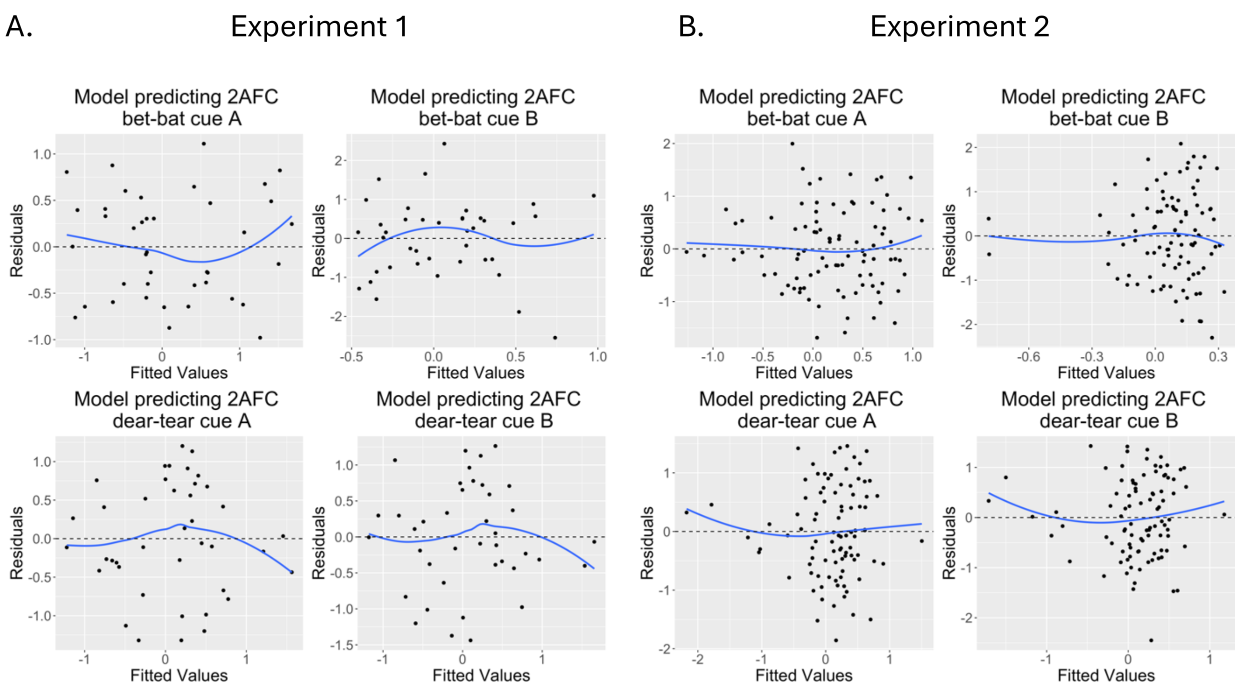
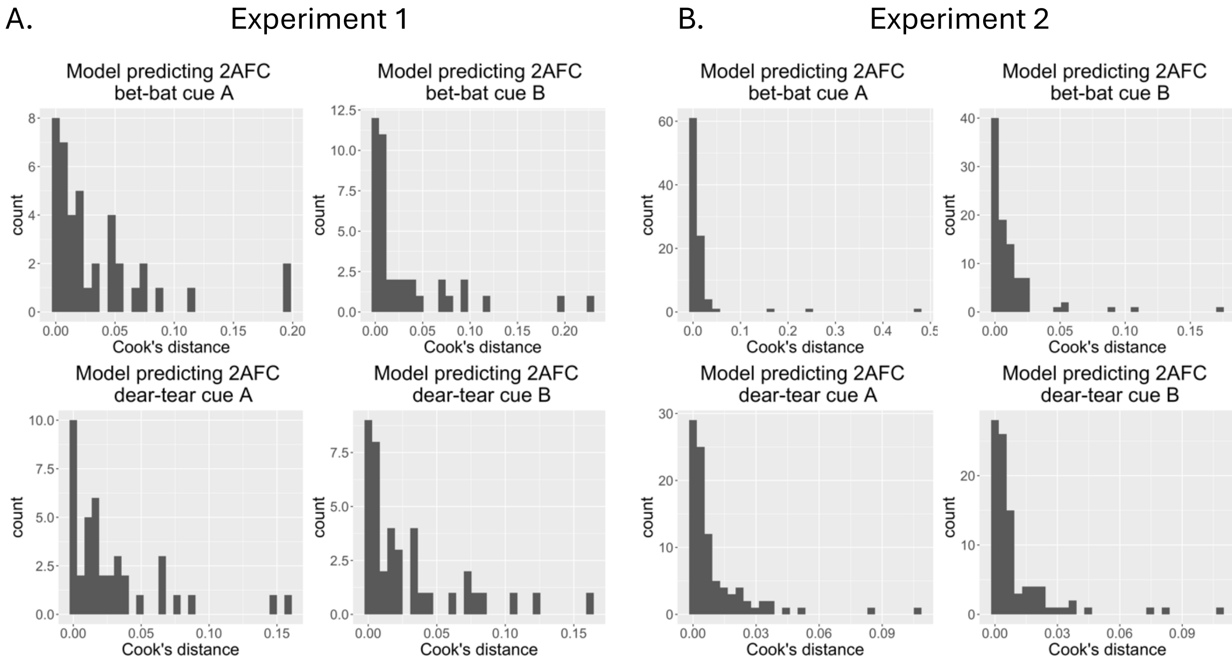


**Figure S.7**

*Plots of Cook's distance for the multivariate multiple regression model predicting 2AFC coefficients for each contrast and acoustic cue*




## Hypothesis 2 - Multiple Regression

As reported in the *Hypothesis 2 - Multiple Regression* section in the main article, hypothesis 2 was tested using multiple regression. Model validation plots for both experiments are displayed in Figure S.4 (QQ-plot of residuals), Figure S.5 (plot of fitted values against residuals), and Figure S.6 (plot of Cook's distance per participant). The residuals appear to be relatively normally distributed based on the Q-Q plots and to have roughly constant variance based on the fitted-residual plots. While some participants have higher influence than others based on Cook's distance, none seem to be problematically influential (one common cut-off is Cook's distance >1, which all participants fall well below).

Note that for Experiment 2, the model validation plots (S.4-S.6) are for the model excluding an influential participant. A summary of the original model including this participant is reported in Table S.5, and model validation plots for that model are displayed in Figure S.7. It is clear that one participant is particularly influential in this model (Cook's distance > 1.5).

**Figure S.8**

*Q-Q plot of residuals for the multiple regression model predicting oddity A scores*

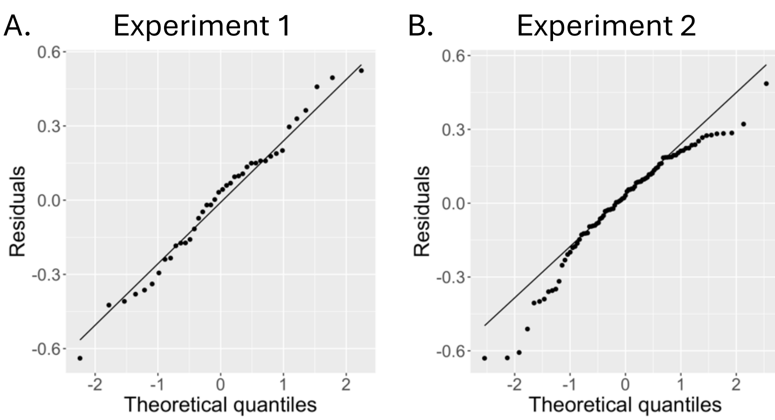



**Figure S.9**

*Fitted-residual plot for the multiple regression model predicting oddity A scores*

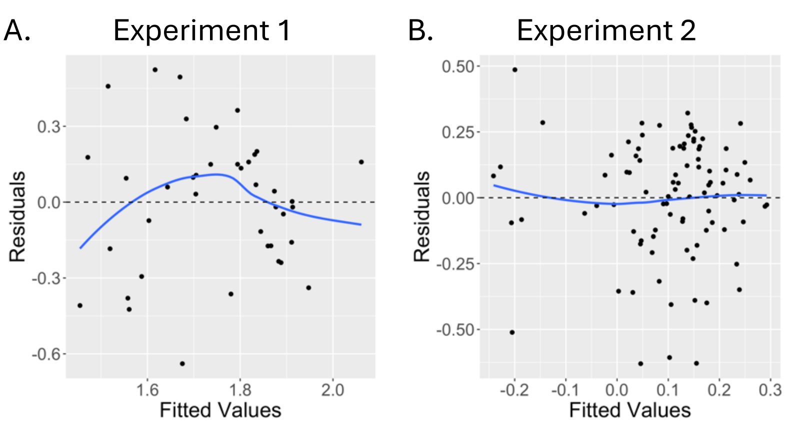


**Figure S.10**

*Plot of Cook's distance for the multiple regression model predicting oddity A scores*

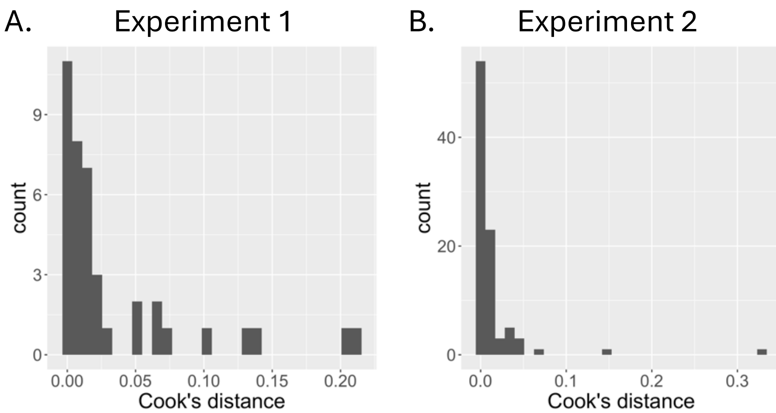



**Table S.6**

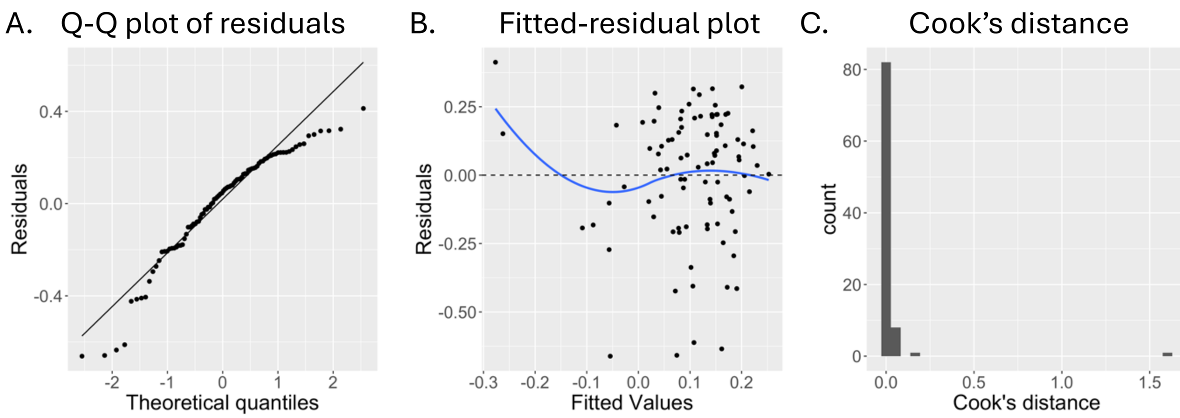
*Summary of the multiple regression model predicting oddity A scores for Experiment 2, before exclusion of an influential participant*

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 0.094 | 0.025 | 3.708 | < 0.001 |
| 2AFC principal comp. 1 | -0.020 | 0.022 | -0.939 | 0.350 |
| 2AFC principal comp. 2 | -0.003 | 0.024 | -0.120 | 0.905 |
| VAS principal comp. 1 | 0.042 | 0.023 | 1.799 | 0.076 |
| VAS principal comp. 2 | 0.023 | 0.024 | 0.959 | 0.340 |
| AX-CPT bin score | -0.037 | 0.029 | -1.309 | 0.194 |
| Backwards digit span | -0.041 | 0.030 | -1.391 | 0.168 |
| *Multiple $R^2$ = 0.136, Adjusted $R^2$ = 0.075, Residual SE = 0.232 (df = 85), n = 92* | | | | |

*Note.* Model equation: Oddity A score ~ 2AFC principal component 1 + 2AFC principal component 2 + VAS principal component 1 + VAS principal component 2 + AX-CPT bin score + Backwards digit span.

**Figure S.11**

*Model validation plots for the multiple regression model predicting oddity A scores, prior to excluding an influential participant*




## Bayesian Analyses

For both Hypothesis 1 and Hypothesis 2, exploratory Bayesian regression models were fitted with the same structure as the frequentist models described above and in the main text. These analyses are not reported in more detail here because, as pointed out by a helpful anonymous reviewer with expertise in Bayesian analyses, the informativeness of our approach was limited by the use of flat priors. However, details of these analyses can be found in the Supplemental R Markdown document on the OSF for the sake of transparency, and the analyses supported the qualitative conclusions drawn from our preregistered frequentist analyses.

## Dimensionality Reduction

As reported in the *Dimensionality Reduction of 2AFC and VAS Variables* section in the main article, all 12 slope and consistency variables from the native phonetic perception tasks were put into a PCA analysis. The correlations between the original variables and the first five principal components derived from them are reported in Table S.7. Biplots are shown in Figure S.12.

**Table S.7**

*Correlations between the 12 native perception variables and the first five principal components extracted from them*

**Experiment 1**

| | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| 2AFC bet-bat acoustic cue A coefficient | 0.391 | 0.326 | -0.017 | 0.061 | 0.024 |
| 2AFC bet-bat acoustic cue B coefficient | 0.169 | -0.051 | 0.273 | -0.372 | -0.658 |
| 2AFC dear-tear acoustic cue A coefficient | 0.395 | -0.289 | -0.054 | -0.067 | -0.051 |
| 2AFC dear-tear acoustic cue B coefficient | 0.392 | -0.245 | -0.057 | 0.082 | -0.095 |
| 2AFC bet-bat rotated logistic slope | -0.217 | -0.372 | -0.075 | 0.241 | -0.526 |
| 2AFC dear-tear rotated logistic slope | -0.102 | 0.143 | -0.048 | -0.814 | -0.025 |
| 2AFC bet-bat consistency | -0.394 | -0.328 | 0.021 | 0.062 | -0.056 |
| 2AFC dear-tear consistency | -0.385 | 0.335 | -0.023 | -0.043 | -0.074 |
| VAS bet-bat slope | -0.285 | -0.196 | -0.477 | -0.151 | -0.025 |
| VAS dear-tear slope | -0.123 | 0.460 | -0.192 | 0.235 | -0.245 |
| VAS bet-bat consistency | -0.176 | -0.293 | 0.477 | -0.117 | 0.414 |
| VAS dear-tear consistency | -0.144 | 0.182 | 0.646 | 0.169 | -0.194 |
| *Percent variance explained* | 38% | 17% | 13% | 9% | 8% |

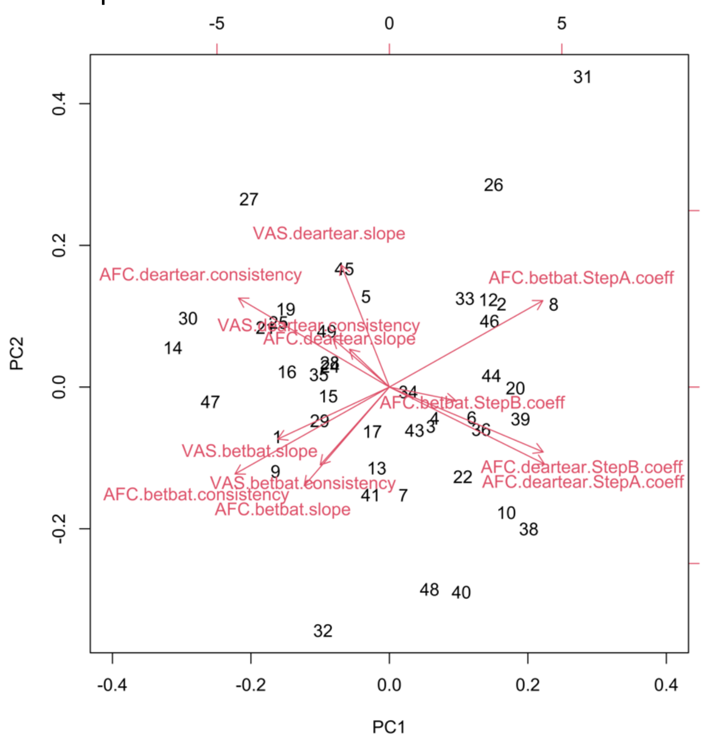
**Experiment 2**

| | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| 2AFC bet-bat acoustic cue A coefficient | -0.324 | -0.414 | -0.092 | 0.022 | -0.116 |
| 2AFC bet-bat acoustic cue B coefficient | -0.096 | -0.292 | 0.149 | -0.385 | 0.507 |
| 2AFC dear-tear acoustic cue A coefficient | -0.413 | 0.296 | -0.089 | -0.026 | 0.210 |
| 2AFC dear-tear acoustic cue B coefficient | -0.396 | 0.250 | -0.090 | -0.008 | 0.263 |
| 2AFC bet-bat rotated logistic slope | 0.130 | 0.326 | 0.363 | 0.068 | 0.448 |
| 2AFC dear-tear rotated logistic slope | 0.085 | -0.316 | -0.004 | -0.488 | 0.348 |
| 2AFC bet-bat consistency | 0.366 | 0.374 | 0.136 | -0.039 | 0.113 |
| 2AFC dear-tear consistency | 0.395 | -0.340 | 0.072 | -0.026 | -0.042 |
| VAS bet-bat slope | -0.066 | 0.058 | 0.715 | -0.296 | -0.399 |
| VAS dear-tear slope | -0.033 | -0.261 | 0.437 | 0.604 | 0.267 |
| VAS bet-bat consistency | 0.331 | 0.224 | -0.197 | -0.268 | -0.022 |
| VAS dear-tear consistency | 0.359 | -0.104 | -0.240 | 0.286 | 0.220 |
| *Percent variance explained* | 34% | 19% | 10% | 9% | 8% |

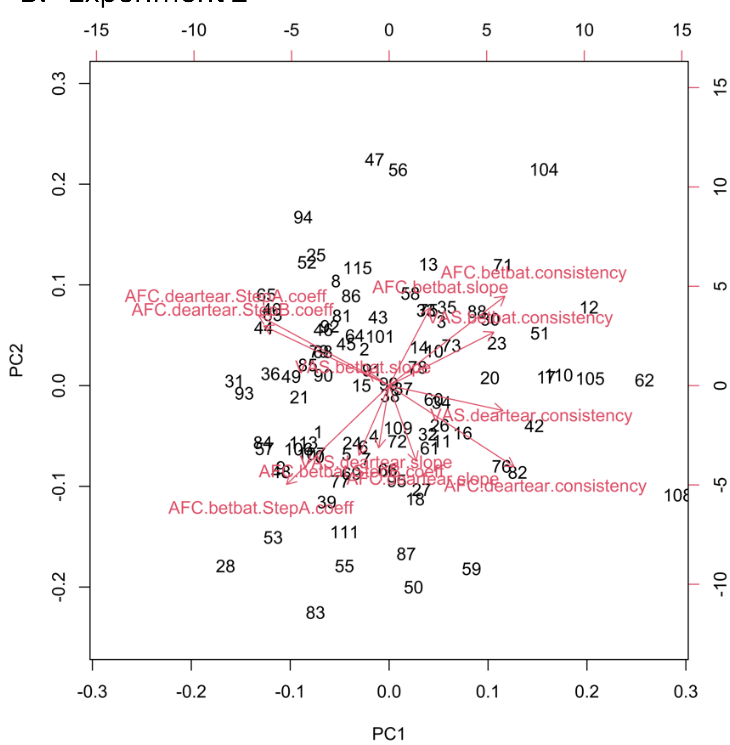
**Figure S.12**

*Biplots of the first and second principal components derived from all native phonetic perception measures, for both experiments*

A. Experiment 1




B. Experiment 2



**Supplemental References**

Apfelbaum, K. S., Kutlu, E., McMurray, B., & Kapnoula, E. C. (2022). Don't force it! Gradient speech categorization calls for continuous categorization tasks. *The Journal of the Acoustical Society of America*, *152*(6), 3728-3745. https://doi.org/10.1121/10.0015201

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *J. Acoust. Soc. Am.*, *144*(3), EL172-EL177. https://doi.org/10.1121/1.5052025.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological bulletin*, *145*(5), 508. https://doi.org/10.1037/bul0000192.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40-57. https://doi.org/10.1016/j.wocn.2016.08.006.

Nunnally, J. C. (1978). Psychometric theory (2d ed., Ser. McGraw-Hill series in psychology). McGraw-Hill.

## Link Between Chapters 1 and 2

The goal of Chapter 1 was to characterize the nature of individual differences in native and non-native phonetic perception on a behavioural level, and to determine the relationships among these differences. Until now, it had remained unclear whether listeners' performance on common native phonetic perception tasks—specifically, 2AFC and VAS tasks—reflected identical or distinct constructs, and whether differences in native phonetic perception predicted differences in non-native phonetic perception. The first chapter established that the native and non-native phonetic perception measures under study all appear to be interrelated through behavioural response consistency—that is, the similarity of listeners' behavioural responses across repeated presentations of a given stimulus. In particular, listeners with more consistent responses to native phonemes on the VAS task tended to have more consistent responses and steeper identification slopes on the 2AFC task. More consistent VAS responses also predicted better identification of unfamiliar non-native phonemes. The use of behavioural measures (e.g., sensitivity to non-native sound contrasts, identification functions for continua of native sounds, and correspondence between these identification functions and listeners' actual responses) enabled a first glimpse into the cognitive processes underlying the perception of native and non-native speech sounds.

With the importance of behavioural response consistency established, Chapter 2 turns to the question of how such consistency might originate. In particular, we aimed to determine whether the considerable individual differences in response consistency observed in Chapter 1 might relate to differences in the consistency of neural responses to sound. The same behavioural measures as in Chapter 1 were employed in order to reinforce our conclusions about relationships between performance on phonetic perception tasks; however, the additional inclusion of a neural activity measure in Chapter 2 enabled us to directly explore relationships between brain activity and behaviour. The neural measure in question was the Frequency Following Response (FFR), an auditory evoked potential recorded using electroencephalography (EEG). As outlined in detail in Chapter 2, the FFR is a particularly interesting measure of neural sound encoding for our purposes because (i) it shows robust individual differences across healthy young adults and (ii) it can quantify the similarity of a listener's neural responses across repeated presentations of a given stimulus (i.e., neural response consistency). By investigating individual differences in speech perception at both neural and behavioural levels, we can provide a richer picture of these differences while exploring how they may arise.

# Chapter 2: Individual Differences in the Consistency of Neural and Behavioural Responses to Speech Sounds


## Abstract

There are documented individual differences among adults in the consistency of speech sound processing, both at neural and behavioural levels. Some adults show more consistent neural responses to speech sounds than others, as measured by an event-related potential called the frequency-following response (FFR); similarly, some adults show more consistent behavioural responses to native speech sounds than others, as measured by two-alternative forced choice (2AFC) and visual analog scaling (VAS) tasks. Adults also differ in how successfully they can perceive non-native speech sounds. Interestingly, it remains unclear whether these differences are related within individuals. In the current study, native English-speaking adults completed native phonetic perception tasks (2AFC and VAS), a non-native (German) phonetic perception task, and an FFR recording session. From these tasks, we derived measures of the consistency of participants' neural and behavioural responses to native speech as well as their non-native perception ability. We then examined the relationships among individual differences in these measures. Analysis of the behavioural measures revealed that more consistent responses to native sounds predicted more successful perception of unfamiliar German sounds. Analysis of neural and behavioural data did not reveal clear relationships between FFR consistency and our phonetic perception measures. This multimodal work furthers our understanding of individual differences in speech processing among adults, and may

eventually lead to individualized approaches for enhancing non-native language acquisition in adulthood.

**Keywords**: frequency following response, phonetic perception, non-native perception, individual differences, consistency

## Introduction

Adults differ in their perception of speech sounds. These individual differences provide insight into nuances of speech perception that would otherwise be obscured at the group level, and they are of theoretical interest because their sources and functions have yet to be well specified. In relating these individual differences to other factors, such as auditory processing (e.g., Won et al., 2016) and executive function (e.g., Kapnoula et al., 2017), we can better understand the broader structure and functioning of speech perception as well as the mechanisms supporting it. Beyond holding theoretical value, research on variability in speech perception could lead to tailored interventions that would benefit clinical and non-clinical populations alike. Such interventions could include customized language training for adults acquiring a new language, or personalized diagnosis and treatment for speech and hearing disorders such as dyslexia, aphasia, and age-related hearing loss.

Strikingly, even in their native language, healthy young adults show differences in how they respond to phonemes—the fundamental units that make up spoken language. For instance, one commonly used experimental paradigm is the two-alternative forced choice (2AFC) task; in the context of phonetic perception, participants might be presented with a continuum of sounds (e.g., gradually ranging from *bet* to *bat*) and asked to label each sound as belonging to one of two categories (either *bet* or *bat*). In another paradigm called a visual analog scaling (VAS) task, participants might be presented with the same continuum of sounds, but instead of choosing between two response options, they would select a point along a continuous line (e.g., a slider with 100 possible responses ranging between *bet* and *bat*) to indicate their perception of each stimulus. In both cases, by plotting a participant's responses to each stimulus against the actual changes in the acoustic properties of the stimulus, an identification slope is obtained. Participants differ greatly in their identification slopes, with some showing shallower slopes (indicating responses that change more gradually as the stimulus changes) while others show steeper slopes (indicating responses that change more suddenly as the stimulus changes). They also differ in the consistency of their responses, with some participants showing very similar responses across repeated presentations of the same stimulus and others showing variable responses. The reasons for these differences are not well understood. By investigating how these differences relate to each other and what their potential sources might be, we can clarify what these common speech perception tasks are measuring and better understand the nature of speech perception.

Interestingly, while identification slopes on 2AFC and VAS tasks have sometimes been considered interchangeable measures of the same construct (e.g., Ou & Yu, 2022), it has been demonstrated that participants' slopes are not related across the two tasks (Honda et al., 2024; Kapnoula et al., 2017). By way of illustration, a participant could show a steep 2AFC slope and a shallow VAS slope even though the same stimuli were presented in both tasks. This outcome is possible because the two measures in fact appear to be tapping into distinct constructs: while VAS slopes seem to reflect the gradiency of phonetic perception (i.e., the ability to distinguish gradual, fine-tuned phonetic differences in speech sounds), 2AFC slopes seem to reflect the consistency of phonetic perception (i.e., the ability to reliably give the same response when presented with the same stimulus; Honda et al., 2024; Kapnoula et al., 2017). This conclusion is supported by the finding that the steepness of participants' 2AFC slopes is predicted by the consistency of their VAS responses, with more consistent responders showing steeper slopes (Honda et al., 2024; Kapnoula et al., 2017). Studies have also shown that participants who are illiterate (Serniclaes et al., 2005) or who have a language impairment such as developmental dyslexia (Joanisse et al., 2000; Serniclaes et al., 2001) tend to have shallower 2AFC slopes, likely due to inconsistent response patterns. Such findings provide further evidence that 2AFC slopes appear to measure consistency. It therefore seems that even though slopes from 2AFC and VAS tasks reflect distinct concepts (consistency and gradiency, respectively), performance on the two tasks is linked through the underlying construct of consistency, with greater consistency promoting steeper 2AFC slopes and more similar VAS responses to a given stimulus across trials.

In addition to differences in native speech perception, there are well-documented differences between adults in the perception of non-native speech sounds. For example, there is wide variability in (i) how accurately listeners identify or discriminate non-native sounds (Hattori & Iverson, 2010; Lengeris & Hazan, 2010), (ii) how they assimilate non-native sounds to native categories (Mayr & Escudero, 2010), and (iii) which acoustic cues they rely on when categorizing non-native sounds (Schertz et al., 2015). While some adults are highly successful, many others have difficulty distinguishing non-native speech sounds even after receiving training and feedback (Bradlow et al., 1997; Hanulíková et al., 2012; Strange & Dittman, 1984). Although some factors have been found to account in part for these differences, such as native language background (Flege et al., 1997) and musical ability (Slevc & Miyake, 2006), predictors of successful non-native perception at the individual level remain unclear.

An interesting question that arises is whether differences in non-native perception are linked to the aforementioned differences in native perception. The latest research suggests that native speech perception measures (e.g., those derived from 2AFC and VAS tasks) do indeed predict non-native perception outcomes. One study examined the relationship between perception of native speech sounds on a VAS task, and non-native discrimination ability (Fuhrmeister et al., 2023). The authors found that more consistent VAS responses to native sounds were associated with more accurate discrimination of non-native sounds (Fuhrmeister et al., 2023). Preliminary work by another group has similarly reported a positive relationship between the consistency of native perception on a VAS task and the accuracy of non-native identification (Kapnoula & Samuel, 2023). Finally, an association between consistent VAS responses to native sounds and successful non-native perception was again reported by Honda et al. (2024). These emerging results suggest that further research would be valuable in confirming whether response consistency may be a key factor linking native and non-native perception.

While the consistency of behavioural responses seems to underlie differences in performance both across native perception tasks and across native and non-native perception tasks, the sources of this consistency remain unclear. Fuhrmeister and Myers (2021) studied the structural neural correlates of consistency, finding that less gyrification in the bilateral transverse temporal gyri was associated with more consistent responses on a native VAS task. This finding establishes that differences in behavioural consistency may arise in part from differences in neural anatomy, but whether and how differences in neural activity also contribute remains to be elucidated. Just as behavioural response consistency can be quantified as the similarity of *behavioural responses* across repeated presentations of a stimulus, neural response consistency can be quantified as the similarity of *neural responses* (e.g., brainwaves as measured by electroencephalography; EEG) across repeated stimulus presentations. Recent studies with children have linked inconsistent or atypical neural representations of speech with dyslexia (Destoky et al., 2020; Keshavarzi et al., 2022; Power et al., 2016) and with a dyslexia risk gene (Neef et al., 2017). Given that dyslexia generally involves deficits in phonological awareness (the ability to recognize, identify, or manipulate any phonological unit within a word, such as a phoneme or syllable; Ziegler & Goswami, 2005), these results hint at the potential importance of consistent neural activity in speech processing. Nonetheless, it remains to be seen whether the consistency of neural activity predicts the consistency of behavioural responses to speech in healthy adult populations. In the most closely related investigation so far, Ou & Yu (2022) examined the relationship between adults' 2AFC slopes and measures of their subcortical and cortical auditory encoding. They found that steeper 2AFC slopes were associated with more faithful subcortical speech encoding—that is, with EEG peak latencies that more closely resembled the actual voice onset times of the eliciting stimuli (Ou & Yu, 2022). Their measure of subcortical encoding was about the fidelity rather than the consistency of neural activity, but it is possible that the two concepts are related, in which case these results are compatible with the idea that steep 2AFC slopes reflect consistency in perception. Nevertheless, the authors did not examine a direct index of the consistency of neural responses, and in fact they did not treat their behavioural outcome as a measure of consistency either—instead, they referred to 2AFC slopes as a measure of gradiency (but see Honda et al., 2024). More work is thus needed to specifically examine potential links between neural response consistency and behavioural response consistency.

In the current study, we set out to investigate a measure of neural activity that could be at the basis of behavioural response consistency: the frequency following response (FFR). The FFR is an auditory evoked potential that is generated by the auditory system and is phase-locked to individual cycles of a periodic auditory stimulus (Skoe & Kraus, 2010). It is recorded while participants listen to repeated presentations of a stimulus such as the speech syllable “da”, and it retains many waveform properties of the stimulus, including the fundamental frequency (F0), first formant (F1), and formant transitions (Krishnan, 2002; Plyler & Krishnan, 2001). Because the FFR retains properties of the stimulus that elicited it, it can even be heard as intelligible speech when converted from a neural signal to an audio signal (Galbraith et al., 1995). Given its phase-locked nature, it might at first glance seem that the FFR would only relate to the processing of a stimulus’ frequencies (especially of the F0). However, it is in fact a rich signal from which many measures can be derived (see

Krizman & Kraus, 2019 for a comprehensive overview of FFR measures and analyses). Timing-based measures (e.g., peak latencies, pitch tracking, and phase-locking of the FFR to stimulus frequencies) reflect the neural timing precision of the auditory system; magnitude-based measures (e.g., FFR amplitude at the stimulus F0, F1, and harmonics) reflect the strength of auditory frequency encoding; and fidelity-based measures (e.g., response consistency and stimulus-to-response correlation) reflect the stability and synchrony of stimulus-evoked neural firing (Krizman & Kraus, 2019). These measures have been found to relate not only to pitch processing (e.g., Krizman et al., 2015) but also to a variety of linguistic and experiential variables that are not strictly frequency-related, including consonant perception (Omote et al., 2017), reading ability (Hornickel & Kraus, 2013), and maternal education level (Skoe et al., 2013). The FFR is thus an informative and multifaceted response that has been associated with a wide range of factors, as elaborated upon further below.

For a variety of theoretical and practical reasons, it is interesting and relevant to study the FFR as a potential source of individual differences in phonetic perception. First, the FFR reflects the robustness of sound encoding: the more consistent it is and the better it represents the frequencies present in the eliciting stimulus, the more robustly it is encoding sound (Kraus et al., 2017; Krizman & Kraus, 2019). In addition, the FFR has recently been revealed to have multiple subcortical and cortical sources, making it an aggregate measure reflecting the response of the entire auditory system (Coffey et al., 2019). Accordingly, it is shaped by experiential factors that affect auditory system activity (e.g., short-term auditory training or long-term language and musical experience; Coffey et al., 2019). Furthermore, the FFR is a robust measure of auditory processing due to its high test-retest reliability (Hornickel et al., 2012; Song et al., 2011a). While other auditory evoked potentials are less reliable (Bidelman et al., 2018) and are therefore typically averaged across many participants in order to arrive at a sufficiently clear neural response, the FFR is reliable enough to be studied at the individual level. Indeed, there are group- and individual-level differences in the robustness of sound encoding as measured by the FFR (e.g., Chandrasekaran & Kraus, 2010; Coffey et al., 2016a). Importantly for the investigation of consistency, it is possible to derive a measure of neural response consistency from the FFR, which quantifies how similar an individual's neural response to a stimulus is over time (Krizman & Kraus, 2019). To our knowledge, such a measure of consistency has not been established for other auditory evoked potentials. In sum, the FFR is a uniquely fitting candidate for identifying neural sources of individual variability in the consistency of phonetic perception—it is reliable, it shows robust individual differences, and it provides a measure of neural response consistency.

At the group level, studies have revealed that certain populations have enhanced neural encoding of sound relative to others, as reflected by the FFR. For instance, representation of stimulus frequencies in the FFR has been found to be more robust for musicians compared to non-musicians (Bidelman & Krishnan, 2010; Musacchia et al., 2007; Parbery-Clark et al., 2009), for tone language speakers compared to non-tone language speakers (Krishnan et al., 2009; Llanos et al., 2017), for bilinguals compared to monolinguals (Krizman et al., 2012; Skoe et al., 2017), and even for simultaneous bilinguals compared to sequential bilinguals (Krizman et al., 2015). In terms of neural response consistency, more consistent FFRs have been found for children with good reading skills than for children with poor reading skills (Hornickel & Kraus, 2013), for adolescents with higher maternal education backgrounds than for those with lower maternal education backgrounds (Skoe et al., 2013), for musicians than for non-musicians (Parbery-Clark et al., 2013; Skoe & Kraus, 2013), and for bilinguals than for monolinguals (Krizman et al., 2014). In addition to these group-level differences, correlations have suggested individual differences in the FFR based on experience. Years of musical practice are positively correlated with the strength of phase-locking, pitch tracking, and representation of stimulus frequencies in the FFR (Bones et al., 2014; Musacchia et al., 2007; Wong et al., 2007). Similarly, years of bilingual experience correlate positively with the strength of F0 representation in the FFR (Krizman et al., 2015). It therefore appears that enriched auditory experience, whether linguistic or musical in nature, is associated with enhanced neural sound encoding.

Even in studies that attempt to control for some of these experiential factors (e.g., by including only participants with minimal musical training or second language experience), individual differences in the FFR have been found. These differences appear to be behaviourally relevant because they relate to performance on a variety of auditory perception and native speech perception tasks. For example, strength of F0 representation in the FFR has been positively correlated with speech-in-noise perception (Parbery-Clark et al., 2011; Song et al., 2011b), perceptual bias (tendency to perceive the missing fundamental; Coffey et al., 2016a), auditory spatial selective attention (Ruggles et al., 2012), and pitch discrimination (Krishnan et al., 2012; Krishnan et al., 2010; Marmel et al., 2013). In a longitudinal study, improvements and declines in children's speech-in-noise perception were associated with F0-related FFR enhancements and degradations, respectively (Thompson et al., 2017). Furthermore, individual differences in the encoding of vowel formant information in the FFR have been shown to relate to performance on a vowel identification task (Won et al., 2016). In a similar vein, it has been found that differences in the relative encoding strength of formant versus duration cues in the FFR are predictive of differences in the relative behavioural weighting of those same cues on a 2AFC task (Ou et al., 2023). Such individual differences in the FFR are found even among healthy participants with normal hearing thresholds, and may be due to differences in the number of auditory nerve fibres encoding sound (Shinn-Cunningham et al., 2017). Together, these findings support the notion that the FFR feeds into and reflects higher-level auditory perception.

A handful of studies have examined the FFR in relation to non-native speech perception. Studies training non-native lexical tone perception suggest that increased exposure to non-native tones is associated with increased FFR robustness. For example, F0 tracking in the FFR has been shown to improve after nine sessions of Mandarin lexical tone discrimination training (Chandrasekaran et al., 2012). Another training study reported that the FFRs of native English speakers showed improved F0 tracking for Mandarin tones, but only once participants had been extensively trained to reach native-like levels of behavioural discrimination for the tones (Reetzke et al., 2018). Accordingly, neural sound encoding as measured by the FFR might sometimes predict non-native perceptual success; however, the timescale of this relationship and its applicability beyond lexical tone perception are unclear, and it remains to be seen whether the FFR might predict the consistency of native perception. Other work with the FFR found that the strength of F0 phase-locking to non-native syllables was positively correlated with successful non-native consonant (but surprisingly, not vowel) perception (Omote et al., 2017). Follow-up studies subsequently found that more robust FFR phase-locking to the F1 was linked to better non-native pronunciation during spontaneous speech (Saito et al., 2018), and that more robust FFR phase-locking to the F2 was linked to more accurate non-native vowel perception (Kachlicka et al., 2019). These results hint at the potential for the FFR to predict non-native speech perception outcomes; however, note that different FFR measures were found to be significant in each case, so the relationship between the FFR and non-native perception remains unclear. Furthermore, the participants recruited were already highly proficient in their non-native language, so it is possible that the results were affected by experiential factors relating to the nature and extent of non-native exposure; as mentioned earlier, the FFR tends to be more robust in people with greater bilingual experience (Krizman et al., 2015). It is also worth specifying that none of these studies examined native phonetic perception in relation to the other factors, nor did they address the potential role of FFR consistency in phonetic perception.

The current study aimed to fill various gaps in the literature that have been alluded to above. In particular, performance on different native perception tasks—namely 2AFC and VAS tasks—appears to be related through consistency in behavioural responses, and performance on native and non-native perception tasks also appears to be related through the same behavioural consistency (Honda et al., 2024). The origins of this behavioural response consistency are still unidentified, but one plausible origin is neural response consistency as measured by the FFR. We therefore set out to determine whether consistent FFRs do indeed relate to consistent responses on phonetic perception tasks. Participants completed behavioural measures of native and non-native perception, as well as an EEG session during which the FFR was recorded. We employed both native and non-native phonetic perception tasks, given that few studies have used both behavioural and neurophysiological measures to look at individual differences in speech perception and that none have looked jointly at native and non-native perception. Additionally, prior work on non-native perception and the FFR may have had confounding experiential factors relating to non-native exposure; we addressed this limitation by studying differences in how listeners perceive non-native sounds to which they have not previously been exposed.

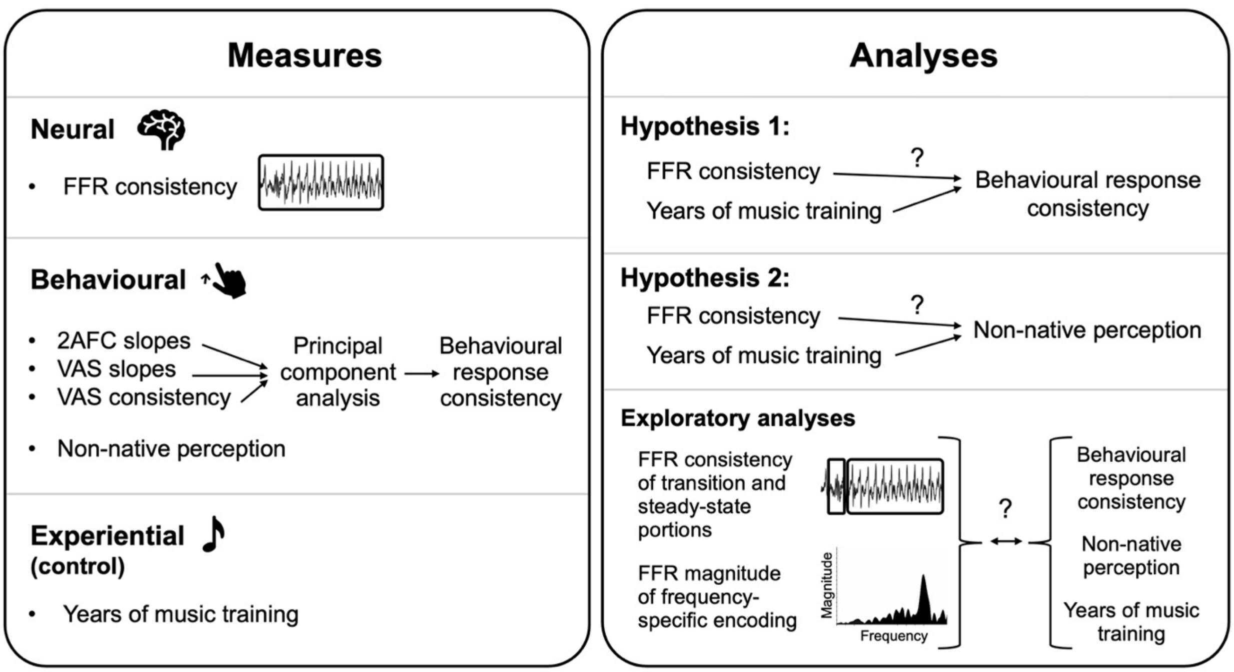
We hypothesized that the FFR, native phonetic perception, and non-native phonetic perception would be related within individuals. More specifically, we predicted that consistent behavioural responses to native speech sounds originate at the level of neural encoding. If this is the case, then more consistent FFRs would predict more consistent VAS responses and steeper 2AFC slopes (which reflect more consistent perception as described above; Honda et al., 2024; Kapnoula et al., 2017) on native phonetic perception tasks. Given that more consistent native perception has previously been associated with better non-native perception, we additionally predicted that consistent neural sound encoding promotes the successful perception of new language sounds. If this is the case, then more consistent FFRs would predict better performance on a non-native phonetic perception task.

## Methods

Data for this study were collected between July 2020 and July 2022. The raw data, the code needed to process and analyze it, and the full analysis output are all publicly available on the OSF (https://osf.io/nec8t/?view_only=bb062dcdd4664e9bbd629a6101eeb672). Note that the behavioural tasks were identical to Honda et al. (2024) but new participants were recruited for this study. Figure 1 provides an overview of the measures and analyses employed in this study, which are elaborated upon below.

**Figure 1**

*Summary of measures and analyses*




*Note.* Summary of measures (left) and analyses (right) for the current study. Calculation of FFR consistency is outlined in detail in Section 2.5.2, and calculation of behavioural measures is outlined in Section 2.5.3. Analyses are described in detail in Section 2.5.4.

## Participants

75 native English speakers (16 males, 59 females; mean age: 23.6) were recruited from the greater Montreal area. All 75 participated in an online behavioural session and were invited to an in-person EEG session; a subset of 40 ended up completing the EEG session. All participants completed a pre-screening questionnaire to ensure that they met our eligibility criteria (aged 18-35; right-handed; no history of literacy, language, cognitive, or hearing impairments; no history of head injury; speaker of a North American variety of English; self-identified monolingual with little to no knowledge of any languages other than English). The behavioural session had a duration of approximately 1-1.5 hours and the EEG session had a duration of approximately 30-45 minutes. Compensation of $17 for the online behavioural portion and $20 for the in-person EEG portion was provided. Participants signed a consent form approved by the Institutional Review Board of the Faculty of Medicine and Health Sciences of McGill University.

## Questionnaires

Handedness was assessed using the Edinburgh Inventory (Oldfield, 1971). A questionnaire adapted from the *Language history questionnaire* (LHQ 2.0; Li et al., 2013) and the *Montreal Music History Questionnaire* (MMHQ; Coffey et al., 2011) was used to collect information about demographics, language history and proficiency, and musical experience.

## Electroencephalography

The FFR was recorded while participants listened to multiple repetitions of the synthesized speech syllable /da/ (Chandrasekaran & Kraus, 2010). The syllable was 150 ms long and comprised a 10 ms consonant burst, a 30 ms formant transition, and a 110 ms steady-state vowel. In the steady-state portion, the stimulus had an F0 of 98 Hz, and the first through fifth formants (F1-F5) had values of 720, 1240, 2500, 3600, and 4500 Hz respectively. This syllable has been extensively used in previous FFR studies and is known to evoke characteristic peaks in response to its transient (/d/) and periodic (/a/) segments (Coffey et al., 2021; Skoe & Kraus, 2010). The stimulus was presented binaurally via insert earphones (EARTONE 3A) at 80 dB SPL, in alternating polarities (waveform shifted by 180° to reverse polarity) in order to better differentiate the lower-frequency and higher-frequency components of the FFR and to minimize stimulus artifact and the cochlear microphonic (Chandrasekaran & Kraus, 2010). There were 4000 total stimulus presentations (2000 per polarity), with an interstimulus interval of 97 ms. A vertical montage of 5 electrodes was used: an active electrode at Cz, two grounds on the forehead, and two references on the mastoids. Responses were recorded using BioSemi ActiABR software, at a sampling rate of 16384 Hz. Participants were instructed to simply relax and listen to the sounds, and the session took approximately half an hour including electrode application and removal.

**Behavioural Tasks**

Behavioural tasks were identical to Honda et al. (2024). Participants had five behavioural tasks to complete: two that measured native phonetic perception, one that measured non-native phonetic perception, one that measured sustained attention, and one that measured working memory. The phonetic perception tasks are described briefly below, and further details of all five tasks can be found in Honda et al. (2024). The sustained attention (Continuous Performance Task; Conners et al., 2003) and working memory (backwards digit span; Wechsler, 2008) measures were not included in the primary analyses of this experiment as the results of the previous study did not find any relationship between these control cognitive variables and the speech tasks; these measures were only included in behavioural data analyses whose goal was to directly replicate previous work (Honda et al., 2024). The behavioural tasks were completed online at home using the platform Gorilla.sc. Participants used their own headphones and underwent a headphone screening to standardize sound presentation and ensure an acceptable listening environment (Woods et al., 2017). When participants came to the lab in-person for the EEG session, an audiometric screening was conducted to ensure normal hearing thresholds in both ears (<35 dB at .5, 1, 2, 4, and 8 kHz).

#### ***Native Phonetic Perception Tasks***

Two native perception tasks provided measures of participants' identification slopes and response consistency. In these tasks, participants listened to minimal pairs that varied in different phonological contrasts (*bet-bat* and *dear-tear*; publicly available stimuli at https://osf.io/369my/, from Clayards, 2018). Each minimal pair was manipulated to vary orthogonally in two acoustic cues, forming a continuum of 25 stimuli per pair. The same stimuli were used in both native perception tasks. On each trial of the 2AFC task, participants chose between two options on the screen via mouse click to indicate what they heard (e.g., *bet* or *bat*; side of the screen counterbalanced across participants). On each trial of the VAS task, the screen displayed a slider with the two members of the minimal pair on opposite ends, and participants indicated where along the continuous slider they perceived the stimulus to be (from 0 to 100, with ends of the slider counterbalanced across participants).

### *Non-Native Perception Task*

The non-native perception task involved differentiating German vowels and consonants (øː vs. œ, yː vs. ʏ, ʃ vs. ç) which are known to be perceptually difficult speech sounds for native English speakers (Mayr & Escudero, 2010). German words containing these sounds were presented in a 3-interval oddity (3-I oddity) task. On each trial of the task, participants were presented with three German words spoken by three different voices. Participants then indicated which stimulus (if any) sounded different by clicking "1", "2", "3", or "None".

## Analyses

### *Data Exclusion*

For each participant's FFR, a signal-to-noise ratio was obtained by calculating the root mean square (RMS) of the EEG signal from 0 to 160 ms post-stimulus onset (signal) and dividing this by the RMS of the signal from -50 to 0 ms pre-stimulus onset (noise). Participants were excluded if their signal-to-noise ratio was <1.5 or >10, as this was taken to indicate a signal contaminated by noise or by artifacts that were not neural in origin. Six participants' FFR data were excluded on the basis of such signal quality issues, leaving a final sample of 34. During the response consistency analyses, participants were further excluded if they did not have at least 3400 (out of 4000) usable trials after artifact rejection, resulting in the exclusion of a further three participants from analyses involving FFR response consistency values.

Participants were excluded from the analyses of the behavioural data if they reported having exposure to German ($n = 1$), given that this could bias scores on the non-native perception task. They were also excluded on a task-by-task basis according to performance-based criteria established by Honda et al. (2024). Final numbers of participants included in each analysis are provided at the bottom of the table reporting that analysis.

***EEG Preprocessing***

EEG signals were bandpass filtered offline between 80-2000 Hz to isolate the FFR from lower-frequency evoked potentials and electrical line noise. Signals were epoched from -50 to 180 ms relative to stimulus onset and baselined to the 10 ms pre-stimulus onset. Artifact rejection with a threshold of $\pm 35$ μV was applied (number of trials remaining per participant after artifact rejection: mean = 3953 [out of 4000] and SD = 79, after excluding three participants as mentioned above). Subject averages were obtained by averaging epochs of each polarity and then summing the positive and negative polarity averages. Such adding of responses to both polarities emphasizes the FFR's temporal envelope (lower frequencies), making encoding of the stimulus' F0 more evident (Krizman & Kraus, 2019). These subject averages were then converted from the time domain to the frequency domain via fast Fourier transform (FFT), and spectral amplitude at the stimulus' F0 (98 Hz) was extracted as a basic measure of the strength of periodicity encoding (Skoe & Kraus, 2010). These processing steps and analyses were carried out using BrainVision Analyzer (2019; Brain Products GmbH, Germany). Our primary measure of interest was response consistency, which quantifies the stability of the FFR between trials for a given participant (Krizman & Kraus, 2019). To calculate response consistency, a participant's trials are randomly split in half and the average of one half is correlated with the average of the other half (Krizman & Kraus, 2019). We employed a bootstrapping method whereby this process is repeated 200 times, and the correlation values from each iteration are averaged to provide a final measure of response consistency. Such a bootstrapped approach is more rigorous than calculating the correlation value from a single iteration (Krizman & Kraus, 2019), and has been used in various other studies (e.g., Otto-Meyer et al., 2018; Parbery-Clark et al., 2013; Tierney & Kraus, 2013; White-Schwoch et al., 2015). Note that in order to ensure the soundness of our approach, we calculated response consistency values using different numbers of iterations (100, 300, 500, and 1000) and correlated these values with the ones obtained from our original 200 iterations; in all cases the correlation values were > 0.99, indicating highly robust output independent of the number of iterations. These consistency analyses were carried out in MatLab (version R2022a, The MathWorks Inc., USA).

***Preparatory Analyses of Behavioural Data***

Preparatory analyses for the behavioural data were identical to those reported in Honda et al. (2024); more details on the measures derived from each task can be found there. Identification slopes were calculated from both native phonetic perception tasks. For the 2AFC task, a mixed-effects logistic regression model was constructed for each minimal pair, with participants' 2AFC responses as the outcome and the two manipulated acoustic cues (formant frequency and vowel duration for *bet-bat*, voice onset time and onset F0 for *dear-tear*) as the predictors. The models included random intercepts and random slopes per participant, and the random slopes coefficients were extracted as the primary measure of interest (as in Clayards, 2018 and Kong & Edwards, 2015). A larger random slope coefficient for a given acoustic cue indicates that the participant responses were more influenced by changes in the stimuli compared to the group average. This analysis was performed using the lme4 package (Bates et al., 2015) in R (R Development Core Team, 2020). For the VAS task, slopes were calculated by fitting the rotated logistic developed by Kapnoula et al. (2017) to participants' average response to each stimulus. Slopes from the rotated logistic conceptually resemble the random slopes derived from 2AFC logistic regression just described, but they model gradiency independently of acoustic cue use (because our stimuli vary in two acoustic cues). The rotated logistic is modelled on a four-parameter logistic function containing estimates for minimum asymptote, maximum asymptote, slope, and crossover point. Its distinguishing feature is an additional parameter ($\theta$) representing the angle of the crossover point—the coordinate space gets rotated to be orthogonal to this angle so that the slope parameter provides a single gradiency measure that is independent of the two acoustic cues making up the space. For each participant, one slope value was derived from responses to the *bet-bat* continuum and another from responses to the *dear-tear* continuum. A smaller slope value from the rotated logistic reflects a steeper slope, that is, less gradient responses. This analysis was performed using MatLab (version 2015a, The MathWorks Inc., USA).

Behavioural response consistency was calculated using data from the VAS task. For each participant and trial, a residual value was obtained by calculating the difference between the participant's actual VAS response and their predicted response value based on the rotated logistic. The standard deviation of the resulting residuals was averaged for each minimal pair to provide estimates of consistency for the *bet-bat* and the *dear-tear* continua.

Differences in non-native speech perception ability were quantified using the non-parametric sensitivity index *A* (a corrected version of *A'*; Zhang & Mueller, 2005), which was calculated for each contrast in the 3-I oddity task. An *A* score of 0.5 denotes null discrimination and a score of 1.0 denotes perfect discrimination.

***Primary Analyses***

Prior to relating behavioural and neural measures, analyses were run to relate the behavioural measures of native and non-native speech perception. These analyses were identical to those reported in Honda et al. (2024), with the aim of determining whether patterns observed in this new dataset would be consistent with previous work. Specifically, canonical correlation and multivariate multiple regression were used to evaluate the relationships among native perception measures (2AFC slopes, VAS slopes, and VAS consistency), while multiple regression was used to determine the relationship between native and non-native perception. A principal component analysis (PCA) was run separately on the four 2AFC variables and the four VAS variables in order to distill the native perception measures to four principal components (two from each task); these components were used as predictors in the multiple regression relating native and non-native perception. Detailed descriptions of the analysis approach for the behavioural data can be found in Honda et al. (2024) and in the associated preregistration document (https://doi.org/10.17605/OSF.IO/9DKGQ).

Next, we conducted a PCA to derive a single measure of behavioural response consistency from the eight native speech perception measures. Finally, the following primary analyses tested our specific hypotheses:

1. *Consistent behavioural responses to native speech sounds originate at the level of neural encoding. If this is the case, then more consistent FFRs would predict steeper 2AFC slopes and more consistent VAS responses on a native phonetic perception task.* To test this, we ran a multiple regression analysis with FFR response consistency as the predictor of interest and music experience as a control predictor (music experience/ability has been found to relate to speech processing and perception, as mentioned in the introduction). Both predictors were standardized by subtracting the mean and dividing by the standard deviation. The outcome variable was the first principal component derived from our PCA, which reflected the consistency of participants' responses to native speech sounds – see below for details.
2. *Consistent neural sound encoding promotes the successful perception of new language sounds. If this is the case, then more consistent FFRs would predict better performance on a non-native phonetic perception task.* To test this, we ran another multiple regression analysis with FFR response consistency as the predictor of interest and music experience as a control predictor. Both predictors were standardized by subtracting the mean and dividing by the standard deviation. The outcome variable was non-native phonetic perception scores (*A* scores from the oddity task, averaged across contrasts). This model was run independently from the

one for the first hypothesis because it aimed to answer a distinct question; we were separately interested in assessing potential relationships between the FFR and behavioural consistency and between the FFR and non-native perception.[1]

Beyond these analyses, exploratory correlations and regressions were run to further investigate nuances of the potential relationships between the behavioural and neural data. Previous work has occasionally found significant relationships between the FFR and other variables only when a specific portion of the FFR is examined (the response to the initial transition portion of the stimulus vs. the response to the steady-state vowel portion; e.g., Chandrasekaran et al., 2009; Krizman et al., 2012; Krizman et al., 2015; Parbery-Clark et al., 2012). Accordingly, we performed the same regression analyses as described above, but with separate models in which the transition vs. steady-state portions of the FFR were entered as predictors. In addition, because frequency-specific magnitude is an FFR measure that has been more commonly examined than response consistency (Krizman & Kraus, 2019), we calculated each participant's FFR magnitude at the F0, F1, F2, and the first five harmonics (H1-H5) and explored whether these values correlated with our behavioural speech perception measures. Specifically, we ran exploratory Pearson's correlations to see whether these magnitude values related to behavioural response consistency (as derived from the PCA of native speech measures described above), to non-native perception scores, or to music experience. Finally, we correlated FFR response

[1]Initially, these primary analyses had FFR consistency as the dependent variable, with music experience and behavioural response consistency/non-native perception as the predictors. During the submission and revision process, analyses were modified to instead have the FFR as a predictor given the directionality of our hypotheses. For transparency, the original analyses are reported in the R Markdown document on this project's OSF page; qualitative conclusions from these analyses were the same as those from the updated analyses reported here.

consistency with the signal-to-noise ratio (SNR) and the root mean square (RMS) of the FFR in order to explore whether the consistency measure related to the clarity and power of the signal.

## Results

### Electroencephalography

There were considerable individual differences in the robustness of the FFR, in line with previous research (Coffey et al., 2016a). Figure 2A shows overall variation in consistency values, while Figure 2B shows example participants with consistent and inconsistent FFRs.

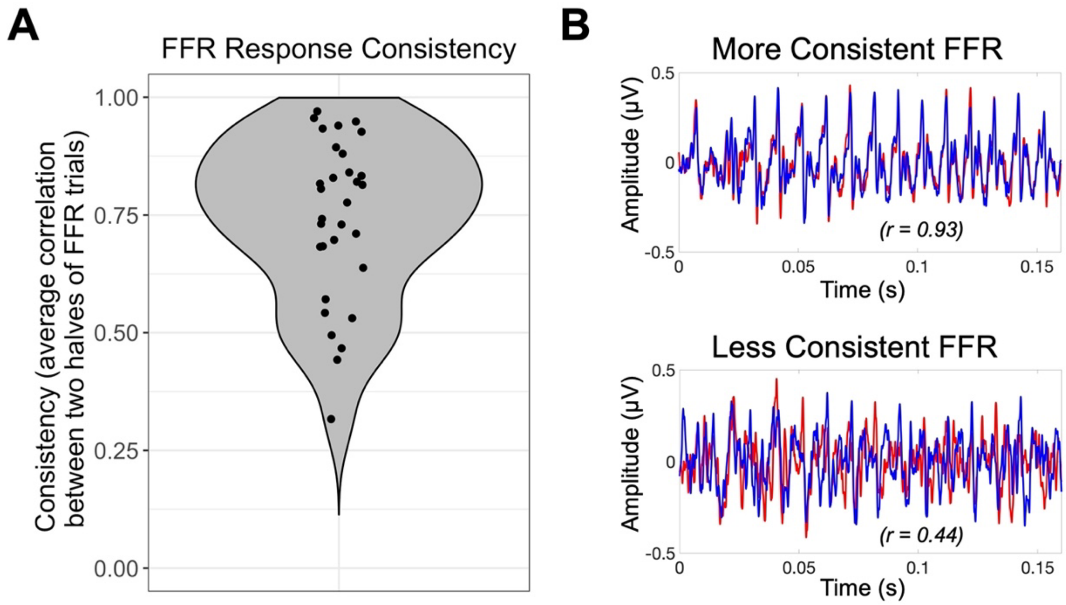
### Behavioural Data

#### *Descriptive Overview*

As anticipated based on previous work (e.g., Clayards, 2018; Kapnoula et al., 2017; Honda et al., 2024), considerable variability in performance on the 2AFC and VAS tasks was observed. Identification slopes for both tasks are displayed in Figure 3, revealing some participants with steep slopes and others with shallow slopes. Participants also differed in their response consistency on the VAS task, as displayed in Figure 4. Regardless of their identification slopes, some participants' responses were much more closely clustered around the same value for a given stimulus (i.e., consistent) compared to other participants with more scattered responses. Individual variation in performance was also observed for the non-native perception task, as displayed in Figure 5. Participants performed above chance (> 25%) and within the range expected based on previous work (e.g., Rauber et al., 2005; Silveira, 2011), though their accuracy varied greatly.

**Figure 2**

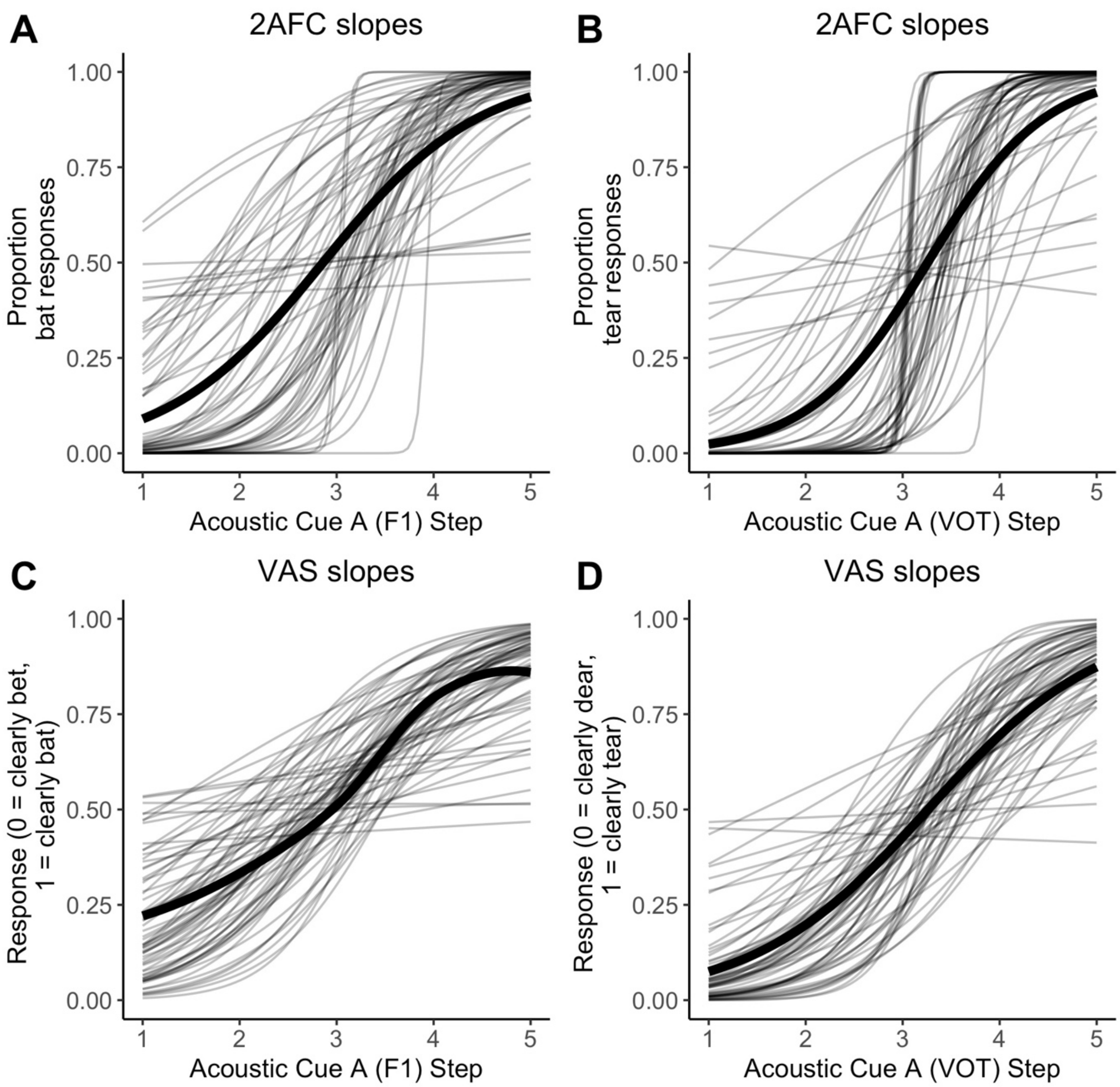
*Examples of variability in FFR consistency*




*Note.* (A) Violin plot of FFR response consistency values for all participants. Dots represent individual participants (they are horizontally jittered for visualization purposes), and the plot is trimmed at 1 because consistency values cannot exceed 1. (B) Two example participants, with a consistent FFR (top) and inconsistent FFR (bottom). Each participant's trials were randomly split in half and each half was averaged to produce a line (red for one half, blue for the other half). Greater overlap between the two lines indicates a more stable signal across trials and therefore greater consistency. The two halves are much more similar for the top participant ($r$ = 0.93) than for the bottom one ($r$ = 0.44).

**Figure 3**

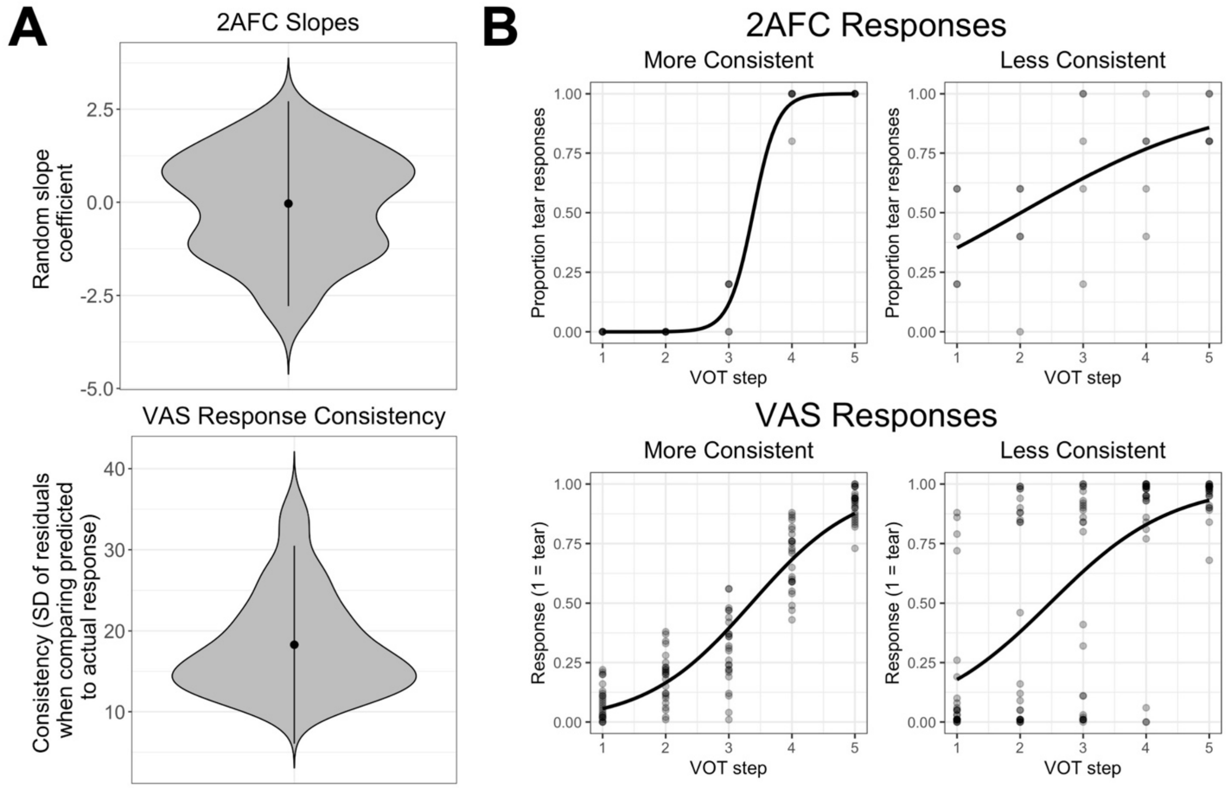
*Group and individual responses on the native perception tasks (2AFC and VAS)*




*Note.* (A) 2AFC bet-bat responses by cue A, (B) 2AFC dear-tear responses by cue A, (C) VAS bet-bat responses by cue A, and (D) VAS dear-tear responses by cue A. Thin lines are logistic curves fit to each individual participant for each step of acoustic cue A, and thick lines are logistic curves fit to the whole dataset. VAS responses varied continuously from 0-100, but were transformed to range from 0-1 for the purposes of fitting logistic curves to the data for these plots (regular logistic regression was used here for visualization purposes, rather than the rotated logistic function fit to the VAS data as described in the *Preparatory analyses of behavioural data* section). Note that while a handful of participants appear to be responding at chance, we relied on preregistered exclusion criteria from Honda et al. (2024) such that individuals were excluded if their endpoint responses were more than 2SD from the mean; these participants did not meet the criteria for exclusion.

**Figure 4**

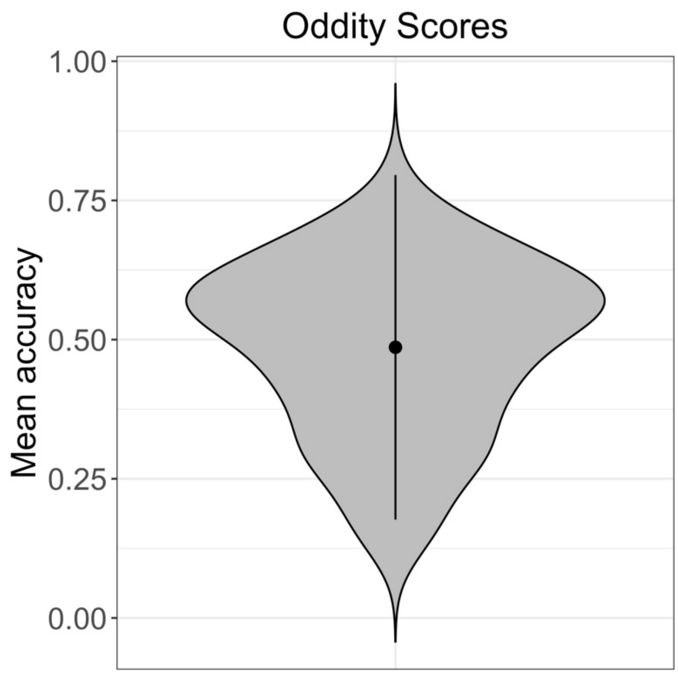
*Individual variability in behavioural response consistency*




*Note.* (A) Violin plots of 2AFC random slope coefficients (top) and VAS response consistency values (bottom) for all participants, averaged across both contrasts. The dot and vertical line denote the mean and standard deviation respectively. (B) Two example participants demonstrating consistent vs. inconsistent response patterns on the 2AFC and VAS tasks. For the 2AFC task (top), each dot is the participant's average response across the five presentations of a given stimulus—in some cases, responses overlap perfectly as indicated by more opaque dots. For the VAS task (bottom), each dot is a participant's response on a given trial. Lines are logistic curves fit to responses; dots clustered closely around the fitted curve indicate more consistent responses. VAS responses varied continuously from 0-100, but were transformed to range from 0-1 for the purposes of fitting logistic curves to the data for these plots (regular logistic regression was used here for visualization purposes, rather than the rotated logistic function fit to the VAS data as described in the *Preparatory analyses of behavioural data* section). The same two example participants are shown across both tasks, illustrating that participants tend to respond with similar consistency regardless of the task.

**Figure 5**

*Violin plot of scores on the 3-I Oddity non-native perception task*



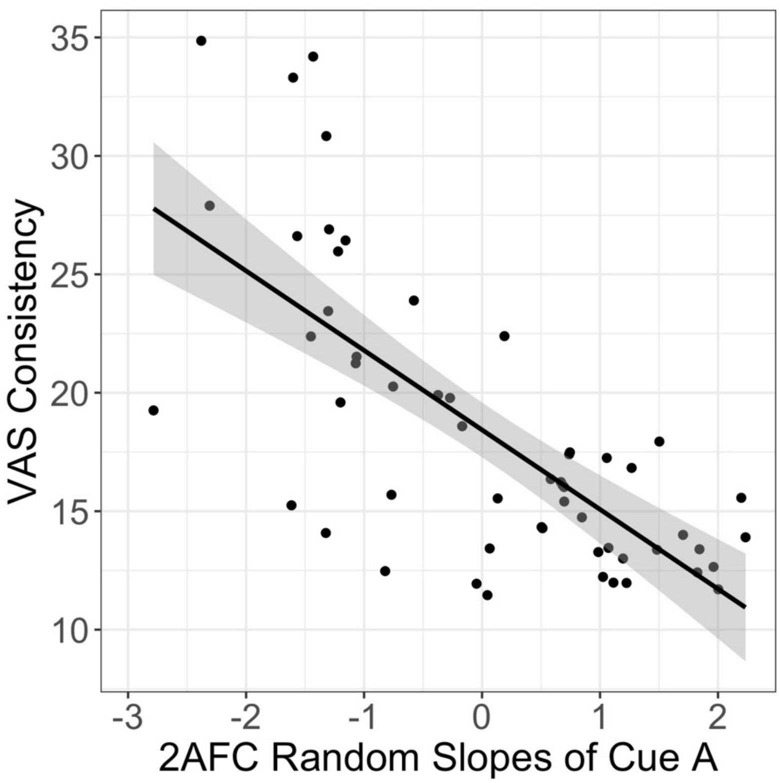

*Note.* Accuracy is averaged across contrasts. The dot and vertical line denote the mean and standard deviation, respectively.

### ***Relationships Among Behavioural Speech Perception Measures***

Replicating the results of Honda et al. (2024), canonical correlation analyses revealed a significant relationship between 2AFC slopes and VAS consistency ($r_c$ = 0.80, $p <$ 0.001 along the first canonical dimension) but not between 2AFC slopes and VAS slopes ($r_c$ = 0.35, $p$ = 0.35 along the first canonical dimension). The complete canonical coefficients and their significance can be found in Supplemental Table S.1, along with loadings of the original variables onto the canonical dimensions in Supplemental Table S.2. Follow-up multivariate multiple regression further confirmed this finding by demonstrating that 2AFC slopes were significantly predicted by the consistency of VAS *dear-tear* responses ($V$ = 0.32, $p$ = 0.003) but not by the VAS slope measures. In particular, more consistent VAS responses were associated with steeper 2AFC slopes, reinforcing Honda et al.'s (2024) claim that 2AFC slopes reflect the consistency of phonetic perception. A summary of the multivariate model can be found in Supplemental Table S.3. Figure 6 illustrates the relationship between 2AFC slopes and VAS consistency.

**Figure 6**

*Scatterplot of the relationship between 2AFC slopes for acoustic cue A and VAS consistency*




*Note.* 2AFC slope values and VAS consistency values are averaged across both contrasts. Higher 2AFC slope values indicate steeper slopes, while higher VAS consistency values indicate less consistent responses. Each dot represents a single participant, and a line of best fit with 95% CI is shown.

As for the relationship between native and non-native perception, multiple regression revealed that the first principal component from the 2AFC variables (reflecting the steepness of 2AFC slopes across acoustic cues and contrasts) and the attention scores significantly predicted performance on the non-native 3-I oddity task (2AFC component: $\hat{\beta}$ = -0.12, $p$ = 0.02; attention scores: $\hat{\beta}$ = -0.12, $p$ = 0.05). These relationships were such that steeper 2AFC slopes and better attention were associated with better non-native perception. A summary of the principal components derived from the 2AFC and VAS variables can be found in Supplemental Table S.4, and full model output can be found in Supplemental Table S.5. Rather than 2AFC slopes, Honda et al. (2024) had previously found that VAS response consistency was predictive of non-native perception. However, 2AFC slopes and VAS consistency have now been shown to relate to each other both in the present experiment and across both experiments reported by Honda et al. (2024); it therefore makes sense that the consistency of responses to native sounds is the underlying factor of importance for successful non-native perception in both cases. For these analyses, as in Honda et al. (2024), Cook's distance was calculated in order to check for influential participants. One participant was found to have higher influence than the others for this model; when they were excluded from the multiple regression model as an exploratory analysis, there were no longer any significant predictors of non-native perception. As such, we caution against interpreting the aforementioned findings too strongly—however, close examination of the participant's performance on the native and non-native tasks revealed reasonable response patterns within the anticipated range. For the sake of transparency, the model excluding the participant is reported in Supplemental Table S.6.

**Relating Neural and Behavioural Measures – Primary Analyses**

By running a PCA on the eight native speech perception measures, we derived a single measure of behavioural response consistency. The first principal component derived from the analysis was made up primarily of the four 2AFC random slope coefficients and the two VAS consistency measures (Table 1)—that is, all of the measures that are expected to reflect the consistency of participants' behavioural responses. In contrast, the second principal component was made up primarily of the two VAS slope measures, which reflect the gradiency of native speech perception. This served to reinforce our previous conclusion that consistency and gradiency are separate constructs and that together they explain most of the variance in performance across 2AFC and VAS tasks. The first principal component (which accounted for 52% of the variance) was extracted for use in our subsequent primary analyses, as a measure of behavioural response consistency. In order to facilitate interpretation, the component's sign was reversed (multiplied by -1) so that larger values reflected greater behavioural consistency.

**Table 1**

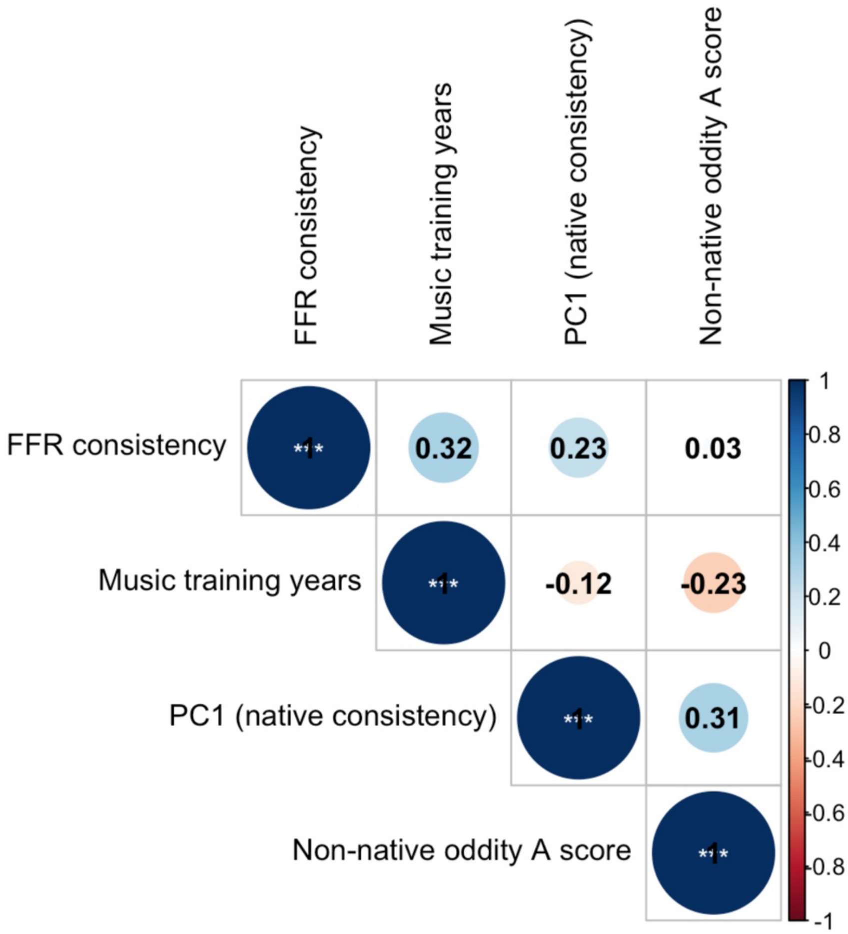
*Correlations between the original native speech perception variables and the first two principal components extracted from them*

| | PC1 | PC2 |
|---|---|---|
| 2AFC bet-bat acoustic cue A | -0.43 | -0.05 |
| 2AFC bet-bat acoustic cue B | -0.37 | -0.07 |
| 2AFC dear-tear acoustic cue A | -0.42 | -0.15 |
| 2AFC dear-tear acoustic cue B | -0.36 | -0.13 |
| VAS bet-bat consistency | 0.40 | -0.08 |
| VAS dear-tear consistency | 0.43 | 0.07 |
| VAS bet-bat slope | -0.15 | 0.66 |
| VAS dear-tear slope | -0.07 | 0.71 |
| *Percent variance explained* | 52% | 15% |

Prior to conducting the main analyses of interest, pairwise Pearson correlations were computed between the primary neural and behavioural measures of interest for the sake of preliminary visualization. A plot of these correlations is displayed in Figure 7. Although there were no significant correlations between any of the variables, the majority of the correlations were in the direction that would be anticipated based on our hypotheses and on previous work; namely, the correlations between neural consistency and the other three variables (behavioural consistency, music experience, and non-native perception) were all positive, as was the correlation between behavioural consistency and non-native perception. It is possible that true relationships exist between these variables but that the current study did not have the power to detect them (e.g., due to the true effect size being small or to the logistical limitations of recruiting large numbers of EEG participants).

**Figure 7**

*Correlation plot showing Pearson correlation values between each pair of variables of primary interest*




*Note.* Circle size indicates correlation strength and circle colour indicates correlation strength and sign. Asterisks denote statistically significant correlations (*** = p < .001, ** = p < .01, * = p < .05).

For the regression model with behavioural response consistency as the outcome variable, neither FFR consistency ($\hat{\beta}$ = 0.44, *p* = 0.15) nor music experience ($\hat{\beta}$ = -0.32, *p* = 0.30) were significant predictors. See Table 2 for a full model summary. Contrary to our expectations, we therefore did not find evidence of a relationship between the consistency of neural and behavioural responses to speech sounds. Nonetheless, it is worth noting that the regression coefficient for FFR consistency was in the anticipated (positive) direction, with greater neural response consistency being associated with greater behavioural response consistency; despite a lack of statistical significance, we cannot necessarily rule out the existence of a true relationship which the current study may have been unable to detect.

For the regression model with non-native speech perception as the outcome variable, FFR consistency ($\hat{\beta}$ = 0.01, $p$ = 0.59) and music experience ($\hat{\beta}$ = -0.03, $p$ = 0.20) were again not significant predictors. A full model summary is displayed in Table 3. Somewhat surprisingly, there is therefore no evidence of a relationship between more consistent neural responses and better non-native speech perception. However, the coefficient for FFR consistency was again in the anticipated positive direction. Note that due to our limited sample size, these analyses should be considered preliminary. As we mention in the Discussion, it is possible that true relationships exist between our variables but that the present study was not able to detect them.

**Table 2**

*Summary of the regression model predicting behavioural response consistency from FFR consistency*

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* | *95% CI* |
|---|---|---|---|---|---|
| (Intercept) | 1.13 | 0.28 | 4.06 | <.001 | [0.55, 1.70] |
| FFR consistency | 0.44 | 0.30 | 1.49 | 0.15 | [-0.17, 1.05] |
| Music training | -0.32 | 0.30 | -1.07 | 0.30 | [-0.94, 0.30] |

*n = 28*

**Table 3**

*Summary of the regression model predicting non-native perception from FFR consistency*

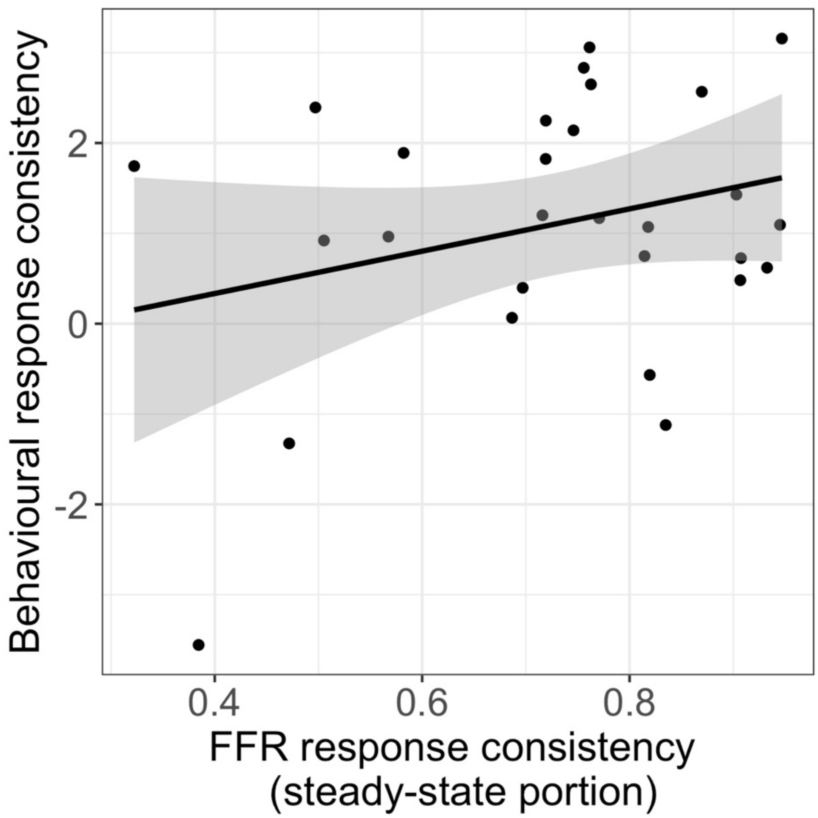
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* | *95% CI* |
|---|---|---|---|---|---|
| (Intercept) | 0.21 | 0.02 | 10.06 | <.001 | [0.17, 0.26] |
| FFR consistency | 0.01 | 0.02 | 0.55 | 0.59 | [-0.03, 0.06] |
| Music training | -0.03 | 0.02 | -1.33 | 0.20 | [-0.08, 0.02] |
| *n = 28* | | | | | |

**Relating Neural and Behavioural Measures – Exploratory Analyses**

For the exploratory regressions with the consistency of the transition portion of the FFR as a predictor, neither the model with behavioural consistency as the outcome nor the one with non-native perception as the outcome revealed any significant predictors. For the models with the steady-state portion of the FFR as a predictor, the model with non-native perception as the outcome again showed no significant predictors. However, the model with behavioural response consistency as the outcome had FFR response consistency as a marginal predictor ($\hat{\beta}$ = 0.53, $p$ = 0.08). The observed relationship was in the expected direction such that more consistent FFR responses in the steady-state portion were associated with greater behavioural response consistency. A scatterplot of the relationship is displayed in Figure 8, and full output for the model can be found in Table 4.

**Figure 8**

*Relationship between neural response consistency and behavioural response consistency*




*Note.* Scatterplot of the marginal relationship between behavioural response consistency (quantified by a PCA as described above) and neural response consistency. Higher consistency values indicate more consistent neural and behavioural responses. Each dot represents a single participant, and a line of best fit with 95% CI is shown.

**Table 4**

*Summary of the regression model predicting behavioural response consistency from the consistency of the steady-state portion of the FFR*

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* | *95% CI* |
|---|---|---|---|---|---|
| (Intercept) | 1.13 | 0.27 | 4.17 | <.001 | [0.57, 1.70] |
| FFR consistency (steady-state portion) | 0.53 | 0.30 | 1.80 | 0.08 | [-0.08, 1.14] |
| Music training | -0.37 | 0.30 | -1.24 | 0.23 | [-0.98, 0.25] |

*n* = 28

The exploratory correlations between frequency-specific magnitude of the FFR (F0, F1, and H1-H5 encoding strength), phonetic perception, and music experience did not reveal significant relationships. A handful of relationships were initially marginal or significant as determined by Pearson's correlations (e.g., between F0 encoding and music experience, and between H1 encoding and the consistency of native perception), however significance no longer held after implementing the Benjamini-Hochberg procedure to correct for multiple comparisons. Our final exploratory correlations revealed that FFR response consistency was positively correlated with both the SNR ($r = 0.73$, $p < 0.001$) and RMS ($r = 0.55$, $p = 0.001$) of the signal; participants with more consistent neural responses also tended have a clearer and more powerful FFR signal. Full details and output of all exploratory analyses can be found in the R Markdown document on this project's OSF page (https://osf.io/nec8t/?view_only=bb062dcdd4664e9bbd629a6101eeb672), as mentioned above.

## Discussion

This study aimed to investigate individual differences in the neural encoding of speech and in behavioural performance on native and non-native speech perception tasks, and to better understand these differences by exploring the relationships among them. We were particularly interested in whether the consistency of neural responses to speech (as measured by the FFR) predicted (i) the consistency of behavioural responses to native speech and (ii) sensitivity to non-native speech sound contrasts on a behavioural task. Somewhat surprisingly, we did not find strong evidence of relationships between neural speech encoding and performance on any of the speech perception tasks. However, patterns of performance on the behavioural speech perception tasks were similar to those observed in previous work (Honda et al., 2024; Kapnoula et al., 2017), strengthening our understanding of individual differences in speech perception.

**Behavioural Responses to Native and Non-Native Speech Sounds**

Our behavioural results helped to clarify how individual differences in performance on different native and non-native phonetic perception tasks are related. We found that 2AFC and VAS slopes were not related within individuals, but that steeper 2AFC slopes were related to more consistent VAS responses. This is a direct replication of Honda et al. (2024) and the same pattern as reported by Kapnoula et al. (2017). In replicating these two prior studies, one of which had found only marginal evidence for a relationship between 2AFC slopes and VAS consistency (Kapnoula et al., 2017), we strengthen the evidence in favour of the emerging view that 2AFC and VAS tasks do not measure the same construct. We can conclude that 2AFC slopes tap into consistency (with shallower slopes reflecting more variable responses across repeated presentations of a stimulus) while VAS slopes tap into gradiency (with shallower slopes reflecting the perception of fine-tuned phonetic differences between stimuli along a continuum), and that performance on the two tasks is related through the consistency of responses. These findings are useful for speech perception researchers to bear in mind when designing new studies; tasks and measures should be selected according to the constructs of primary interest. For instance, if the goal is to measure gradiency, then VAS slopes should be used rather than assuming that 2AFC slopes are an appropriate approximation (e.g., as in Ou & Yu, 2022). Our results additionally highlight behavioural response consistency as an informative measure for studying individual differences in speech perception and for relating performance across tasks.

Reinforcing the importance of behavioural response consistency, we also found that steeper identification slopes on a 2AFC task may predict better labelling of unfamiliar non-native speech sounds. Our findings are thus in line with research that has reported links between consistent native perception and successful non-native perception (Fuhrmeister et al., 2023; Honda et al., 2024; Kapnoula & Samuel, 2023). These links are not necessarily strong (found in exploratory but not preregistered analyses by Honda et al., 2024) and are only just beginning to be explored, but they may prove to be a productive topic of future investigation. Given that some adults are much more successful than others at learning to identify non-native speech sounds (e.g., Bradlow et al., 1997; Hattori & Iverson, 2010) and that speech sound learning supports non-native language learning more broadly (e.g., Jakoby et al., 2011), it is important to identify the predictors of such success. Potential predictors such as consistent native perception could then be targeted by training programs to improve both native and non-native perception, or could be harnessed as pre-screenings to identify adults who might require more support during the language learning process.

**Relationship Between Neural and Behavioural Response Consistency**

Turning to the main objective of the study—the potential relationship between neural encoding consistency (as reflected in the FFR) and phonetic perception—contrary to our hypothesis, our primary analyses did not find strong evidence that neural consistency of sound was linked to behavioural performance on native or non-native phonetic perception tasks. It should be noted that both the zero-order correlations (Pearson) and the regressions that controlled for music experience found a pattern of results consistent with our expectations, but did not reach significance. This could mean that our method to detect the relationship was not sensitive enough, or that the size of the effect was not big enough to be detected with our sample size. The results thus should not be considered definitive in either supporting or failing to support our hypothesis.

Furthermore, exploratory analyses revealed that the consistency of the steady-state portion of the FFR was marginally related to the consistency of behavioural responses to native speech sounds. This is not the first time that relationships between the FFR and other variables have been found when specifically examining the steady-state portion of the FFR. For example, Krizman et al. (2012) previously found that bilinguals had better F0 encoding than monolinguals in quiet as well as in noise, but this was the case only for the steady-state portion of the FFR (not for the transition portion). Similarly, other work has found that simultaneous bilinguals have better F0 encoding and more consistent FFRs than sequential bilinguals, specifically during the steady-state portion of the response (Krizman et al., 2015). Compared to the transient portion of the FFR, the steady-state portion involves greater cortical contributions (Coffey et al., 2016b). It is therefore possible that when studying the FFR in relation to factors such as language exposure or the consistency of phonetic perception, the steady-state portion is the most likely to reveal relationships because it is more influenced by top-down cortical processes that are also relevant to higher-level behavioural and experiential variables. Note, however, that the relationship we found between FFR steady-state consistency and behavioural response consistency did not reach significance; furthermore, *p* values (and marginal ones in particular) should be treated with caution due to sampling variability, because only very small values are likely to be clearly diagnostic of an effect (Cumming, 2008). Further work with larger sample sizes and with a particular focus on the steady-state portion of the FFR may clarify these findings and determine their generalizability.

The overall lack of clear evidence for links between FFR consistency and phonetic perception could be due to a variety of factors. Firstly, our phonetic perception tasks did not require participants to focus on pitch perception. Pitch perception is a behavioural measure that has been commonly found to relate to the FFR (e.g., Bidelman & Krishnan, 2010; Coffey et al., 2016a; Zhang & Gong, 2017), which is logical given that a characteristic feature of the FFR is its representation of pitch. Perhaps brain-behaviour relationships would have been more evident if our native and non-native speech perception tasks had involved pitch perception by testing the perception of lexical tones; for instance, if we had related the FFR to how consistently native Mandarin speakers perceived continua of Mandarin tones and to how well they labelled non-native Thai tones. Nonetheless, we were interested in studying speech perception beyond tone perception, which is why we selected both consonants and vowels as stimuli for our native and non-native perception tasks.

Unlike pitch perception and the FFR, links between phonetic perception and the FFR have very rarely been explored, and the nature of these links is not clear. In one case, participants with better perception of non-native English consonants were found to have better FFR phase-locking to the F0, but this finding did not hold for the perception of

English vowels or for other FFR measures such as phase-locking to harmonics (Omote et al., 2017). In another study, participants with better perception of non-native English vowels were found to have better FFR phase-locking to F2, although this was not the case for phase-locking to the F0 or F1 (Kachlicka et al., 2019). As such, it is possible that phonetic perception measures simply do not show reliable relationships to the FFR—the encoding of periodic sounds by the FFR may be slightly too far removed from the multitude of perceptual and cognitive processes involved in labelling speech sounds (at least, in non-tonal languages that place relatively less importance on pitch). It is conceivable, for instance, that instead of originating at the level of early neural sound encoding as we hypothesized, the behavioural response consistency observed on our native phonetic perception tasks instead originates at a later processing stage with a different neural signature (e.g., during higher-order processing by auditory association areas, or during the decision of which response button to press). Work with cortical event-related potentials might help to clarify this possibility. Prior research has looked at the amplitude of cortical event-related potentials, finding that the amplitude of the N1 (reflecting a late stage of perceptual processing) varies linearly with changes in stimulus voice onset time (Toscano et al., 2010) and also varies as a function of place of articulation for both fricatives and stop consonants (Pereira et al., 2018). Similarly, the amplitude of the P3 (reflecting a post-perceptual categorization phase) has been found to vary in a graded way alongside changes in stimulus voice onset time (Toscano et al., 2010). The N1 and P3 are therefore informative event-related potentials that relate to later stages of phonetic perception, but rather than studying their amplitude at the group level, future research could derive measures of response consistency from them at the individual level in order to relate the consistency of cortical responses to the consistency of behavioural responses. While such individual differences in cortical event-related potentials have rarely been examined, they will likely become a rich area for future research as the field increasingly acknowledges the importance of studying individual variability, and as EEG recording and analysis methods continue to be refined.

Another possibility is that the range of individual differences observed in healthy young adults (as studied here) is not extensive enough to reveal potential relationships between the measures that we used. As an example, a previous study correlated children's FFR response consistency with their phonetic discrimination (Tecoulesco et al., 2020). The researchers found a significant positive relationship between these measures when combining data from a typically developing group with data from a group with autism spectrum disorder, but found no relationship when examining the typically developing group in isolation (Tecoulesco et al., 2020). Prior work had already established that FFR response consistency differs significantly between typically developing children and children with autism spectrum disorder (Otto-Meyer et al., 2018); these differences may have been what enabled the emergence of a relationship between the FFR and phonetic discrimination. Similarly, children with poor reading skills show less consistent FFRs than those with good reading skills (Hornickel & Kraus, 2013), and this variation may account in part for why FFR consistency has been found to relate to tests of emergent literacy (White-Schwoch et al., 2015). The aforementioned brain-behaviour relationships involving FFR response consistency in children may also have emerged because the FFR does not fully mature and stabilize until early adulthood (Skoe et al., 2015). It may therefore be that relationships between the FFR and phonetic perception are more likely to be observed only in more heterogeneous populations that show large variations in performance—for example, when studying participants who are neurodiverse—or at earlier stages of development when differences may be more marked.

Additionally, our results could indicate that any potential relationship between the FFR and phonetic perception does not become evident until after significant non-native language exposure. As reported by Reetzke et al. (2018), native English speakers who are trained to label Mandarin tones show improvements both in behavioural tone discrimination and in neural encoding of the tones in the FFR; however, the FFR improvements are observed well after the behavioural improvements, only once participants have been “over-trained” for several days after having reached native-like discrimination levels. In a similar vein, Song et al. (2008) found that native English speakers’ FFRs showed better tone pitch tracking after a two-week Mandarin tone training program, but only for the pitch contour that was the most acoustically complex. This result lends further support to the idea that FFR-behaviour relationships may emerge only with extensive non-native training. The findings could also suggest that relationships between the FFR and non-native sounds can only be found when the non-native stimuli are particularly unfamiliar to participants; the German sounds used in the current study may not have been distinct enough from English sounds compared to a complex Mandarin tone contour. Furthermore, in the other previous studies that have reported relationships between the FFR and non-native speech perception or production, all participants were already highly proficient in their non-native language (Kachlicka et al., 2019; Omote et al., 2017; Saito et al., 2018); again, the results of these studies may have differed from ours because relationships between the FFR and non-native speech measures only emerge with greater non-native experience. Nevertheless, it is also important to investigate FFR-behaviour relationships in listeners without non-native language experience as done here; this not only controls for differences in non-native exposure, but if relationships are eventually found, this could have implications for the many adults who are at early stages of learning a new language. For instance, perhaps an element of the FFR will later be found to relate to some aspect of early non-native language learning in adults, providing a neural marker of learning success or difficulty which could be used to personalize the learning process. It should also be noted that even in the prior studies relating the FFR to non-native perception or production, the findings have been mixed and not aligned with straightforward hypotheses as specified above (Kachlicka et al., 2019; Omote et al., 2017; Saito et al., 2018). Consequently, one possibility is that the FFR and non-native speech outcomes are simply not strongly related, even in populations with extensive non-native language experience.

An additional point is that we found FFR consistency to be correlated with the power (RMS) and clarity (SNR) of the FFR signal. Consequently, we cannot rule out the possibility that our neural consistency measure reflects additional factors beyond consistency, such as skull thickness or experiential factors that could impact signal strength. Nonetheless, we controlled for language experience by recruiting functionally monolingual participants, and we included music experience in our analyses in order to control for its potential relationship to the FFR. Response consistency is also a well-documented FFR measure that has been described and employed in numerous previous studies (see Krizman & Kraus, 2019, for a summary).

As for our behavioural measures, one potential question that some might raise is whether different individuals use different strategies when responding to speech sounds on a VAS task, influencing performance. There are a few reasons why we believe that this is unlikely to be the case. For one, correlations between individuals' auditory and visual (e.g., for stimuli ranging visually from an apple to a pear) VAS slopes are weak and insignificant in most cases (Kapnoula & McMurray, 2021; Kapnoula et al., 2021). In addition, Kapnoula and McMurray (2021) have shown that gradient phonetic perception on a VAS task relates to gradient phonetic perception as measured by electroencephalography (i.e., smaller P3 event-related potentials, reflecting less strong post-perceptual categorization) and by eye-tracking in a visual world paradigm (i.e., more looks to a competitor item when a phoneme is more ambiguous). Together, this evidence suggests that VAS slopes can indeed measure the construct of gradiency and that VAS responses are not simply a result of participant response strategies that would bias performance across modalities. Recall also that our measure of behavioural response consistency was obtained from a principal component analysis of both 2AFC- and VAS-derived measures, and so even if response strategies affect VAS performance, it is unlikely that they had an impact on our aggregate consistency measure.

A final consideration is that the FFR has now been shown to arise from multiple neural generators, from brainstem nuclei up to higher-level auditory cortical regions; a consequence of this is that the signals from different generators can undergo constructive or destructive interference, with the potential to cancel each other out (Coffey et al., 2016b; Tichko & Skoe, 2017). As such, there is the possibility that a particular FFR generator (e.g., auditory cortex) is related to our native and/or non-native perception measures, but that other generators (e.g., superior olive or inferior colliculus) are less related to speech perception and are in fact cancelling out part of the signal from the related generator. To clarify this, future work could use combinations of functional neuroimaging (fMRI), magnetoencephalography (MEG), and computational modelling to localize specific neural sources of the FFR that may relate to differences in native and non-native phonetic perception.

**Future Directions**

Other potential future directions include relating the FFR not only to phonetic perception but also to phonetic production, or conducting longitudinal learning studies to determine how neural and behavioural measures of speech processing may change and relate to each other over a longer timescale. For instance, adults could undergo testing (FFR recording, native and non-native speech perception tasks, and non-native aptitude measures such as vocabulary and reading comprehension scores) at two timepoints: once at the beginning and once at the end of an intensive non-native language learning course. Perhaps consistent neural sound encoding or native perception at pretest would predict eventual non-native language learning outcomes, or changes in neural sound encoding would predict the extent of improvement in non-native learning over the course of several weeks. As mentioned above, it could also be fruitful to test populations that may tend to show more extremes in variability on neural and behavioural speech processing measures. This could include, for instance, examining relationships between the consistency of neural and behavioural responses to speech in children, or in adults with and without autism spectrum disorder or dyslexia. It is also interesting to note that stable or consistent neural activity may not always be optimal; for example, it is possible that variability in neural activity provides a useful index of uncertainty, with less clear stimuli eliciting more variable neural responses (Waschke et al., 2021). In the current study, stimulus clarity was not a factor since we recorded neural responses to a single syllable (/da/). However, future work could compare the consistency of neural responses to speech sounds varying in clarity (e.g., presented in varying amounts of noise, or varying along a continuum similarly to Ou et al., 2023) and relate this to behavioural consistency, in an effort to further untangle potential brain-behaviour associations and understand the basis of successful phonetic perception.

**Broader Implications**

Although the current study only directly examines phonetic perception and neural encoding of frequency, it could also have implications for other elements of language processing and acquisition. It has recently been demonstrated that basic auditory temporal and spectral processing (as measured by behavioural discrimination tasks and FFR encoding of stimulus frequencies) predicts non-native pronunciation proficiency on a narrative task in experienced learners (Saito et al., 2020). Also, auditory pitch detection has been linked to vocabulary learning and grammatical rule extraction in a new language (Chandrasekaran et al., 2010; Llompart, 2020; Mueller et al., 2012; Wong & Perrachione, 2007). For example, studies have shown that native English listeners' ability to discriminate vowels with different pitch patterns is significantly related to their ability to learn pseudowords that differ in those pitch patterns (Chandrasekaran et al., 2010; Wong & Perrachione, 2007). Relatedly, Llompart (2020) found that non-native phonetic discrimination predicted non-native word identification, and Silbert et al. (2015) found that the ability to differentiate unfamiliar voicing, place, and tone contrasts predicted performance on a foreign word learning task involving those same contrasts. EEG work has also linked non-native phonetic perception to broader non-native language proficiency, revealing that participants with more efficient (shorter latency) mismatch negativity and P3a responses to a non-native vowel also scored higher on a non-native vocabulary and reading comprehension exam (Jakoby et al., 2011). Furthermore, larger mismatch negativity responses to pitch-deviant stimuli were found in participants who successfully learned a syntactic rule in an artificial language compared to ones who did not (Mueller et al., 2012). Together, these findings suggest that basic auditory processing and phonetic perception are important precursors for non-native language acquisition more generally. As such, the individual differences that we observed in phonetic perception and in sound encoding as measured by the FFR could have repercussions for non-native language learning outcomes beyond phonetic perception. Studies should continue to examine individual differences at various levels of language processing, from early auditory encoding to high-level syntactic and semantic processes, in order to untangle relationships between these differences and to understand the kind of profile that might characterize a successful learner.

## Conclusion

Using a multimodal approach combining EEG and behavioural tasks, we aimed to provide new insight into the neural and behavioural factors supporting successful native and non-native speech perception. To our knowledge, this study is the first to incorporate neurophysiological and behavioural measures in order to explore relationships among individual differences in the FFR, native phonetic perception, and non-native phonetic perception. All of the measures that we collected are known to show reliable individual differences (e.g., Coffey et al., 2016a; Kong & Edwards, 2015; Kapnoula et al., 2017); it is important to identify the correlates of such differences in order to gain a more comprehensive understanding of what constitutes a “good listener” in the context of speech perception. Our behavioural results showed that (i) different native perception tasks (2AFC and VAS) are related through consistency in responses and (ii) consistent perception of native sounds predicts successful labelling of non-native speech sounds, bolstering other recent research (Fuhrmeister et al., 2023; Honda et al., 2024; Kapnoula et al., 2017). While our neural results did not reveal clear evidence of relationships between the FFR and phonetic perception, exploratory analyses showed the possibility of a weak relationship between neural response consistency (specifically in the steady-state portion of the FFR) and behavioural response consistency during native phonetic perception. This work contributes to our understanding of the factors at play in the earliest stages of encountering a new language, and of why certain adult learners develop better non-native skills than others. Ultimately, such investigations of individual differences could result in the creation of personalized approaches for facilitating language acquisition based on each individual's neural and perceptual profile. Enhancing non-native language learning in adults is particularly important given that we live in an era of ever-increasing migration and globalization, in which it has become commonplace for many people to communicate in languages other than their native one.

**CRediT Authorship Contribution Statement**

Claire T. Honda: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft. Meghan Clayards: Conceptualization, Methodology, Supervision, Writing – review & editing. Shari R. Baum: Conceptualization, Methodology, Supervision, Writing – review & editing.

**Declaration of Competing Interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Data Availability**

All data from this experiment is publicly available on the Open Science Framework (OSF) and can be accessed at https://osf.io/nec8t/?view_only=bb062dcdd4664e9bbd629a6101eeb672.

**Acknowledgements**

The authors would like to thank Don Nguyen for help with data collection and technical troubleshooting, Morgan Sonderegger for advice about behavioural data analyses, and Emily Coffey for advice about EEG analyses. They would also like to thank the Centre for Research on Brain, Language and Music (CRBLM) for providing EEG equipment and lab space, as well as the Fonds de recherche du Québec – Nature et technologies for providing funding to the CRBLM. This work was supported by an NSERC grant awarded to Shari R. Baum (RGPIN-2023-03239), a SSHRC grant awarded to Meghan Clayards (435-2020-1140), and NSERC CGS M and PGS D grants along with a McGill Faculty of Medicine Max Binz Fellowship awarded to Claire T. Honda.

## References

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,67*(1). https://doi.org/10.18637/jss.v067.i01.

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain research*, *1355*, 112-125. https://doi.org/10.1016/j.brainres.2010.07.100.

Bidelman, G. M., Pousson, M., Dugas, C., & Fehrenbach, A. (2018). Test–retest reliability of dual-recorded brainstem versus cortical auditory-evoked potentials to speech. *Journal of the American Academy of Audiology*, *29*(02), 164-174.

Bones, O., Hopkins, K., Krishnan, A., & Plack, C. J. (2014). Phase locked neural activity in the human brainstem predicts preference for musical consonance. *Neuropsychologia*, *58*, 23-32. https://doi.org/10.1016/j.neuropsychologia.2014.03.011.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

BrainVision Analyzer (Version 2.2.0) [Computer software]. (2019). Gilching, Germany: Brain Products GmbH.

Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: implications for developmental dyslexia. *Neuron*, *64*(3), 311-319.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236-246. https://doi.org/10.1111/j.1469-8986.2009.00928.x.

Chandrasekaran, B., Kraus, N., & Wong, P. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology, 107*, 1325e1336. https://doi.org/10.1152/jn.00923.2011.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456-465. https://doi.org/10.1121/1.3445785.

Clayards, M. (2018). Differences in cue weights for speech perception are correlated for individuals within and across contrasts. *J. Acoust. Soc. Am.*, *144*(3), EL172-EL177. https://doi.org/10.1121/1.5052025.

Coffey, E. B., Arseneau-Bruneau, I., Zhang, X., Baillet, S., & Zatorre, R. J. (2021). Oscillatory entrainment of the frequency-following response in auditory cortical and subcortical structures. *Journal of Neuroscience*, *41*(18), 4073-4087. https://doi.org/10.1523/JNEUROSCI.2313-20.2021.

Coffey, E. B., Colagrosso, E. M., Lehmann, A., Schönwiesner, M., & Zatorre, R. J. (2016a). Individual differences in the frequency-following response: relation to pitch perception. *PloS one*, *11*(3), e0152374. https://doi.org/10.1371/journal.pone.0152374.

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., & Zatorre, R. J. (2016b). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature communications*, *7*(1), 11070.

Coffey, E.B.J., Herholz, S.C., Scala, S. and Zatorre, R.J., 2011, June. Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research. In *The Neurosciences and Music IV: Learning and Memory, Conference. Edinburgh, UK*.

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., ... & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature communications*, *10*(1), 1-10. https://doi.org/10.1038/s41467-019-13003-w.

Conners, C. K., Epstein, J. N., Angold, A., & Klaric, J. (2003). Continuous performance test performance in a normative epidemiological sample. *Journal of abnormal child psychology*, *31*(5), 555-562. https://doi.org/10.1023/A:1025457300409.

Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on psychological science*, *3*(4), 286-300.

Destoky, F., Bertels, J., Niesen, M., Wens, V., Vander Ghinst, M., Leybaert, J., ... & Bourguignon, M. (2020). Cortical tracking of speech in noise accounts for reading strategies in children. *PLoS biology*, *18*(8), e3000840. https://doi.org/10.1371/journal.pbio.3000840.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470. https://doi.org/10.1006/jpho.1997.0052.

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language*, *215*, 104919. https://doi.org/10.1016/j.bandl.2021.104919.

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001213.

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, *6*(17), 2363-2367.

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning*, *62*, 79-109. https://doi.org/10.1111/j.1467-9922.2012.00707.x.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English/r/-/l/perception and production by adult Japanese speakers. In *Second Language Studies: Acquisition, Learning, Education and Technology*.

Honda, C. T., Clayards, M., & Baum, S. R. (2024). Exploring Individual Differences in Native Phonetic Perception and Their Link to Non-Native Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance, 50*(4), 370-394. https://doi.org/10.1037/xhp0001191

Hornickel, J., Knowles, E., & Kraus, N. (2012). Test-retest consistency of speech-evoked auditory brainstem responses in typically-developing children. *Hearing research*, *284*(1-2), 52-58. https://doi.org/10.1016/j.heares.2011.12.005.

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience*, *33*(8), 3500-3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013.

Jakoby, H., Goldstein, A., & Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology*, *48*(11), 1517-1531. https://doi.org/10.1111/j.1469-8986.2011.01227.x.

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of experimental child psychology*, *77*(1), 30-60. https://doi.org/10.1006/jecp.1999.2553.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and language*, *192*, 15-24. https://doi.org/10.1016/j.bandl.2019.02.004.

Kapnoula, E. C., & Samuel, A. G. (2023, May 31-June 2). *Examining the links between L1 phoneme categorization and non-native phonetic learning* [Poster presentation]. XVI International Symposium of Psycholinguistics, Vitoria-Gasteiz, Spain.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594. https://doi.org/10.1037/xhp0000410.

Keshavarzi, M., Mandke, K., Macfarlane, A., Parvez, L., Gabrielczyk, F., Wilson, A., ... & Goswami, U. (2022). Decoding of speech information using EEG in children with dyslexia: Less accurate low-frequency representations of speech, not “Noisy” representations. *Brain and Language*, *235*, 105198. https://doi.org/10.1016/j.bandl.2022.105198.

Kong, E. J., & Edwards, J. (2015, August). Individual differences in L2 learners' perceptual cue weighting patterns. *Proceedings of the International Congress of Phonetic Sciences*.

Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The frequency-following response: a window into human communication. In Kraus, N., Anderson, S., White-Schwoch, T., Fay, R., & Popper, A. (Eds.), *The frequency-following response* (pp. 1-15). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing research*, *166*(1-2), 192-201. https://doi.org/10.1016/S0378-5955(02)00327-1.

Krishnan, A., Bidelman, G. M., & Gandour, J. T. (2010). Neural representation of pitch salience in the human brainstem revealed by psychophysical and electrophysiological indices. *Hearing Research*, *268*(1-2), 60-66. https://doi.org/10.1016/j.heares.2010.04.016.

Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., & Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, *50*(12), 2849-2859. https://doi.org/10.1016/j.neuropsychologia.2012.08.013.

Krishnan, A., Swaminathan, J., & Gandour, J. T. (2009). Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. *J. Cogn. Neurosci.*, *21*(6), 1092-1105. https://doi.org/10.1162/jocn.2009.21077.

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing research*, *382*, 107779. https://doi.org/10.1016/j.heares.2019.107779.

Krizman, J., Marian, V., Shook, A., Skoe, E., & Kraus, N. (2012). Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. *Proceedings of the National Academy of Sciences*, *109*(20), 7877-7881. https://doi.org/10.1073/pnas.1201575109.

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. *Brain and Language, 128*, 34–40. https://doi.org/10.1016/j.bandl.2013.11.006.

Krizman, J., Slater, J., Skoe, E., Marian, V., & Kraus, N. (2015). Neural processing of speech in children is influenced by extent of bilingual experience. *Neuroscience letters*, *585*, 48-53. https://doi.org/10.1016/j.neulet.2014.11.011.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757-3768. https://doi.org/10.1121/1.3506351.

Li, P., Zhang, F., Tsai, E., & Puls, B. (2013). Language history questionnaire (lhq 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17,* 673-680. doi:10.1017/S1366728913000606.

Llanos, F., Xie, Z., & Chandrasekaran, B. (2017). Hidden Markov modeling of frequency-following responses to Mandarin lexical tones. *Journal of Neuroscience Methods*, *291*, 101-112. https://doi.org/10.1016/j.jneumeth.2017.08.010.

Llompart, M. (2020). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition*, 1-16. https://doi.org/10.1017/S1366728920000656.

Marmel, F., Linley, D., Carlyon, R. P., Gockel, H. E., Hopkins, K., & Plack, C. J. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *Journal of the Association for Research in Otolaryngology*, *14*(5), 757-766. https://doi.org/10.1007/s10162-013-0402-3.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279-297. https://doi.org/10.1017/S1366728909990022.

Mueller, J. L., Friederici, A. D., & Männel, C. (2012). Auditory perception at the root of language learning. *Proceedings of the National Academy of Sciences*, *109*(39), 15953-15958. https://doi.org/10.1073/pnas.1204319109.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894-15898. https://doi.org/10.1073/pnas.0701498104.

Neef, N. E., Müller, B., Liebig, J., Schaadt, G., Grigutsch, M., Gunter, T. C., ... & Friederici, A. D. (2017). Dyslexia risk gene relates to representation of sound in the auditory brainstem. *Developmental cognitive neuroscience*, *24*, 63-71. https://doi.org/10.1016/j.dcn.2017.01.008.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97-113. https://doi.org/10.1016/0028-3932(71)90067-4.

Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. *Cortex*, *93*, 146-154. https://doi.org/10.1016/j.cortex.2017.05.005.

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research*, *236*, 733-743. https://doi.org/10.1007/s00221-017-5164-4.

Ou, J., Xiang, M., & Yu, A. C. (2023). Individual variability in subcortical neural encoding shapes phonetic cue weighting. *Scientific Reports*, *13*(1), 9991.

Ou, J., & Yu, A. C. (2022). Neural correlates of individual differences in speech categorisation: evidence from subcortical, cortical, and behavioural measures. *Language, Cognition and Neuroscience*, *37*(3), 269-284. https://doi.org/10.1080/23273798.2021.1980594.

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical–cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience*, *33*(3), 549-557. https://doi.org/10.1111/j.1460-9568.2010.07546.x.

Parbery-Clark, A., Anderson, S., Hittner, E., & Kraus, N. (2012). Musical experience offsets age-related delays in neural timing. *Neurobiology of aging*, *33*(7), 1483-e1.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience*, *29*(45), 14100-14107. https://doi.org/10.1523/JNEUROSCI.3256-09.2009.

Parbery-Clark, A., Strait, D. L., Hittner, E., & Kraus, N. (2013). Musical training enhances neural processing of binaural sounds. *Journal of Neuroscience*, *33*(42), 16741-16747. https://doi.org/10.1523/JNEUROSCI.5700-12.2013.

Pereira, O., Gao, Y. A., & Toscano, J. C. (2018). Perceptual encoding of natural speech sounds revealed by the N1 event-related potential response. *Auditory Perception & Cognition*, *1*(1-2), 112-130.

Plyler, P.N., & Krishnan, A. (2001). Human frequency-following responses: Representation of second formant transitions in normal-hearing and hearing-impaired listeners. *Journal of the American Academy of Audiology, 12*, 523–533. https://doi.org/10.1055/s-0042-1745643.

Power, A. J., Colling, L. J., Mead, N., Barnes, L., & Goswami, U. (2016). Neural encoding of the speech envelope by children with developmental dyslexia. *Brain and Language*, *160*, 1-10. http://dx.doi.org/10.1016/j.bandl.2016.06.006.

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rauber, A. S., Escudero, P., Bion, R. A., & Baptista, B. O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. In *Ninth European Conference on Speech Communication and Technology*.

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology*, *28*(9), 1419-1427. https://doi.org/10.1016/j.cub.2018.03.026.

Ruggles, D., Bharadwaj, H., & Shinn-Cunningham, B. G. (2012). Why middle-aged listeners have trouble hearing in everyday settings. *Current Biology*, *22*(15), 1417-1422.

Saito, K., Sun, H., Kachlicka, M., Robert, J., Nakata, T., & Tierney, A. (2020). Domain-General Auditory Processing Explains Multiple Dimensions of Second Language Acquisition in Adulthood. *Studies in Second Language Acquisition*. https://doi.org/10.1017/S0272263120000467.

Saito, K., Sun, H., & Tierney, A. (2018). Explicit and implicit aptitude effects on second language speech learning: scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, 1-18. https://doi.org/10.1017/S1366728918000895.

Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of phonetics*, *52*, 183-204. https://doi.org/10.1016/j.wocn.2015.07.003.

Serniclaes, W., Sprenger-Charolles, L., Carré, R., & Demonet, J. F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/1092-4388(2001/032).

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition*, *98*(2), B35-B44. https://doi.org/10.1016/j.cognition.2005.03.002.

Shinn-Cunningham, B., Varghese, L., Wang, L., & Bharadwaj, H. (2017). Individual differences in temporal perception and their implications for everyday listening. In Kraus, N., Anderson, S., White-Schwoch, T., Fay, R., & Popper, A. (Eds.), *The frequency-following response* (pp. 159-192). Springer International Publishing. https://doi.org/10.1007/978-3-319-47944-6.

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*, 99-119. https://doi.org/10.1016/j.wocn.2015.03.001.

Silveira, R. (2011). Pronunciation instruction and syllabic-pattern discrimination. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, *27*(1), 21-36. https://doi.org/10.1590/S0102-44502011000100002.

Skoe, E., Burakiewicz, E., Figueiredo, M., & Hardin, M. (2017). Basic neural processing of sound in adults is influenced by bilingual experience. *Neuroscience*, *349*, 278-290. https://doi.org/10.1016/j.neuroscience.2017.02.049.

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and hearing*, *31*(3), 302. https://doi.org/10.1097/AUD.0b013e3181cdb272.

Skoe, E., & Kraus, N. (2013). Musical training heightens auditory brainstem function during sensitive periods in development. *Frontiers in psychology*, *4*, 622. https://doi.org/10.3389/fpsyg.2013.00622.

Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, *25*(6), 1415-1426. https://doi.org/10.1093/cercor/bht311.

Skoe, E., Krizman, J., & Kraus, N. (2013). The impoverished brain: disparities in maternal education affect the neural response to sound. *Journal of Neuroscience*, *33*(44), 17221-17231. https://doi.org/10.1523/JNEUROSCI.2102-13.2013.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter?. *Psychological science*, *17*(8), 675-681. https://doi.org/10.1111/j.1467-9280.2006.01765.x.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of cognitive neuroscience*, *20*(10), 1892-1902. https://doi.org/10.1162/jocn.2008.20131.

Song, J. H., Nicol, T., & Kraus, N. (2011a). Test–retest reliability of the speech-evoked auditory brainstem response. *Clinical Neurophysiology*, *122*(2), 346-355. https://doi.org/10.1016/j.clinph.2010.07.009.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011b). Perception of speech in noise: neural correlates. *Journal of cognitive neuroscience*, *23*(9), 2268-2279. https://doi.org/10.1162/jocn.2010.21556.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & psychophysics*, *36*(2), 131-145. https://doi.org/10.3758/bf03202673.

Tecoulesco, L., Skoe, E., & Naigles, L. R. (2020). Phonetic discrimination mediates the relationship between auditory brainstem response stability and syntactic performance. *Brain and Language*, *208*, 104810. https://doi.org/10.1016/j.bandl.2020.104810.

Thompson, E. C., Carr, K. W., White-Schwoch, T., Otto-Meyer, S., & Kraus, N. (2017). Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. *Hearing research*, *344*, 148-157. https://doi.org/10.1016/j.heares.2016.11.007.

Tichko, P., & Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hearing research*, *348*, 1-15. https://doi.org/10.1016/j.heares.2017.01.014.

Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. *Journal of Neuroscience, 33*(38), 14981-14988.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. J. (2010). Continuous perception and graded categorization: Electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological science*, *21*(10), 1532-1540.

Waschke, L., Kloosterman, N. A., Obleser, J., & Garrett, D. D. (2021). Behavior needs neural variability. *Neuron*, *109*(5), 751-766.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale–Fourth Edition: Technical and interactive manual*. San Antonio, TX: The Psychology Corporation.

White-Schwoch, T., Woodruff Carr, K., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., … & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS biology*, *13*(7), e1002196. https://doi.org/10.1371/journal.pbio.1002196.

Won, J., Tremblay, K., Clinard, C., Wright, R., Sagi, E., & Svirsky, M. (2016). The neural encoding of formant frequencies contributing to vowel identification in normal-hearing listeners. *The Journal of the Acoustical Society of America*, *139*(1), 1–11. https://doi.org/10.1121/1.4931909.

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*(4), 565-585. https://doi.org/10.1017/S0142716407070312.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience*, *10*(4), 420. https://doi.org/10.1038/nn1872.

Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, *79*(7), 2064-2072. https://doi.org/10.3758/s13414-017-1361-2.

Zhang, X., & Gong, Q. (2017). Correlation between the frequency difference limen and an index based on principal component analysis of the frequency-following response of normal hearing listeners. *Hearing Research*, *344*, 255-264. https://doi.org/10.1016/j.heares.2016.12.004.

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika*, *70*(1), 203-212. https://doi.org/10.1007/s11336-003-1119-8.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, *131*(1), 3. https://doi.org/10.1037/0033-2909.131.1.3.

## Supplemental Materials

As noted in the main text, all data and code from these experiments are publicly available on the OSF at https://osf.io/nec8t/?view_only=bb062dcdd4664e9bbd629a6101eeb672. The stimuli from the native phonetic perception tasks can also be accessed through the OSF page for Clayards (2018). In the following supplemental materials, we highlight a few relevant analyses which are part of the more extensive OSF files.

### Relating Behavioural Measures of Native Speech Perception

As described in the main text, canonical correlation analyses were used to relate individual differences in performance on the 2AFC and VAS native perception tasks. Further details about this analysis approach are available in Honda et al. (2024). The canonical correlation coefficients derived from these analyses are displayed in Table S.1. As anticipated, the correlation between 2AFC coefficients and VAS consistency was significant while the correlation between 2AFC coefficients and VAS slopes was not. The canonical coefficients (loadings of the original variables onto the canonical dimensions) are displayed in Table S.2.

Multivariate multiple regression was used to follow up on the canonical correlation analyses. This approach enabled us to predict all four 2AFC coefficients not only from the four VAS variables but also from two control predictors. The resulting multivariate multiple regression model produced regression tables from four separate regression models, fit with each 2AFC coefficient as the response (see supplemental R Markdown document for output). Multivariate tests (Type II MANOVA) were then employed to determine the significance of each predictor across the four models while accounting for the covariances between coefficients. The output of these multivariate tests is displayed in Table S.3. In line with our predictions, the only significant predictor was VAS dear-tear consistency.

**Table S.1**

*Canonical correlation coefficients and their significance for both correlations run*

| **Correlation 1 (2AFC coefficients and VAS slopes)** | | | | | |
|---|---|---|---|---|---|
| Dimension | Canonical Cor. | *F* | *df*1 | *df*2 | *p* |
| 1 | 0.35 | 1.14 | 8 | 100 | 0.35 |
| 2 | 0.20 | 0.71 | 3 | 51 | 0.55 |
| **Correlation 2 (2AFC coefficients and VAS consistency)** | | | | | |
| Dimension | Canonical Cor. | *F* | *df*1 | *df*2 | *P* |
| 1 | 0.80 | 8.77 | 8 | 100 | <0.001 |
| 2 | 0.23 | 0.92 | 3 | 51 | 0.44 |
| *n = 56* | | | | | |

**Table S.2**

*Canonical coefficients for each variable involved in the canonical correlation analyses*

| **Correlation 1 (2AFC coefficients and VAS slopes)** | | | **Correlation 2 (2AFC coefficients and VAS slopes)** | | |
|---|---|---|---|---|---|
| | Canonical Dimension | | | Canonical Dimension | |
| | 1 | 2 | | 1 | 2 |
| 2AFC variables | | | 2AFC variables | | |
| bet-bat acoustic cue A | 0.18 | 1.05 | bet-bat acoustic cue A | -0.46 | 1.17 |
| bet-bat acoustic cue B | -0.15 | -0.01 | bet-bat acoustic cue B | -0.32 | 0.25 |
| dear-tear acoustic cue A | -2.08 | -0.26 | dear-tear acoustic cue A | -0.81 | -1.82 |
| dear-tear acoustic cue B | 2.11 | 0.35 | dear-tear acoustic cue B | 0.47 | 0.46 |
| VAS slope variables | | | VAS consistency variables | | |
| bet-bat slope | -0.99 | 0.73 | bet-bat consistency | 0.22 | -2.03 |
| dear-tear slope | 0.91 | 0.70 | dear-tear consistency | 0.94 | 1.71 |

*Note*. These canonical coefficients are interpreted similarly to linear regression coefficients. For instance, for Correlation 1, a one-unit increase in bet-bat acoustic cue A corresponds to an increase of 0.18 in the first canonical variate for the set of 2AFC variables, when holding other variables constant.

**Table S.3**

*Summary of the multivariate multiple regression model predicting 2AFC coefficients*

| Predictor | Pillai's trace | *F* | Num *df* | Den *df* | *p* |
|---|---|---|---|---|---|
| VAS bet-bat slope | 0.01 | 0.07 | 4 | 40 | 0.99 |
| VAS dear-tear slope | 0.10 | 1.06 | 4 | 40 | 0.39 |
| VAS bet-bat consistency | 0.04 | 0.44 | 4 | 40 | 0.78 |
| VAS dear-tear consistency | 0.32 | 4.74 | 4 | 40 | 0.003 |
| AX-CPT bin score (attention) | 0.06 | 0.67 | 4 | 40 | 0.62 |
| Backwards digit span (memory) | 0.16 | 1.84 | 4 | 40 | 0.14 |
| *n = 50* | | | | | |

*Note.* Model equation: cbind(2AFC bet-bat acoustic cue A slope, 2AFC bet-bat acoustic cue B slope, 2AFC dear-tear acoustic cue A slope, 2AFC dear-tear acoustic cue B slope) ~ VAS bet-bat slope + VAS dear-tear slope + VAS bet-bat consistency + VAS dear-tear consistency + AX-CPT bin score + Backwards digit span.

**Relating Behavioural Measures of Native and Non-Native Speech Perception**

Multiple regression was used to determine whether our measures of native speech perception or control measures related to non-native speech perception ability as quantified by performance on the German oddity task. We first reduced the dimensionality of the native perception measures by running a principal component analysis (PCA) on the four 2AFC coefficients and another PCA on the four VAS measures. The first two components from each PCA, displayed in Table S.4, were then used as predictors in the multiple regression analysis. The output of the multiple regression model predicting non-native perception is displayed in Table S.5. The first principal component from the 2AFC coefficients (reflecting behavioural response consistency) and the AX-CPT bin scores (reflecting sustained attention; calculated based on bin scores described by Hughes et al., 2014) were significant predictors. As mentioned in the main text, an outlier participant was identified by calculating Cook's distance; the summary of an exploratory model without that participant is displayed in Table S.6. For the model without the outlier, there were no significant predictors.

**Table S.4**

*Correlations between the original 2AFC variables and the first two principal components extracted from them (left), and between the original VAS variables and the first two principal components extracted from them (right)*

| | PC1 | PC2 | | PC1 | PC2 |
|---|---|---|---|---|---|
| 2AFC bet-bat acoustic cue A | -0.51 | -0.20 | VAS bet-bat slope | -0.36 | -0.50 |
| 2AFC bet-bat acoustic cue B | -0.44 | -0.76 | VAS dear-tear slope | 0.01 | -0.85 |
| 2AFC dear-tear acoustic cue A | -0.54 | 0.32 | VAS bet-bat consistency | 0.66 | -0.09 |
| 2AFC dear-tear acoustic cue B | -0.51 | 0.52 | VAS dear-tear consistency | 0.66 | -0.18 |
| *Percent variance explained* | 75% | 16% | *Percent variance explained* | 49% | 28% |

**Table S.5**

*Summary of the multiple regression model predicting oddity A scores*

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *p* |
|---|---|---|---|---|
| (Intercept) | 0.11 | 0.06 | 1.98 | 0.05 |
| 2AFC principal comp. 1 | -0.12 | 0.05 | -2.45 | 0.02 |
| 2AFC principal comp. 2 | -0.06 | 0.07 | -0.77 | 0.44 |
| VAS principal comp. 1 | 0.07 | 0.06 | 1.12 | 0.27 |
| VAS principal comp. 2 | -0.01 | 0.06 | -0.24 | 0.81 |
| AX-CPT bin score (attention) | -0.12 | 0.06 | -2.02 | 0.05 |
| Backwards digit span (memory) | -0.10 | 0.06 | -1.55 | 0.13 |
| *Multiple $R^2$ = 0.26, Adjusted $R^2$ = 0.15, Residual SE = 0.36 (df = 40), n = 47* | | | | |

*Note.* Model equation: Oddity A score ~ 2AFC principal component 1 + 2AFC principal component 2 + VAS principal component 1 + VAS principal component 2 + AX-CPT bin score + Backwards digit span.

**Table S.6**

*Summary of the multiple regression model predicting oddity A scores (excluding an influential participant)*

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $p$ |
|---|---|---|---|---|
| (Intercept) | 0.17 | 0.03 | 6.56 | < 0.001 |
| 2AFC principal comp. 1 | -0.03 | 0.02 | -1.42 | 0.16 |
| 2AFC principal comp. 2 | -0.01 | 0.03 | -0.45 | 0.66 |
| VAS principal comp. 1 | -0.04 | 0.03 | -1.24 | 0.22 |
| VAS principal comp. 2 | -0.02 | 0.03 | -0.74 | 0.46 |
| AX-CPT bin score (attention) | 0.01 | 0.03 | 0.18 | 0.86 |
| Backwards digit span (memory) | 0.04 | 0.03 | 1.27 | 0.21 |
| *Multiple $R^2$ = 0.27, Adjusted $R^2$ = 0.15, Residual SE = 0.17 (df = 39), n = 46* | | | | |

*Note.* Model equation: Oddity A score ~ 2AFC principal component 1 + 2AFC principal component 2 + VAS principal component 1 + VAS principal component 2 + AX-CPT bin score + Backwards digit span.

**Supplemental References**

Honda, C. T., Clayards, M., & Baum, S. R. (2024). Exploring Individual Differences in Native Phonetic Perception and Their Link to Non-Native Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance, 50*(4), 370-394. https://doi.org/10.1037/xhp0001191

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior research methods*, *46*(3), 702-721. https://doi.org/10.3758/s13428-013-0411-5

## Link Between Chapters 2 and 3

In Chapters 1 and 2, we employed behavioural and neural measures to study how adults process and perceive speech sounds. This work provided important contributions to our understanding of individual differences in phonetic perception and of how different speech processing measures relate to one another. One notable outcome, in line with previous work, was that adults showed wide ranges of scores on a non-native perception task; yet even within this variability, most adults were not highly accurate in identifying the non-native sounds. Having characterized the individual differences and the general difficulties that adults show when trying to distinguish sounds in a new language, we were next interested in exploring ways of reducing such difficulties and of facilitating non-native speech sound learning.

Chapter 3 investigates the efficacy of a possible tool for aiding adults in acquiring new speech sounds: transcutaneous auricular vagus nerve stimulation (taVNS). A relatively novel neurostimulation technique, taVNS is increasingly being studied for its potential to improve a wide variety of outcomes, from mood to memory. In Chapter 3, we examine whether the administration of taVNS during non-native speech sound training can facilitate the learning of the trained sounds. It is worth noting that the first two chapters focussed solely on the *perception* of unfamiliar speech sounds, and yet the successful *production* of speech sounds is evidently also a key part of communicating effectively in a new language. Chapter 3 therefore extends our previous work by including measures of both perception and production. Furthermore, taVNS may have modulatory effects on mood and motivation, and an individual's motivation to learn a new language can play a role in the success of their language learning outcomes. Accordingly, Chapter 3 examines language learning motivation in addition to speech perception and production. Overall, while the first two chapters focussed on better understanding the basic nature of speech processing and perception, this final chapter describes more applied research whose goal was to determine whether a novel tool could enhance speech sound learning and learning-related motivation in adults.

# Chapter 3: No Clear Benefit of Transcutaneous Auricular Vagus Nerve Stimulation For Non-Native Speech Sound Learning

## Abstract

Learning to understand and speak a new language can be challenging and discouraging for adults. One potential tool for improving learning is transcutaneous auricular vagus nerve stimulation (taVNS), which modulates perception, memory, and attention systems. It has recently been reported that taVNS can improve English speakers' ability to perceive unfamiliar Mandarin tones. The current project explored the potential benefits of taVNS for language learning beyond tone perception. We studied adults' ability to perceive and produce unfamiliar speech sounds as well as any potential change in language learning motivation from pre- to post-training. Forty-five native English speakers were divided into three groups and were trained to perceive German sounds: one group received stimulation during easier-to-learn sounds (vowels), one group received stimulation during harder-to-learn sounds (fricatives), and a control group received no stimulation. We did not find evidence that taVNS improved perception or production of the German sounds, but there was evidence that it did improve some aspects of motivation. Specifically, the group that received taVNS during easier sounds showed a decrease in feelings of tension/pressure about language learning, while the other groups did not. Overall, the present study does not find that taVNS holds benefits for the acquisition of new speech sounds; however, the field is nascent, and so the potential applications of taVNS for language learning remain to be clarified.

**Keywords**: transcutaneous auricular vagus nerve stimulation, phonetic perception, non-native perception, speech perception training, phonetic production, language learning motivation

## Introduction

Many adult learners struggle to attain native-like performance across various measures of linguistic aptitude (e.g., Abrahamsson & Hyltenstam, 2009; Munro & Mann, 2005). Indeed, language learning outcomes tend to worsen as age of acquisition increases (Abrahamsson & Hyltenstam, 2009; Johnson & Newport, 1989; Kang & Guion, 2006; Pulvermüller & Schumann, 1994; Weber-Fox & Neville, 1996; White et al., 2013). One aspect of language acquisition that poses particular challenges for adult learners is the perception and production of new speech sounds (Díaz et al., 2012; Iverson et al., 2003). While infants implicitly learn to differentiate language sounds through unsupervised exposure (Maye et al., 2002), adults can benefit from explicit instruction in the form of supervised phonetic training paradigms (McCandliss et al., 2002; Iverson et al., 2005); and even after training, many adults still show relatively poor differentiation of new speech sounds in perception and production (Hanulíková et al., 2012; Strange & Dittman, 1984). The maturation of the brain has been argued to be a contributing factor in these age-related learning differences (see Stowe & Sabourin, 2005, for a review). One potential means of increasing the plasticity of the brain and improving learning in adulthood is through the use of neurostimulation techniques such as vagus nerve stimulation (VNS).

The vagus nerve is the longest cranial nerve in the body, reaching from the medulla down to the colon and innervating multiple organs along the way (Yuan & Silberstein, 2016a). The nerve's afferent fibers send sensory input to the vagal nuclei, which then pass the information along to various brain regions implicated in memory, perception, arousal, and affect, including the locus coeruleus, raphe nucleus, amygdala, thalamus, hippocampus, and nucleus accumbens (Berthoud & Neuhuber, 2000; Frangos et al., 2015; Sawchenko, 1983; Yuan & Silberstein, 2016a). By modifying the activity of the vagus nerve through stimulation, it is therefore possible to alter the activity of multiple brain areas and impact their associated functions (Frangos et al., 2015; Yuan & Silberstein, 2016b)—an approach that has advantages over other methods that only modulate localized neural activity (e.g., transcranial magnetic stimulation, transcranial direct current stimulation, or direct chemical stimulation; Bandler, 1969; Hallett, 2000; Thair et al., 2017). Indeed, in both animals and humans, VNS via an implanted electrode can improve memory (Clark et al., 1995; Clark et al., 1999; Ghacibeh et al., 2006; Sun et al., 2017), likely by modulating synaptic plasticity in the hippocampus (Zuo et al., 2007). Similarly, VNS has been shown to increase arousal and alertness as quantified by behavioural and neural measures (Collins et al., 2021; Rizzo et al., 2003). Such increases in arousal appear to be the result of enhanced excitatory activity in the locus coeruleus and other subcortical structures, leading to widespread activation throughout the cortex (Collins et al., 2021). In addition, VNS can improve positive affect by promoting the release of serotonin and noradrenaline from the raphe nucleus and locus coeruleus (Austelle et al., 2022; Elger et al., 2000). The positive effects of VNS on mood are further attested to by its approved use as a treatment for major depressive disorder (Austelle et al., 2022). In terms of modulating perception, rodent and human studies have also demonstrated that VNS can be paired with auditory stimuli to induce lasting, stimulus-specific plasticity in the auditory cortex (De Ridder et al., 2014; Engineer et al., 2015; Lai & David, 2021; Shetake et al., 2012).

Although VNS shows great promise as a neuromodulatory technique, it involves surgically implanting electrodes in the neck, rendering it invasive and inaccessible to the majority of the population. More recently, trancutaneous auricular vagus nerve stimulation or taVNS—a non-invasive counterpart to VNS—has been introduced as a similarly effective means of modulating neural activity (Frangos et al., 2015; Van Leusden et al., 2015; Yakunina et al., 2017). The auricular branch of the vagus nerve passes just under the skin of the outer ear (cymba concha, cymba cavum, external acoustic meatus, and tragus), and so taVNS can be administered in a straightforward and accessible way by placing electrodes against the ear (Badran et al., 2018; Butt et al., 2020; Frangos et al., 2015; Yakunina et al., 2017).

The brain regions affected by taVNS are similar to those affected by VNS; these include the locus coeruleus, raphe nucleus, amygdala, insula, thalamus, hippocampus, and nucleus accumbens (Badran et al., 2018; Frangos et al., 2015; Yakunina et al., 2017). As with VNS, taVNS has been found to modulate human perception and cognition. For example, studies have shown that taVNS can improve memory (Jacobs et al., 2015; Sun et al., 2021; Thakkar et al., 2023), arousal (Chen et al., 2023; Sharon et al., 2021), mood (Ferstl et al., 2022), tinnitus symptoms (Shim et al., 2015), and interoception (Villani et al., 2019), as well as decrease reaction times (Chen et al., 2021).

Learning a new language depends crucially on the ability to attend to and remember newly learned information. Thus, taVNS may hold promise for accelerating language learning due to its effects on arousal and memory. Indeed, preliminary evidence suggests that taVNS can enhance memory for spoken and written word lists under some conditions (Giraudier et al., 2020; Kaan et al., 2021), and can enhance reading skills in adults learning a new orthography (Thakkar et al., 2020). The potential for taVNS to aid language learning is also implied by work showing that arousal state modulates phonetic perception (Schuerman et al., 2022). Beyond its effects on arousal and memory, taVNS may enhance language learning by increasing auditory plasticity in the adult brain. As previously mentioned, invasive VNS can lead to stimulus-specific plasticity in the auditory cortex (De Ridder et al., 2014; Engineer et al., 2015; Shetake et al., 2012); and studies of event-related potentials have found that delivering taVNS during auditory perception tasks can enhance auditory preattentive processing (N1 amplitude; Rufener et al., 2023), selective attention (P3 amplitude; Rufener et al., 2018), and lexico-semantic encoding (N400 amplitude; Phillips et al., 2021), as well as decrease auditory processing time (P3 latency; Rufener et al., 2018). Perhaps taVNS therefore induces complementary benefits to arousal, memory, and auditory processing, which work in concert to facilitate the acquisition of a new language.

Important evidence that taVNS may accelerate language learning in adulthood comes from recent work by Llanos et al. (2020). The authors demonstrated that taVNS, in conjunction with perceptual training, can enhance native English speakers' ability to label certain Mandarin tones (Llanos et al., 2020). They divided participants into three groups: one that received stimulation during easier-to-perceive tones, one that received stimulation during harder-to-perceive tones, and a control group that received no stimulation. They found that taVNS specifically enhanced the labelling of easier-to-perceive tones for the group that was stimulated during those tones; this is likely because taVNS can increase arousal, and arousal specifically enhances memory for perceptually salient (i.e., easier-to-perceive) stimuli (Llanos et al., 2020; Mather et al., 2016; Mather & Sutherland, 2011). Nonetheless, a follow-up study by the same research group did not find any overall differences in tone learning performance between stimulated groups and a control group (McHaney et al., 2023). Exploratory analyses did reveal that participants' tone labelling accuracy increased faster for trials during which taVNS was administered compared to trials without taVNS, and that this effect was most pronounced when taVNS amplitude was low (McHaney et al., 2023). However, these effect sizes were small and there was no main effect of stimulation, so taVNS did not improve overall accuracy (McHaney et al., 2023). Around the same time, Pandža et al. (2020) investigated native English speakers' ability to associate meaning with Mandarin pseudowords that differed in lexical tone. They found that taVNS during Mandarin lexical tone training led to enhanced performance on a subsequent meaning recognition task (Pandža et al., 2020). Additionally, they found that taVNS before training was associated with greater decreases in reaction time on the recognition task and with more accurate performance on a recall task; however, these findings did not hold when taVNS was administered during (rather than before) training. Similar mixed results were obtained by Phillips et al. (2021): taVNS before or during Mandarin lexical tone training did not improve performance on a word learning task, but taVNS before training sped up reaction times on a recognition task, and taVNS during training led to improved recognition of mismatch trials on the recognition task. As such, taVNS appears to have some potential for enhancing lexical tone learning, but effects are not always consistently found; the extent of its efficacy and whether this may generalize beyond the learning of lexical tones remains unclear. Furthermore, while taVNS has shown some benefits for the *perception* of unfamiliar language sounds, it is not yet known whether those benefits extend to the *production* of unfamiliar language sounds.

Apart from facilitating language learning itself, taVNS may improve the subjective learning experience through its effects on mood and motivation. Calloway et al. (2020) found that participants who received taVNS prior to being trained on a language learning task showed greater reductions in negative affect from pre- to post-training compared to control participants who did not receive stimulation. Improvements in mood could subsequently impact motivation and learning outcomes, given that mood is a factor affecting learners' perceptions of success and failure during language learning (Williams et al., 2004). taVNS also seems to play a role in motivation, having been shown to increase adults' motivation to obtain food rewards (Neuser et al., 2020). Yet, to our knowledge, the potential impacts of taVNS on language learning motivation have not been investigated to date.

The current project had the broad aim of determining whether taVNS can enhance the learning of unfamiliar non-tonal speech sound contrasts. More specifically, we had 3 objectives: to determine whether taVNS during non-native speech perception training (1) enhances the *perception* of the trained sounds, (2) enhances the subsequent *production* of the same speech sounds, and (3) enhances *motivation* associated with language learning. The first objective was addressed by running a conceptual replication of Llanos et al. (2020)

using unfamiliar phonemic contrasts rather than lexical tones. In doing so, we hoped to clarify and extend the previous equivocal findings on taVNS and language learning. To this end, 45 native English speakers were trained on a perceptual labeling task for German front rounded vowels and fricatives. During training, 15 of the participants received stimulation paired with the vowels ("easier" phonemic contrast), 15 received stimulation paired with the fricatives ("harder" phonemic contrast), and 15 received no stimulation (control group). The second and third objectives involved examining the potential impacts of taVNS on elements of language learning that have thus far remained unexplored. To accomplish this, participants completed a German speech production task and a motivation questionnaire pre- and post-training. We anticipated that taVNS would enhance perception and production of the unfamiliar contrasts relative to the control group, and that its effects would be greatest for vowel learning since the vowel contrast is more perceptually salient than the consonant one. We also predicted that participants in the taVNS groups would show greater increases in language learning motivation from pre to post compared to controls.

## Materials and Methods

### Participants

Forty-five adults (33 females, 10 males, 2 preferring not to answer) were recruited from the Montreal area. This number was chosen based on the similar work of Llanos et al. (2020), who recruited 36 participants; they had 12 participants per group, and we decided to obtain 15 per group in order to try to replicate and extend their results. All participants identified as monolingual English speakers and were unfamiliar with German. Participants were aged 18-35 (mean: 23.0) with normal hearing thresholds in both ears as determined by an audiometric screening, and with no history of literacy, language, or cognitive impairments. People with medical implants, with metal braces, or who were pregnant were excluded for safety reasons. At the beginning of the experiment, participants were randomly assigned to one of three groups: taVNS-vowel (N = 15), taVNS-fricative (N = 15), and Control (N = 15). Participants signed an informed consent form and received monetary compensation ($40). The duration of the entire study was approximately 1.5 hours. The research protocol was approved by the Institutional Review Board of the Faculty of Medicine and Health Sciences of McGill University.

**Tasks**

***Demographic Information***

Participants completed a questionnaire about demographics, language history and proficiency, and musical experience, since these factors could influence speech processing. The questionnaire was adapted from the *Language history questionnaire* (LHQ 2.0; Li et al., 2013) and the *Montreal Music History Questionnaire* (MMHQ; Coffey et al., 2011). One-way ANOVAs confirmed that the extent of second language (L2) experience and musical experience did not differ significantly across groups. A summary of these ANOVAs and of demographic information for each group can be found in Supplemental Table 1.

***Non-Native Speech Perception Training***

To address our first objective, participants were trained to distinguish a German consonant contrast (palatal vs. postalveolar fricative; ç vs. ʃ) and a German vowel contrast (tense vs. lax high front rounded vowel; yː vs. ʏ) which are known to be perceptually challenging sounds for native English speakers (Mayr & Escudero, 2010). English speakers tend to perceive both German ç and ʃ as English ʃ (Moulton, 1962), whereas they tend to perceive German yː and ʏ as English uː and ʊ respectively (Mayr & Escudero, 2010; Strange et al., 2004). In line with Best's Perceptual Assimilation Model, which predicts that non-native sounds will be better discriminated when they are assimilated to two different native categories than when they are assimilated to a single native category (Best, 1991), previous work has shown that native English speakers perceive the German vowel contrast more accurately than the fricative one (Honda et al., 2024). The 10 German minimal pairs used in the training task are displayed in Table 1. To construct the stimuli, four native German speakers were recorded (two males, two females). The resulting sound files were edited to leave 20 ms before and after each production, and maximum amplitudes were normalized across speakers using GoldWave version 6.15 (GoldWave Inc., 2015). Each speaker produced each minimal pair once, resulting in a total of 80 speech stimuli (4 speakers x 2 contrasts x 10 words).

**Table 1**

*German minimal pairs used in the non-native speech perception training task*

| Consonant contrast | | Vowel contrast | |
|---|---|---|---|
| Palatal fricative (ç) | Postalveolar fricative (ʃ) | Tense high front rounded vowel (yː) | Lax high front rounded vowel (ʏ) |
| Fichte /fɪçtə/ | fischte /fɪʃtə/ | Brühl /bʁyːl/ | brüll /bʁʏl/ |
| Kirche /kɪəçə/ | Kirsche /kɪəʃə/ | Düne /dyːnə/ | dünne /dʏnə/ |
| Löchern /løçɪən/ | löschern /løʃɪən/ | fühlen /fyːlən/ | füllen /fʏlən/ |
| selig /zelɪç/ | seelisch /zelɪʃ/ | Hüte /hyːtə/ | Hütte /hʏtə/ |
| Wicht /vɪçt/ | wischt /vɪʃt/ | Wüste /vyːstə/ | wüsste /vʏstə/ |

The training procedure was based on that of Llanos et al. (2020), who used a forced-choice task to present stimuli in six training blocks and one generalization block. During training, half of our German stimuli (N = 40, from two speakers—one male, one female) were presented in six blocks, with each stimulus being presented once per block. On each trial, participants heard a stimulus and indicated which phoneme it contained by choosing between two options via mouse click (side of the screen counterbalanced across participants). The palatal vs. postalveolar fricatives were represented by the symbols "ç"/ "sh" and the tense vs. lax vowels were represented by the symbols "üː"/"ü" to facilitate learning without needing to teach participants the International Phonetic Alphabet. Visual feedback ("Correct"/"Incorrect") was provided immediately after each trial. Feedback lasted 1000 ms, and there were 500 ms between the end of feedback and the onset of the following stimulus. After the six training blocks, participants completed a Generalization block during which they labeled the other half of the stimuli (N = 40, from the other two speakers). There was no feedback or stimulation during the Generalization block. To avoid physical interference with the stimulation electrodes placed on the left ear, audio was delivered monaurally through the right ear with an insert earphone. The Training and Generalization blocks were programmed and presented using E-Prime 3.0.

***Electrical Stimulation Procedure***

Transcutaneous stimulation of the vagus nerve occurred during the perception training task. Replicating the procedure used in Llanos et al. (2020), stimulation was delivered through the cymba concha and cavum concha of the outer ear, at a level below each participant's perceptual threshold (as described further below). The participant's left ear was first cleaned with an alcohol swab. Next, silicon putty was molded to the shape of the participant's ear. The molded putty had an indentation across the middle caused by the crus of the helix, demarcating the cymba concha and cavum concha on either side. Two Ag-AgCl disc electrodes were then embedded in the putty in the centre of the areas corresponding to the cymba concha (cathode) and cavum concha (anode) and covered with conductive gel. Finally, the mold was pressed into place in the ear. The same experimenter performed the electrode setup for all participants to ensure maximal consistency in the procedure. Electrical stimulation was generated with a BIOPAC STMISOLA Constant Current Isolated Linear Stimulator. Consistent electrode contact was ensured by monitoring the stimulator's "Protect" light, which turns on only when contact is lost (this occurred on occasion during electrode setup and calibration, in which case the setup steps were repeated, but did not occur during training). Stimulation waveforms consisted of 14 biphasic square-wave pulses (150 μs pulse width), delivered at a rate of 25 Hz and with an amplitude no higher than 3 mA for safety reasons. The pulse train began at the onset of the auditory stimulus and continued for 560 ms. These stimulation parameters were selected based on Llanos et al. (2020) who found significant taVNS effects on speech sound learning using the same pulse width, frequency, and amplitude specifications, and they also closely resemble the parameters used in other work (e.g., Engineer et al., 2015; McHaney et al., 2023). Pulses were generated using Matlab (Mathworks, v. 2017a) and transmitted to the stimulator via a Measurement Computing USB-1208HS DAQ card.

Before the non-native speech perception training, each participant's perceptual threshold for the taVNS was identified through a calibration procedure. During calibration, individual pulse trains were delivered with the same parameters described above, starting at 0.1 mA and increasing in steps of 0.2 mA until the participant indicated feeling the stimulation. Amplitude was then decreased in steps of 0.1 mA until the participant no longer felt the stimulation. Each participant's threshold was recorded as the amplitude at which they could reliably begin to feel the stimulation across at least two repetitions of this procedure. During training, stimulation was delivered with a pulse amplitude 0.2 mA below the participant's perceptual threshold. There were no significant differences in pulse amplitude between the two groups that received stimulation (taVNS-vowel: M = 0.61 mA, SD = 0.35 mA; taVNS-fricative: M = 0.67 mA, SD = 0.45 mA; two-sample t-test: $t_{26} = -0.41$, $p = 0.69$). The control group underwent the same setup and threshold determination procedures so that all participants were blind to the condition to which they were assigned.

***Non-Native Speech Production Task***

To address our second objective, participants completed a non-native speech production task before and after the speech perception training. We wanted to test how well the perceptual training transferred to participants' general ability to produce the trained sounds accurately, and not simply their ability to imitate words they had been trained on. We therefore used a consonant-vowel-consonant (CVC) syllable production task and compared pre- and post-training productions. Each trial of the task involved a familiarization component followed by a production component. Participants were first familiarized with a non-native speech sound by hearing it presented in an isolated syllable (/çə/ and /ʃə/ for the fricatives, /y:l/ and /ʏl/ for the vowels). Speech sounds were produced by a third male native German speaker, different from the ones used in training. In order to reduce transfer of learning between the production task and the subsequent perceptual training task, we tried to maximise differences between the stimuli by using isolated syllables and a different voice (Baker & Trofimovich, 2006; Bradlow et al., 1997). The auditory presentation of the non-native speech sounds was accompanied by visual presentation of letters on a screen so that participants learned to associate each sound with a simplified version of its corresponding orthography (palatal vs. postalveolar fricatives represented by "ç"/ "sh" and tense vs. lax vowels represented by "üː"/"ü" as described in the training task). In this way, the production task also served to familiarize participants with the phonemes that would subsequently be trained. On each trial, after being familiarized with the native exemplar of the speech sound, participants were prompted to produce the sound within a CVC syllable. There were six different CVCs to produce for each sound (6 x 4 sounds = 24 productions total), and CVCs were written using English orthography combined with the letters that had been associated with the German sounds (ç, sh, üː, ü). Trials were blocked so that all six trials for a given sound appeared together, and block order was randomized. Table 2 displays the full list of CVCs used. Participants' productions were recorded using a headset microphone (Logitech, Switzerland). For each participant, all 24 productions were recorded before and after the training phase. The production task was programmed and presented using E-Prime 3.0.

**Table 2**

*German CVCs used in the non-native speech production task*

| Consonant contrast | | Vowel contrast | |
|---|---|---|---|
| Palatal fricative (ç) | Postalveolar fricative (ʃ) | Tense high front rounded vowel (yː) | Lax high front rounded vowel (ʏ) |
| kiç<br>/kɪç/ | kish<br>/kɪʃ/ | hüːn<br>/hyːn/ | hün<br>/hʏn/ |
| liç<br>/lɪç/ | lish<br>/lɪʃ/ | püːs<br>/pyːs/ | püs<br>/pʏs/ |
| geeç<br>/giç/ | geesh<br>/giʃ/ | füːm<br>/fyːm/ | füm<br>/fʏm/ |
| veeç<br>/viç/ | veesh<br>/viʃ/ | müːp<br>/myːp/ | müp<br>/mʏp/ |
| deç<br>/dɛç/ | desh<br>/dɛʃ/ | lüːm<br>/lyːm/ | lüm<br>/lʏm/ |
| tayç<br>/teɪç/ | taysh<br>/teɪʃ/ | küːt<br>/kyːt/ | küt<br>/kʏt/ |

Recordings of participants' productions were subsequently presented to three native German speakers. The native speakers completed ratings of the recordings at home on Gorilla Experiment Builder (www.gorilla.sc; Anwyl-Irvine et al., 2020), using their own headphones. Productions were presented in three sessions of approximately one hour each; all three sessions were completed within one week, with a minimum break of one hour between sessions to avoid fatigue. Within each session, productions from one third of participants (five per group) were randomly mixed, and productions from pre- and post-training were randomly mixed. For each production, the native speakers indicated which sound they heard via mouse click (2-alternative forced choice [2AFC] between the two sounds that make up the contrast). In addition, the three native speakers rated the quality of each production using a 7-point Likert scale (1 = poor, 7 = native-like). These quality ratings provided a more fine-grained measure of pronunciation ability. While acoustic analyses could also have been used to rate participants' productions, native speaker ratings were chosen because they provide a global accuracy measure that accounts for a variety of acoustic and articulatory dimensions which would be difficult to examine individually in isolation.

***Language Learning Motivation***

To address our third objective, we measured participants' motivation to learn foreign languages using a modified version of the Intrinsic Motivation Inventory (IMI: McAuley et al., 1989). Four items from three subscales (Interest/Enjoyment, Perceived Competence, Tension/Pressure) were used (following Saito, 2021), as shown in Table 3. Participants indicated how true the items were for them on a 7-point Likert scale (1 = not at all true, 7 = very true). Higher scores represent higher motivation for Interest/Enjoyment and Perceived Competence, while lower scores represent higher motivation for Tension/Pressure. The IMI was administered pre- and post-training.

**Table 3**

*Intrinsic Motivation Inventory (IMI), modified for the current study*

| Interest/Enjoyment |
|---|
| I enjoy foreign language exercises very much. |
| I think foreign language exercises are boring activities. (R) |
| I would describe foreign language exercises as very interesting. |
| Foreign language exercises do not hold my attention at all. (R) |
| Perceived competence |
| I think I am pretty good at learning foreign languages. |
| After practicing foreign languages for awhile, I feel pretty competent. |
| I am satisfied with my performance at learning foreign languages. |
| Foreign language exercises are an activity that I can’t do very well. (R) |
| Tension/Pressure |
| I feel very tense while practicing foreign languages. |
| I am very relaxed in practicing foreign languages. (R) |
| I do not feel nervous at all while practicing foreign languages. (R) |
| I feel pressured while practicing foreign languages. |

*Note.* An R after an item indicates a reverse item.

## Analyses and Results

The raw data and all code used to process and analyze it is publicly available on the OSF (https://osf.io/fdsaz/?view_only=d4f9ea6ef9804606b4972b3488981dc3). Details and output of all analyses described below can be found in the R Markdown document on the same OSF page. For all regression models reported in this experiment, the maximal random effects structure was used (Barr et al., 2013) unless the model failed to converge, in which case random effects were removed one at a time until model convergence was achieved. Note that in cases where our analyses were following previous work by Llanos et al. (2020) and McHaney et al. (2023), our maximal model is reported in the main text but an additional model with exactly the same structure as in the previous work was also fit for the sake of replication. The additional models’ output can be found in the R Markdown document. All analyses were carried out in R (R Development Core Team, 2010): mixed-effects logistic models were fit using the package *lme4* (Bates et al., 2015), linear mixed-effects models were fit using the package *lmertest* (Kuznetsova et al., 2017), and ordinal mixed-effects models were fit using the package *ordinal* (Christensen, 2022).

## Effects of Stimulation on Perception

### *Accuracy Improvement During Training – Analyses*

To assess the potential effects of taVNS on the learning of non-native sounds, a mixed-effects logistic regression model was fit similarly to that found in Llanos et al. (2020). The dependent variable was trial-level responses (correct/incorrect) for each participant during the training blocks. Fixed effects consisted of group (taVNS-vowel, taVNS-fricative, and Control = reference level), trial number (1 to 240; centered and divided by 2 SD), and contrast (fricative = 0.5, vowel = -0.5), along with all two- and three-way interactions among those three variables. For this analysis, the maximal model that converged included by-subject random intercepts and by-subject random slopes of trial number, contrast, and the interaction between trial number and contrast, without correlations between random effects. The group-by-trial interaction revealed whether the taVNS groups showed greater improvement over the course of training compared to the control group.

Following McHaney et al. (2023), another mixed-effects logistic regression model was fit to determine whether participants' performance during the training blocks depended on the amplitude of taVNS received or the type of trial (stimulated vs. unstimulated). The dependent variable was trial-level accuracy for participants in the two stimulation groups. There were fixed effects of trial type (stimulated = 0.5, unstimulated = -0.5), trial number, amplitude (centered and divided by 2 SD), and all two- and three-way interactions among these variables. Random effects consisted of by-subject and by-stimulus random intercepts, by-subject random slopes of trial type and trial number, and by-stimulus random slopes of trial number and amplitude, without correlations between random effects. A linear mixed-effects regression model was also fit predicting the retention of correct stimulus-response associations across blocks (as done in Llanos et al., 2020), to determine whether stimulation improved retention over time. The dependent variable for this model was the percentage of trials correctly labelled on both the current block and the previous block, starting at block 2. Fixed effects consisted of group, block (2-6; block 2 = reference level), and contrast, as well as all two- and three-way interactions among them. Random effects consisted of by-subject random intercepts and by-subject random slopes of contrast, without correlations between random effects.

***Accuracy Improvement During Training – Results***

Figure 1 displays accuracy over the course of training, at the individual and group level. We can see that accuracy improved for all three groups and for both speech sounds over the course of training. For the model predicting trial-level accuracy during training, there was a significant effect of trial number, indicating that the control group improved their performance over time across both contrasts ($\hat{\beta}$ = 0.619, $p$ < .001); we can conclude that training resulted in learning even when no stimulation was administered. No other significant effects were found; see Table 4 for a full model summary. Importantly, there were no significant group-by-trial number interactions; this demonstrates that although all three groups showed improved performance over time, the stimulation groups did not show greater improvement compared to the control group.

**Table 4**

*Summary of the mixed-effects logistic regression model predicting accuracy on the training task*

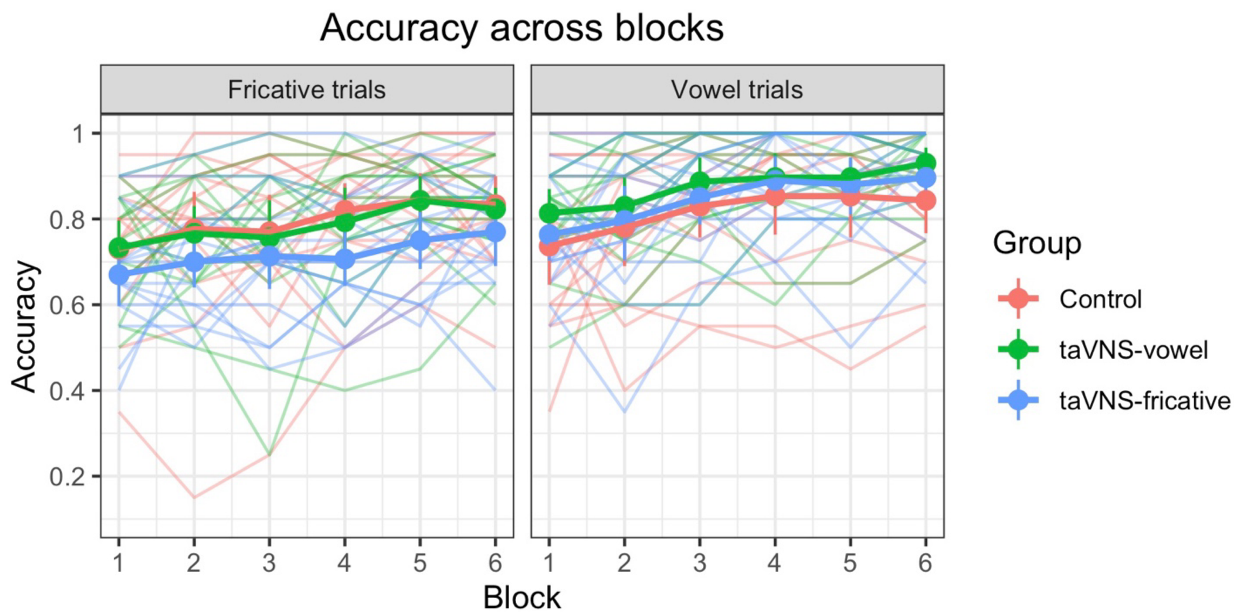
| **Fixed effects** | | | | |
|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $z$ | $p$ |
| Intercept | 1.723 | 0.200 | 8.614 | <.001 |
| Group taVNS-vowel | 0.167 | 0.283 | 0.590 | 0.555 |
| Group taVNS-fricative | -0.178 | 0.282 | -0.633 | 0.527 |
| Trial Number | 0.619 | 0.121 | 5.107 | <.001 |
| Contrast | -0.296 | 0.271 | -1.094 | 0.274 |
| Group taVNS-vowel:Trial Number | 0.073 | 0.173 | 0.424 | 0.672 |
| Group taVNS-fricative:Trial Number | -0.004 | 0.168 | -0.026 | 0.979 |
| Group taVNS-vowel:Contrast | -0.582 | 0.384 | -1.515 | 0.130 |
| Group taVNS-fricative:Contrast | -0.674 | 0.381 | -1.769 | 0.077 |
| Trial Number:Contrast | -0.062 | 0.218 | -0.286 | 0.775 |
| Group taVNS-vowel:Trial Number:Contrast | -0.36 | 0.315 | -1.141 | 0.254 |
| Group taVNS-fricative:Trial Number:Contrast | -0.461 | 0.304 | -1.515 | 0.130 |

| **Random effects** | | | |
|---|---|---|---|
| Group | Term | Variance | SD |
| Subject | Intercept | 0.555 | 0.745 |
| | Trial Number | 0.079 | 0.280 |
| | Contrast | 0.929 | 0.964 |
| | Trial Number:Contrast | 0.172 | 0.414 |

*Note.* Number of observations: 10800, groups: subject (45). *p*-values calculated using the Laplace approximation. Model equation: Accuracy ~ Group * Trial Number * Contrast + (1 + Trial Number * Contrast || Subject).

**Figure 1**

*Accuracy over time on the training task (from the first to the last block), for individuals and groups*




*Note.* Thin translucent lines represent individual participants' data, whereas thick solid lines represent aggregate group data. Vertical lines denote 95% confidence intervals for group accuracy on each block. Note the high levels of individual variability. Overall, accuracy improved over the course of training for all groups, and there were no significant group differences.

For the model including stimulation amplitude and trial type as fixed effects, there was again only a significant effect of trial number ($\hat{\beta}$ = 0.684, $p < .001$); performance on the perception task improved during training regardless of the stimulation condition, and these improvements in performance did not depend on stimulation amplitude. The lack of a significant main effect of trial type (stimulated vs. unstimulated) also indicates that stimulation failed to improve training performance. See Supplemental Table 2 for full model output. Supplemental Figure 1 shows the similar accuracy trajectory over time for stimulated and unstimulated trials. For the model predicting retention of correct stimulus-response associations, stimulation did not significantly predict retention rates. The only significant predictor was block (*p* = .011 for block 3 retention compared to block 2 retention, *p* < .005 for the other three blocks compared to block 2 retention), indicating better retention as training progressed (i.e., learning). See Supplemental Table 3 for model output and Supplemental Figure 2 for each group's retention rates across blocks. Overall, these analyses converge on the conclusion that performance on the training task improved over time but was not affected by stimulation.

***Reaction Times During Training – Analyses***

Given that taVNS has in some cases been shown to decrease reaction times (RTs; Chen et al., 2021; Pandža et al., 2020; Phillips et al., 2021) and increase post-error slowing (PES; Sellaro et al., 2015), analyses were run to determine the potential effects of stimulation on RTs and PES during training. For the RT analysis, trials on which participants responded incorrectly were removed (19% of trials) along with trials where RTs were < 200ms or > 2.5 SD above the participant's mean (a further 2% of trials), following Giannakopoulou et al. (2017). The distribution of raw RT values was positively skewed, so RTs were log transformed. A mixed-effects linear regression model was fit with the resulting cleaned and transformed RTs as the dependent variable. Fixed effects consisted of group, trial number, contrast, and all two- and three-way interactions between them. Random effects consisted of by-stimulus and by-subject random intercepts, as well as by-subject random slopes of trial number, contrast, and the interaction between trial number and contrast, without correlations between random effects. To calculate PES values, the RT on a trial following an error was subtracted from the RT on the trial preceding that error. A linear mixed-effects regression model was then fit predicting PES values, with group as a fixed effect and by-stimulus random intercepts as the maximal random effects structure promoting convergence and non-singularity.

***Reaction Times During Training – Results***

For the RT model, the only significant predictor was trial number ($\hat{\beta}$ = -233.656, $p$ < .001), indicating that participants in the control group became faster at responding to both contrasts as the task progressed. The lack of group-by-trial number interaction indicates that stimulation did not affect this decrease in reaction times. See Supplemental Table 4 for model output and Supplemental Figure 3 for each group's RTs plotted against trial number. For the PES model, no significant predictors were found; stimulation did not increase post-error slowing. See Supplemental Table 5 for model output and Supplemental Figure 4 for PES values per group.

***Accuracy During Generalization Block – Analyses***

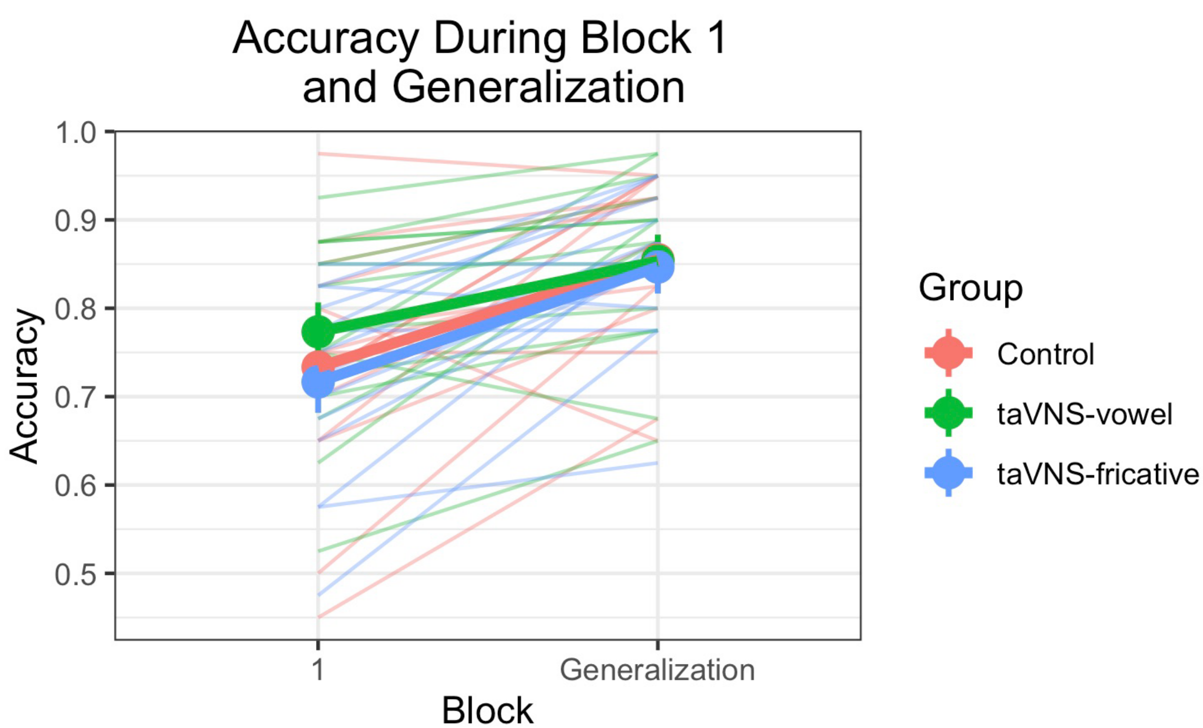
As in Llanos et al. (2020), a mixed-effects logistic regression model was fit with the dependent variable being trial-level accuracy in the generalization block and in block 1 of training. Fixed effects were group, block (generalization, block 1 = reference level), and contrast, along with all two- and three-way interactions between them. Random effects consisted of by-stimulus and by-subject random intercepts, along with by-subject random slopes of block, contrast, and the interaction between block and contrast. The group-by-block interaction enabled us to determine whether stimulation groups showed better generalization of their learning after accounting for baseline performance during block 1 of training.

***Accuracy During Generalization Block – Results***

Accuracy during block 1 and the generalization block is displayed in Figure 2. Table 5 shows the output of the model predicting accuracy during block 1 and the generalization block. There were no significant effects of group in block 1, demonstrating that groups did not differ in baseline performance across contrasts at the beginning of training. Block was the only significant predictor ($\hat{\beta}$ = 1.022, $p < .001$), indicating that performance was better during the generalization block than during block 1 of training—in other words, participants successfully learned and generalized their learning to new voices, regardless of group.

**Figure 2**

*Accuracy on block 1 of training and on the generalization block, for individuals and groups*




*Note.* Thin translucent lines represent individual participants' data, and thick solid lines represent aggregate group data. Vertical lines denote 95% confidence intervals for group accuracy on each block. Note the high levels of individual variability. Overall, accuracy improved from block 1 to the generalization block, and there were no significant group differences.

**Table 5**

*Summary of the mixed-effects logistic regression model predicting accuracy on block 1 and the generalization block*

**Fixed effects**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $z$ | $p$ |
|---|---|---|---|---|
| Intercept | 1.213 | 0.224 | 5.418 | <.001 |
| Group taVNS-vowel | 0.244 | 0.259 | 0.941 | 0.347 |
| Group taVNS-fricative | -0.115 | 0.256 | -0.449 | 0.654 |
| Block Generalization | 1.022 | 0.276 | 3.706 | <.001 |
| Contrast | -0.060 | 0.380 | -0.157 | 0.876 |
| Group taVNS-vowel:Block Generalization | -0.257 | 0.283 | -0.908 | 0.364 |
| Group taVNS-fricative: Block Generalization | -0.003 | 0.278 | -0.009 | 0.993 |
| Group taVNS-vowel:Contrast | -0.422 | 0.396 | -1.067 | 0.286 |
| Group taVNS-fricative:Contrast | -0.472 | 0.388 | -1.218 | 0.223 |
| Block Generalization:Contrast | 0.202 | 0.524 | 0.385 | 0.700 |
| Group taVNS-vowel:Block Generalization:Contrast | -0.005 | 0.513 | -0.010 | 0.992 |
| Group taVNS-fricative:Block Generalization:Contrast | 0.056 | 0.502 | 0.111 | 0.912 |

**Random effects**

| Group | Term | Variance | SD | Correlation | | |
|---|---|---|---|---|---|---|
| Stimulus | Intercept | 0.661 | 0.813 | | | |
| Subject | Intercept | 0.341 | 0.584 | | | |
| | Block Generalization | 0.196 | 0.443 | -0.15 | | |
| | Contrast | 0.528 | 0.727 | -0.21 | 0.23 | |
| | Block Generalization:Contrast | 0.368 | 0.607 | 0.32 | -0.24 | -0.48 |

Number of observations: 3600, groups: stimulus (80), subject (45). *p*-values calculated using the Laplace approximation. Model equation: Accuracy ~ Group * Block * Contrast + (1 + Block * Contrast | Subject) + (1 | Stimulus).

### *Effects of Stimulation on Production – Analyses*

Interrater reliability measures were obtained for the native German speakers' ratings of the production data. Fleiss' Kappa was calculated for the 2AFC ratings and the intraclass correlation coefficient (ICC; two-way random-effects model) was calculated for the Likert ratings (Gisev et al., 2013; Koo & Li, 2016). These calculations revealed acceptable reliability for both rating types (Fleiss' Kappa = 0.384; ICC = 0.594) based on established guidelines ( "fair" to "substantial" according to Landis & Koch, 1977).

For the 2AFC ratings, a mixed-effects logistic regression model was fit with trial-level accuracy as the dependent variable. Fixed effects were group, time (post, pre = reference level), contrast, and all two- and three-way interactions among them, as well as rater (first rater = reference level). Random effects consisted of by-subject random intercepts and by-subject random slopes of time, contrast, and the interaction between time and contrast, without correlations between random effects. The group-by-time interaction revealed whether stimulation groups showed greater increases in production accuracy from pre to post compared to the control group.

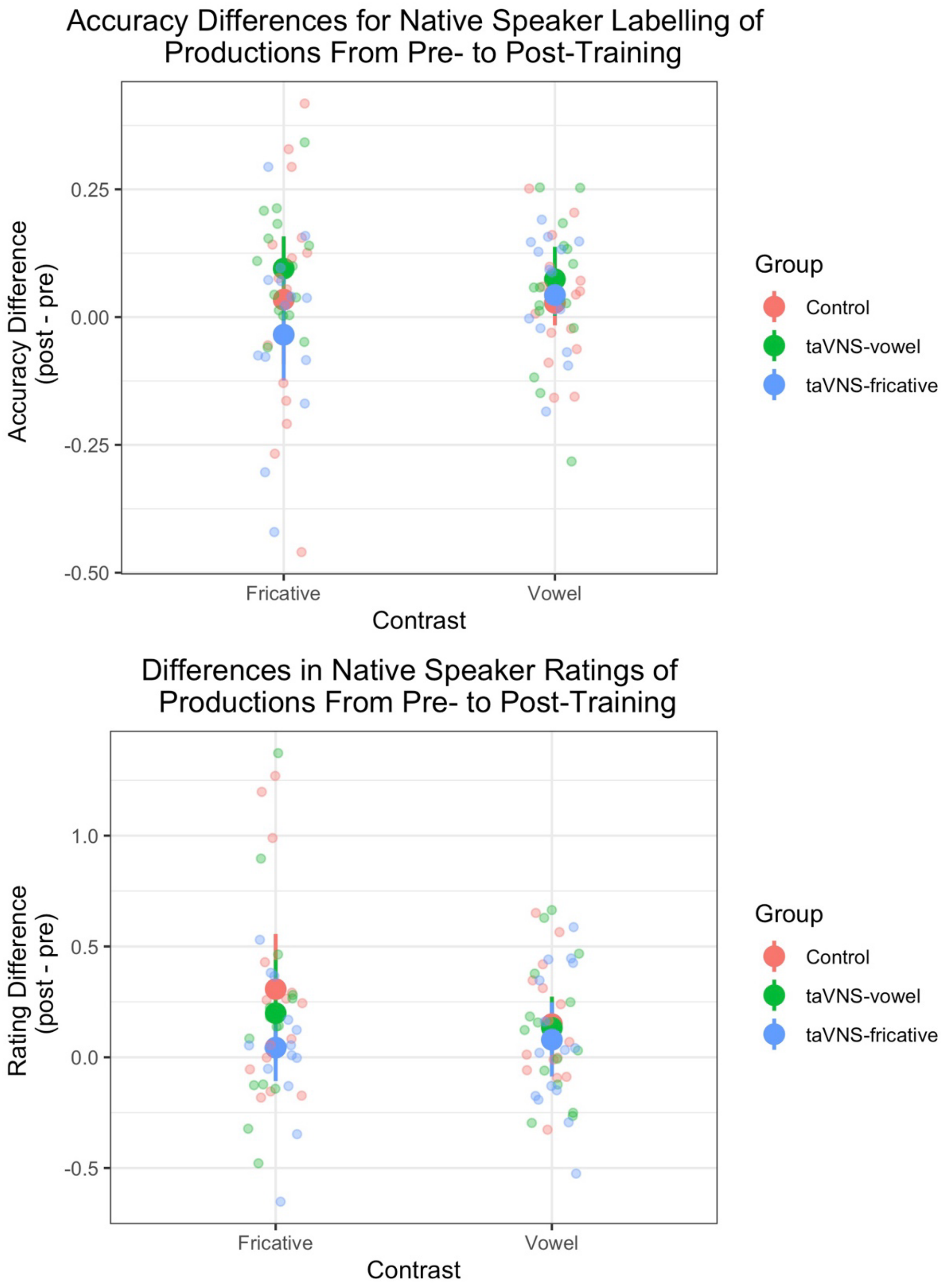
For the Likert ratings, an ordinal mixed-effects regression model was fit. Fixed effects were the same as for the model predicting 2AFC ratings. Random effects consisted of by-subject random intercepts and by-subject random slopes of time, contrast, and the interaction between time and contrast. The group-by-time interaction revealed whether native speakers' ratings of the productions increased more from pre to post for the stimulation groups compared to the control group.

***Effects of Stimulation on Production – Results***

Figure 3 shows the differences in ratings of productions from pre- to post-training, for the 2AFC ratings (A) and the Likert ratings (B). For the model predicting 2AFC ratings, rater was a significant predictor, with raters 2 and 3 tending to rate pre-training productions from participants in the control group more accurately compared to rater 1 ($\hat{\beta}$ = 0.162, $p$ = .014 for rater 2; $\hat{\beta}$ = 0.215, $p$ = .001 for rater 3). Table 6 displays the full output of the model. For the model predicting Likert ratings, rater was again a significant predictor, with rater 3 giving higher ratings than rater 1 ($\hat{\beta}$ = 0.785, $p$ < .001) for pre-training productions from control group participants. Contrast was also a significant predictor, with rater 1 giving higher ratings to fricatives than to vowels ($\hat{\beta}$ = 0.612, $p$ < .001). This effect was driven by high ratings for the fricative ʃ since it also occurs in English. Finally, time was also a significant predictor ($\hat{\beta}$ = 0.230, $p$ = .005) and did not interact with other predictors, revealing that ratings of production quality increased from pre- to post-training regardless of group. Full model output can be found in Table 7. Across both models, the lack of group effects or of group-by-time interactions suggests that the administration of taVNS during perceptual training did not specifically improve the subsequent production of the trained sounds.

**Figure 3**

*Native German speakers' ratings of participants' productions pre- and post-training*




*Note.* (A) Accuracy on a 2-alternative forced choice task. (B) Ratings on a 7-point Likert scale (1 = poor, 7 = native-like). Thin translucent lines represent individual participants' data, whereas thick solid lines represent aggregate group data. Vertical lines denote 95% confidence intervals. There was a significant increase in Likert ratings but not in accuracy judgments of productions from pre- to post-training, and there were no significant differences between groups.

**Table 6**

*Summary of the mixed-effects logistic regression model predicting accuracy of native speakers' ratings of participants' productions*

| **Fixed effects** | | | | |
|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *z* | *p* |
| Intercept | 0.432 | 0.117 | 3.697 | <.001 |
| Group taVNS-vowel | -0.027 | 0.157 | -0.174 | 0.862 |
| Group taVNS-fricative | 0.194 | 0.157 | 1.234 | 0.217 |
| Time Post | 0.19 | 0.124 | 1.532 | 0.125 |
| Contrast | 0.318 | 0.229 | 1.391 | 0.164 |
| Rater 2 | 0.162 | 0.066 | 2.454 | 0.014 |
| Rater 3 | 0.215 | 0.066 | 3.237 | 0.001 |
| Group taVNS-vowel:Time Post | 0.282 | 0.177 | 1.589 | 0.112 |
| Group taVNS-fricative:Time Post | -0.153 | 0.176 | -0.871 | 0.384 |
| Group taVNS-vowel:Contrast | 0.038 | 0.324 | 0.116 | 0.908 |
| Group taVNS-fricative:Contrast | 0.016 | 0.325 | 0.050 | 0.960 |
| Time Post:Contrast | 0.125 | 0.279 | 0.446 | 0.655 |
| Group taVNS-vowel:Time Post:Contrast | 0.165 | 0.398 | 0.415 | 0.678 |
| Group taVNS-fricative:Time Post:Contrast | -0.487 | 0.396 | -1.232 | 0.218 |

| **Random effects** | | | |
|---|---|---|---|
| Group | Term | Variance | SD |
| Subject | Intercept | 0.120 | 0.347 |
| | Time | 0.097 | 0.311 |
| | Contrast | 0.532 | 0.729 |
| | Time:Contrast | 0.632 | 0.795 |

*Note.* Number of observations: 6450, groups: subject (45). *p*-values calculated using the Laplace approximation. Model equation: Accuracy ~ Group * Time* Contrast + Rater + (1 + Time * Contrast || Subject).

**Table 7**

*Summary of the ordinal mixed-effects regression model predicting native speakers' Likert ratings of participants' productions*

**Fixed effects**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $z$ | $p$ |
|---|---|---|---|---|
| Intercept 1\|2 | -1.279 | 0.113 | -11.284 | <.001 |
| Intercept 2\|3 | -0.301 | 0.112 | -2.680 | 0.007 |
| Intercept 3\|4 | 0.418 | 0.113 | 3.712 | <.001 |
| Intercept 4\|5 | 1.150 | 0.113 | 10.142 | <.001 |
| Intercept 5\|6 | 1.922 | 0.115 | 16.756 | <.001 |
| Intercept 6\|7 | 3.187 | 0.121 | 26.362 | <.001 |
| Group taVNS-vowel | 0.147 | 0.150 | 0.979 | 0.327 |
| Group taVNS-fricative | 0.146 | 0.150 | 0.973 | 0.330 |
| Time Post | 0.230 | 0.081 | 2.824 | 0.005 |
| Contrast | 0.785 | 0.168 | 4.667 | <.001 |
| Rater 2 | -0.012 | 0.056 | -0.208 | 0.835 |
| Rater 3 | 0.612 | 0.054 | 11.360 | <.001 |
| Group taVNS-vowel:Time Post | -0.056 | 0.114 | -0.492 | 0.622 |
| Group taVNS-fricative:Time Post | -0.176 | 0.115 | -1.532 | 0.125 |
| Group taVNS-vowel:Contrast | -0.064 | 0.237 | -0.270 | 0.787 |
| Group taVNS-fricative:Contrast | -0.228 | 0.237 | -0.966 | 0.334 |
| Time Post:Contrast | 0.198 | 0.165 | 1.203 | 0.229 |
| Group taVNS-vowel:Time Post:Contrast | -0.102 | 0.232 | -0.438 | 0.661 |
| Group taVNS-fricative:Time Post:Contrast | -0.224 | 0.233 | -0.964 | 0.335 |

**Random effects**

| Group | Term | Variance | SD | Correlation | | |
|---|---|---|---|---|---|---|
| Subject | Intercept | 0.122 | 0.349 | | | |
| | Time | 0.008 | 0.087 | 0.922 | | |
| | Contrast | 0.232 | 0.482 | -0.228 | -0.588 | |
| | Time:Contrast | 0.040 | 0.201 | 0.888 | 0.640 | 0.245 |

Note. Number of observations: 6450, groups: subject (45). *p*-values calculated using the Laplace approximation. Model equation: Response ~ Group * Time * Contrast + Rater + (1 + Time * Contrast | Subject).

***Effects of Stimulation on Motivation – Analyses***

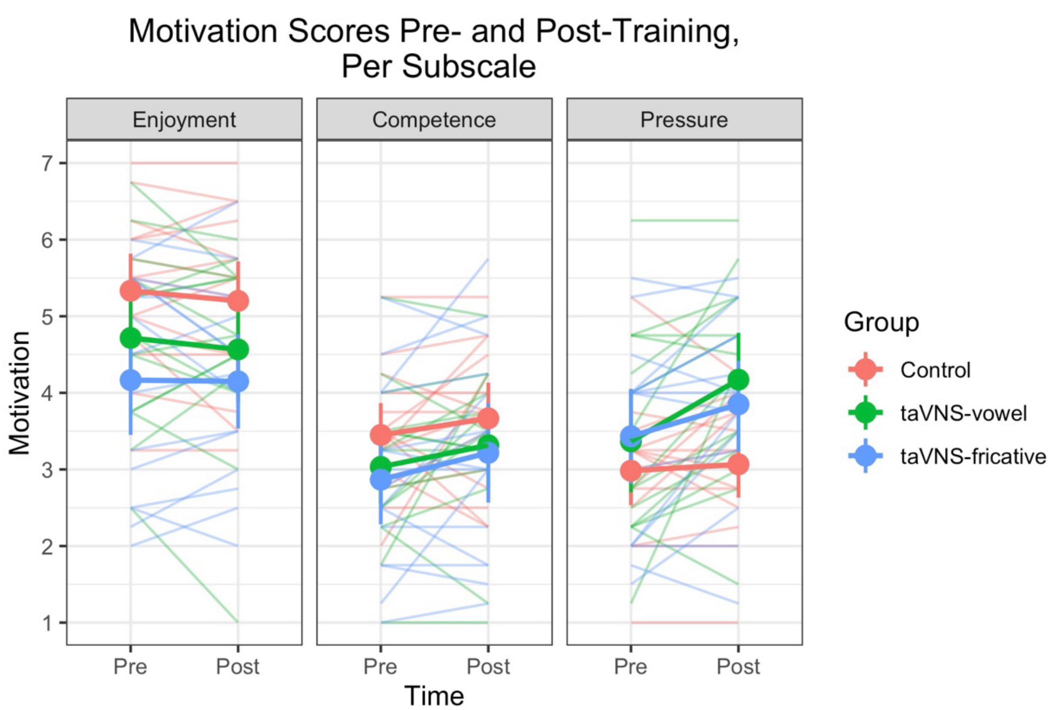
To determine the potential effects of taVNS on language learning motivation, ordinal mixed-effects models were fit as in Saito (2021). Fixed effects consisted of group, time (post, pre = reference level), and the interaction between the two, while random effects consisted of by-item random intercepts, by-subject random intercepts, and by-subject random slopes of time. The group-by-time interaction revealed whether stimulation groups showed greater motivation increases from pre to post compared to the control group. A first model was fit predicting responses to all items of the IMI. Three follow-up models were then fit predicting responses to each subscale: one model had an identical structure to the overall model, while the other two had only by-subject random intercepts to avoid singular fits. These models allowed us to investigate whether different aspects of motivation might be differentially affected by stimulation. Given that this follow-up analysis involved fitting more than one model to the same dataset, Bonferroni corrections were performed on the resulting p-values by dividing alpha by the number of comparisons being made (0.05/3 = 0.017).

***Effects of Stimulation on Motivation – Results***

Individual- and group-level motivation scores on each subscale pre- and post-training are displayed in Figure 4. As was the case for the perception and production measures, considerable individual variability in motivation scores was found. There were no significant predictors for the model predicting responses across all subscales (Supplemental Table 6). For the model predicting responses on the Tension/Pressure subscale, there was a group-by-time interaction such that the taVNS-vowel group showed greater increases in scores (indicating reduced feelings of tension and pressure) compared to the control group ($\hat{\beta}$ = 1.042, *p* = .029). Although this interaction was not significant after Bonferroni correction (*p* < .017), post-hoc pairwise comparisons between each group's scores pre- and post-training revealed that the taVNS-vowel group did show a significant difference from pre to post ($\hat{\beta}$ = -0.846, *p* < .001), which was not the case for the other groups. Stimulation during easier-to-learn sounds may therefore have decreased participants' feelings of tension and pressure associated with language learning. Output of the model and of the post-hoc comparisons can be found in Table 8.

**Figure 4**

*Motivation scores on each subscale from pre- to post-training, for individuals and groups*




*Note.* Higher scores indicate *increased* feelings of enjoyment and competence and *decreased* feelings of pressure. Thin translucent lines represent individual participants' data, whereas thick solid lines represent aggregate group data. Vertical lines denote 95% confidence intervals. Note again the high levels of individual variability. There were no differences between groups except on the pressure subscale, where the taVNS-vowel group showed a significant increase in scores while the other two groups did not.

**Table 8**

*Summary of the ordinal mixed-effects regression model and post-hoc pairwise comparisons predicting participants' motivation ratings on the pressure/tension subscale*

**Fixed effects**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *z* | *p* |
|---|---|---|---|---|
| Intercept 1\|2 | -2.074 | 0.496 | -4.185 | <.001 |
| Intercept 2\|3 | -0.359 | 0.473 | -0.759 | 0.448 |
| Intercept 3\|4 | 1.003 | 0.473 | 2.120 | 0.034 |
| Intercept 4\|5 | 2.003 | 0.481 | 4.165 | <.001 |
| Intercept 5\|6 | 3.250 | 0.503 | 6.464 | <.001 |
| Intercept 6\|7 | 4.781 | 0.564 | 8.473 | <.001 |
| Time Post | 0.069 | 0.337 | 0.205 | 0.837 |
| Group taVNS-vowel | 0.578 | 0.657 | 0.879 | 0.379 |
| Group taVNS-fricative | 0.545 | 0.656 | 0.831 | 0.406 |
| Time Post:Group taVNS-vowel | 1.042 | 0.478 | 2.181 | 0.029 |
| Time Post:Group taVNS-fricative | 0.560 | 0.477 | 1.174 | 0.240 |

**Random effects**

| Group | Term | Variance | SD |
|---|---|---|---|
| Subject | Intercept | 2.323 | 1.524 |

**Post-hoc pairwise comparisons**

| Contrast | $\hat{\beta}$ | $SE(\hat{\beta})$ | *z* ratio | *p* |
|---|---|---|---|---|
| Pre – Post (Control group) | -0.048 | 0.232 | -0.205 | 0.837 |
| Pre – Post (taVNS-vowel group) | -0.846 | 0.253 | -3.343 | <.001 |
| Pre – Post (taVNS-fricative group) | -0.472 | 0.252 | -1.875 | 0.061 |

*Note.* Number of observations: 358, groups: subject (45). *p*-values calculated using the Laplace approximation. Model equation: Rating ~ Time * Group + (1 | Subject).

## Discussion

The goal of the present study was to investigate whether the administration of taVNS during non-native speech perception training could enhance (1) participants' perception of the trained sounds, (2) their subsequent production of those same sounds, and (3) their language learning motivation. Native English speakers underwent training to perceive unfamiliar German vowels and fricatives, and were randomly assigned to one of three groups: taVNS-vowel (stimulation during easier-to-perceive sounds), taVNS-fricative (stimulation during harder-to-perceive sounds), or Control (no stimulation). Participants completed a German speech production task and a language learning motivation questionnaire before and after training. Contrary to our expectations, we did not find clear benefits of taVNS for perception, production, or motivation. However, it is possible that taVNS during training may have specifically improved language learning motivation related to pressure/tension. In particular, the taVNS-vowel group showed reduced feelings of pressure from pre- to post-training, indicating that stimulation during easier-to-perceive sounds may alleviate certain negative feelings associated with learning a new language in adulthood. The fact that taVNS was found to affect feelings of pressure but not the other motivation subscales measured (enjoyment and competence) is in line with previous work showing that taVNS prior to language learning tasks decreased negative affect but did not increase positive affect (Calloway et al., 2020). Nevertheless, as elaborated upon below, the potential benefits of taVNS for language learning and motivation remain largely uncertain; further research is called for to clarify and extend these findings.

**Lack of Evidence for taVNS-Related Improvement in Non-Native Phonetic Perception**

While we had hypothesized that taVNS might improve the perception of non-native phonemic contrasts, there are several potential reasons why this did not end up being the case. Given that music experience has been shown to predict the success of non-native sound learning (e.g., Perfors & Ong, 2012; Slevc & Miyake, 2006), one potential concern is that differences in our groups' musical backgrounds could have influenced our results. While an ANOVA revealed no significant differences between groups as mentioned earlier, the taVNS-fricative group had the fewest years of music training and the control group had the most (see Supplemental Table 1). To verify whether these numeric differences could be playing a role in the efficacy of stimulation, a mixed-effects logistic regression model was fit. The model had the same structure as the aforementioned model predicting accuracy during training, but included years of music training (centered and divided by 2 SD) as an additional fixed effect. As with the model that did not include music training, trial number was the strongest predictor, indicating learning over time ($\hat{\beta}$ = 5.195, $p$ < .001); and music training was a significant predictor ($\hat{\beta}$ = 2.376, $p$ = .018), pointing to the benefits of music experience for speech sound learning (see R Markdown document on the OSF page for full model output). However, no significant main effects of group or group-by-trial interactions were found; even after accounting for music training, taVNS did not improve overall accuracy or accelerate learning across trials.

An additional concern is that participants generally showed quite accurate performance on the training task, even from the beginning (see Figure 1). It is possible that stimulation-related effects were not observed because there was limited room for improvement in performance. To partly alleviate this concern, an additional mixed-effects logistic regression model was fit with the same structure as the main model predicting accuracy during training, but with block 1 accuracy (centered and divided by 2 SD, and excluding accuracy on block 1 trials) as an additional fixed effect. As with the other training models, trial number was the strongest predictor ($\hat{\beta}$ = 3.939, $p < .001$), and there were no significant main effects of group or group-by-trial interactions (see R Markdown document). Thus, although many participants had limited room for improvement during training, it does not appear that any potential effects of stimulation depended on initial accuracy.

Having addressed these concerns through additional analyses, we may turn to other possible explanations for our null findings. Previous work with taVNS and speech sound learning has focussed only on tonal contrasts (Llanos et al., 2020; McHaney et al., 2023; Pandža et al., 2020; Phillips et al., 2021). It is possible that taVNS does not benefit the learning of phonemic contrasts in the same way that it may benefit the learning of tonal contrasts. Evidence suggests that during the processing of novel phonemic contrasts, learners show neural activity in the same regions of the left hemisphere that are activated by native phonemic contrasts (Golestani & Zatorre, 2004). In contrast, during the processing of novel tonal contrasts, learners show neural activity in regions of the right hemisphere that are commonly associated with nonlinguistic pitch processing (Hsieh et al., 2001). Perhaps our results differ from those of previous taVNS and speech sound learning studies due to differences in how novel tonal versus phonemic contrasts are processed. Furthermore, it is likely that native English speakers show different perceptual assimilation patterns for the phonemic contrasts used in the current study compared to the tonal contrasts used in previous work (So & Best, 2014; Strange et al., 2004); differences in assimilation could affect the discriminability of the contrasts (Best & Tyler, 2007), leading to differing effects of taVNS on the acquisition of tonal versus phonemic contrasts. Relatedly, in the current work it was only feasible to test participants with one language background on their acquisition of one subset of non-native phonemic contrasts. Future research could test participants with a variety of backgrounds and demographic characteristics on their acquisition of different speech sounds in order to determine whether any potential taVNS effects may be moderated by factors such as participant language experience, age, or the relationship between the native and non-native languages. Another interesting avenue for future work would be to examine how taVNS affects measures of linguistic and non-linguistic memory and attention, both in the auditory and visual domains (e.g., pairing taVNS with reading or sign language training as well as with other non-linguistic cognitive tasks), in order to disentangle the specificity of stimulation-related effects and determine the optimal situations for taVNS use. It will be important for researchers to carefully design the methodology of future experiments with a view to detecting and differentiating the particular mechanism(s) of taVNS-related improvement that are expected to be at play (e.g., enhancements to auditory plasticity vs. arousal vs. memory).

It should also be noted that although the present study was conceived as a conceptual replication of previous work on tonal contrast learning (Llanos et al., 2020), some differences in study design may account in part for discrepancies between our findings and those of prior studies. For instance, in Llanos et al. (2020), stimulation began 300 ms prior to stimulus onset, and stimuli consisted of single syllables in which the trained speech sound reliably occurred in the same position. In the present work, mono- and disyllabic stimuli were used in which the trained speech sound could occur on either the first or the second syllable; this variability in the position of the trained sound may have reduced the salience of our non-native contrasts. Differences in stimulation timing and in stimulus salience may therefore in part explain some of the disparities between our findings and those of Llanos et al. (2020). Nonetheless, participants' overall high accuracy on our training task (as mentioned above) suggests that the non-native contrasts were salient enough to be effectively acquired. Beyond this, our training task provided participants with two response options, whereas Llanos et al. (2020) provided participants with four response options. As such, participants were more likely to respond correctly by chance in our study, which may have reduced our ability to detect taVNS-related improvements in perception because accuracy scores during training were less variable. Note, however, that Figures 1 and 2 show widespread variability in performance all the same. Finally, in the current work and in Llanos et al. (2020), a single training session was administered; in contrast, in Pandža et al. (2020) and Phillips et al. (2021), two training sessions were administered on separate days. Perhaps taVNS is more effective when paired with speech sound training that spans more than a single session and that includes an opportunity for memory consolidation in between. This being said, given that even the two-day training paradigms used in previous work bear limited resemblance to true language learning which occurs over long timescales, longitudinal work is needed to assess the effects of taVNS on more naturalistic learning.

It is plausible that taVNS does hold potential for improving speech sound learning, but that the optimal stimulation parameters for language learning have not yet been identified. Administration of taVNS entails the selection of various parameters including stimulation amplitude, pulse width, frequency, duration, and timing relative to stimulus presentation. Other work with taVNS has employed a wide range of parameters. For example, stimulation amplitude has sometimes been much higher than in the present study (> 4 mA; Jacobs et al., 2015; Kaan et al., 2021; Liu et al., 2018); frequency has ranged from 5 Hz (Thakkar et al., 2020) to 300 Hz (Phillips et al., 2021); stimulation has sometimes been delivered before (Calloway et al., 2020), during (Llanos et al., 2020; McHaney et al., 2023), or after (Clark, 1999) a learning task; stimulation has been administered in short targeted bursts (< 1 s duration; De Ridder et al., 2014; Engineer et al., 2015; Llanos et al., 2020; McHaney et al., 2023) as well as continuously over an extended time period (> 10 min duration; Calloway et al., 2020; Kaan et al., 2021; Ventura-Bort et al., 2018); and stimulation electrodes have been placed on the cymba concha and cymba cavum (Llanos et al., 2020) or on the outer ear canal (Phillips et al., 2021). While we selected parameters following previous work on taVNS and language learning (Llanos et al., 2020), the field remains nascent and more research is needed to identify the most appropriate taVNS parameters for different desired outcomes. Future work could systematically compare the effects of different taVNS parameters on language learning in order to clarify whether stimulation has benefits and whether certain parameters may be more effective than others.

Another possibility is that taVNS may not be a reliable means of improving the learning of new speech sound contrasts, even tonal ones. The studies that have examined taVNS and tonal contrast learning to date have shown mixed outcomes, as reviewed in the introduction. The present study aimed to conceptually replicate the work of Llanos et al.

(2020), who reported faster rates of Mandarin tone learning for participants receiving taVNS compared to controls. However, the same research group recently published a partial replication of their own work in which they did not find that tone learning rates differed significantly by experimental group (McHaney et al., 2023). Their additional exploratory analyses revealed only modest effects whereby taVNS at lower amplitudes initially increased learning rates during the training task, but without increasing overall accuracy on the task (McHaney et al., 2023). The other existing work on taVNS and tone learning comes from a research group that has administered taVNS under two different conditions (before vs. during Mandarin lexical tone training) and has tested a variety of outcome measures including accuracy and reaction times on word learning, lexical recognition, and recall tasks (Pandža et al., 2020; Phillips et al., 2021). The researchers only found stimulation-related benefits for a few of the many possible combinations of conditions and outcome measures. For example, Pandža et al. (2020) found that taVNS during (but not before) training was significantly associated with greater accuracy (but not decreased reaction times) on certain trials of the recognition task (but not the recall task). In a similar vein, Phillips et al. (2021) found that taVNS before (but not during) training was significantly associated with decreased reaction times (but not increased accuracy) on the recognition task (but not the word learning task). If taVNS were an effective method for improving tone learning, its effects might be expected to emerge in a more robust and uniform way across conditions and studies. These results also suggest that future work should carefully consider and compare different experimental conditions and outcome measures, since effects may depend on the nature of the stimulation and the tasks being administered.

Although we did not find specific benefits of taVNS for non-native speech sound perception, it is worth noting that our analyses converged on the conclusion that participants did in fact improve their perception of the non-native sounds over the course of training, regardless of group. While learning to perceive unfamiliar phonemic contrasts in adulthood is often difficult, it is clear that supervised training paradigms such as this one can facilitate the learning process, as has also been found in previous work (e.g., Bradlow et al., 1997; Giannakopoulou et al., 2017; Iverson et al., 2005; Reetzke et al., 2018). Regardless of the effectiveness of taVNS, it will be fruitful for future research to continue to investigate the optimal paradigms for training non-native perception in adults.

**Lack of Evidence for taVNS-Related Improvement in Non-Native Phonetic Production**

As with non-native perception, no group effects were observed for our non-native production measures; taVNS did not specifically improve production from pre- to post-training. This outcome is perhaps not surprising given that no other studies to date have yet investigated taVNS and non-native production, and that the previous findings about taVNS and non-native perception have been mixed. In the present study, as in prior work, stimulation was delivered during perceptual training. Accordingly, any stimulation-related effects would be anticipated to emerge most notably for perceptual outcome measures; considering that no such effects on perception were found, it is unsurprising that no effects on production were found either.

As discussed above, the perceptual training task resulted in overall improvement in non-native perception regardless of experimental group. On the other hand, improvement in non-native production was less clear. Native speaker ratings of participants' productions did not improve pre- to post-training for the ratings involving a forced choice between the two sounds making up a contrast. For the more fine-grained measure where productions were rated on a seven-point scale from “poor” to “native-like”, statistically significant improvement was found, though the size of the effect was not large (see Figure 3B). Our findings are in line with Sakai and Moorman’s (2018) recent meta-analysis of the effects of perceptual training on non-native production. The authors found that perceptual training resulted in medium-sized effects on perception outcomes and in small effects on production outcomes (Sakai & Moorman, 2018). As such, non-native perception and production are understood to be linked, but perceptual training does not necessarily lead to reliable or significant improvements in production. It should also be noted that our production task differed in format from the perception training, which may account in part for the lack of strong improvement in production post-training. The training task involved listening to non-native words (Table 1), whereas the production task involved hearing and seeing isolated non-native phonemes as exemplars and then producing CVCs (Table 2). These differences arose because the production task was designed so that participants would not directly repeat or imitate the exemplar and so that the stimuli would be feasible to produce for inexperienced learners. In future studies with participants who have greater non-native language experience, perception and production tasks could be made more similar in order to specifically examine the effects of perception training on production. For instance, the same words could be used as stimuli during the perceptual training task and the production task (e.g., Brosseau-Lapré et al., 2013). Since our task did not measure spontaneous speech production, additional work will also be needed to determine the relationship between our training paradigm and more naturalistic production measures.

**Evidence for taVNS-Related Improvement in Language Learning Motivation**

When looking across all items of our motivation questionnaire, we did not find effects of taVNS on language learning motivation. However, when focussing on the items belonging to the tension/pressure subscale of the questionnaire, we did find an effect: from pre- to post-training, the taVNS-vowel group showed a significant decrease in feelings of tension and pressure associated with language learning, which was not the case for the other two groups. Recall that the taVNS-vowel group received stimulation during easier-to-perceive (vowel) sounds. Llanos et al. (2020) also found taVNS effects specifically for the group stimulated during easier-to-perceive non-native sounds—this group showed enhanced learning over the course of non-native perception training.  The authors argued that this finding emerged because taVNS increases arousal, and such modulation of arousal can specifically enhance memory consolidation for more perceptually salient stimuli (Llanos et al., 2020). While we did not find enhanced learning for the taVNS-vowel group, the fact that the group's language learning motivation increased could similarly relate to the perceptual saliency of the stimuli. Perhaps participants naturally felt more capable and relaxed when responding to the vowel trials on the training task because the vowel contrast was more perceptually salient, and so the administration of taVNS during those trials served as a reinforcement signal that modulated neural activity related to affect and reward, in turn leading to decreased feelings of tension post-training.

Prior work supports a role for taVNS in decreasing feelings of tension associated with language learning. There is preliminary evidence that taVNS may improve fear extinction after a fear conditioning task (Burger et al., 2017) and reduce spontaneous negative thoughts after a worry induction task (Burger et al., 2019). The technique may therefore have the potential to lessen participants' overall feelings of fear and of worry. taVNS has additionally been found to increase participants' confidence in their ability to perform a task successfully (Villani et al., 2019). In the particular context of language learning, administration of taVNS prior to a second language learning task has in some cases been demonstrated to reduce negative affect and anxiety (Calloway et al., 2020). All of these findings point to taVNS as a possible means of reducing the stress and tension that can be felt by adults during the language learning process.

At the neural level, these positive effects make sense given that taVNS has been shown to modulate the activity of various brain regions and networks involved in affect and motivation, including the locus coeruleus, raphe nucleus, and limbic system (Badran et al., 2018; Frangos et al., 2015; Yakunina et al., 2017). In clinical contexts, VNS is known to promote the release of serotonin from the raphe nucleus, leading to improved mood (Austelle et al., 2022). The reductions in feelings of pressure and tension observed in our taVNS-vowel group may be attributable in part to such changes in neural activity and in neurotransmitter release. Note, however, that this explanation remains speculative; it was beyond the scope of the current study to measure neural activity patterns or neurotransmitter levels. Future studies could consider including such additional measures to untangle the potential mechanisms whereby taVNS increases language learning motivation. It is also not entirely clear why taVNS would selectively reduce feelings of pressure/tension without affecting the other subscales of motivation measured here (namely, interest/enjoyment and perceived competence). More research is needed to clarify the generalizability of our findings.

In conjunction with language aptitude, motivation is known to be an important factor in predicting language learning outcomes (e.g., Dörnyei, 2001; Gardner, 2000; Saito et al., 2018). Individuals who are motivated—for example, who are willing to expend effort in learning a language, who want to achieve a high level of competence in the language, and who have favourable attitudes towards the learning situation—tend to have greater non-native language achievement (Masgoret & Gardner, 2003). A meta-analysis found that, across different ages and learning environments, the correlation between motivation and second language achievement ranges from around .29 to .39 depending on the particular measure of achievement (Masgoret & Gardner, 2003). This effect size is considered small to medium based on Plonsky and Oswald's (2014) conventions for second language research, or medium to large based on Gignac and Szodorai's (2016) conventions for individual differences research. As such, the role of motivation in language acquisition is non-negligible, and if taVNS truly does impact motivation then this could have important repercussions for adult learners who are struggling to acquire a new language. Future work with taVNS could examine the construct of language learning motivation in greater detail, employing more extensive measures of motivation associated with conceptual frameworks such as the L2 motivational self system (Dörnyei, 2009) and the socio-educational model of second language acquisition (Gardner, 2000).

## Conclusion

In sum, we examined the potential effects of taVNS on the perception and production of non-native phonemic contrasts and on language learning motivation. taVNS had previously shown positive (but inconsistent) effects on the perception of non-native tonal contrasts (Llanos et al., 2020; McHaney et al., 2023; Pandža et al., 2020; Phillips et al., 2021), and so we sought to determine whether such effects might extend to the perception of phonemic (non-tonal) contrasts. To our knowledge, this is also the first time that taVNS has been investigated in relation to non-native phonetic production or language learning motivation. Overall, no clear effects of taVNS on non-native perception or production emerged. Our results hint at a potential benefit of taVNS for language learning motivation—in particular, stimulation during the learning of easier-to-perceive sounds may decrease feelings of tension and pressure associated with language learning. Nevertheless, taVNS did not increase overall motivation across the three subscales of our motivation questionnaire or across our two stimulated groups, so its efficacy is not clear. On the whole, while taVNS is a promising technique with a multitude of potential applications, from treatment of epilepsy (Liu et al., 2018) to relief of tinnitus symptoms (Shim et al., 2015), its usefulness in the context of language learning remains to be determined. Research with taVNS is still just beginning to emerge, and the stimulation parameters and outcome measures used in previous work have been heterogeneous; going forward, it will be important to systematically compare a variety of stimulation conditions and language acquisition outcomes in an effort to more conclusively determine any possible uses of taVNS in language learning contexts. Improving both language acquisition and motivation is an especially important endeavour given the plurality of adults who are now learning new languages in our diverse and globalized world.

### Conflict of Interest

NB is the co-founder of Revai working on developing taVNS for cognitive enhancement. While she assisted with designing the study as outlined under the Author contributions section, she was not involved in analyzing or visualizing the data.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

### Author Contributions

CTH was responsible for conceptualization, methodology, investigation, formal analysis, visualization, and writing the original draft of the article. NB, MC, and SRB were responsible for conceptualization, methodology, supervision, and reviewing and editing the article.

### Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a grant awarded by the Natural Sciences and Engineering Research Council of Canada (NSERC) to SB, a grant awarded by the Social Sciences and Humanities Research Council (SSHRC) to MC, and an

NSERC Postgraduate Scholarship- Doctoral (PGS D) grant along with a Mitacs Accelerate Award awarded to CH.

### Data Availability Statement

The original contributions presented in the study are included in the article/Supplemental material, further inquiries can be directed to the corresponding author.

### Supplementary Material

The Supplemental Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2024.1403080/full#supplementary-material

## References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language learning*, *59*(2), 249-306. https://doi.org/10.1111/j.1467-9922.2009.00507.x

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: an online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Austelle, C. W., O'Leary, G. H., Thompson, S., Gruber, E., Kahn, A., Manett, A. J., ... & Badran, B. W. (2022). A comprehensive review of vagus nerve stimulation for depression. *Neuromodulation: Technology at the Neural Interface*, *25*(3), 309-315. https://doi.org/10.1111/ner.13528

Badran, B. W., Dowdle, L. T., Mithoefer, O. J., LaBate, N. T., Coatsworth, J., Brown, J. C., ... & George, M. S. (2018). Neurophysiologic effects of transcutaneous auricular vagus nerve stimulation (taVNS) via electrical stimulation of the tragus: a concurrent taVNS/fMRI study and review. *Brain stimulation*, *11*(3), 492-500. https://doi.org/10.1016/j.brs.2017.12.009

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of l2 vowels: the role of individual differences. *International Review of Applied Linguistics in Language Teaching*, *44*(3), 231–250. https://doi.org/10.1515/IRAL.2006.010

Bandler, R. J. (1969). Facilitation of aggressive behaviour in rat by direct cholinergic stimulation of the hypothalamus. *Nature*, *224*(5223), 1035-1036.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software,67*(1). https://doi.org/10.18637/jss.v067.i01

Berthoud, H. R., & Neuhuber, W. L. (2000). Functional and chemical anatomy of the afferent vagal system. *Autonomic Neuroscience*, *85*(1-3), 1-17.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, *167*(224), 233-277.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, *101*(4), 2299-2310.

Brosseau-Lapré, F., Rvachew, S., Clayards, M., & Dickson, D. (2013). Stimulus variability and perceptual learning of nonnative vowel categories. *Applied Psycholinguistics*, *34*(3), 419-441.

Burger, A. M., Van der Does, W., Thayer, J. F., Brosschot, J. F., & Verkuil, B. (2019). Transcutaneous vagus nerve stimulation reduces spontaneous but not induced negative thought intrusions in high worriers. *Biological Psychology*, *142*, 80-89. https://doi.org/10.1016/j.biopsycho.2019.01.014

Burger, A. M., Verkuil, B., Fenlon, H., Thijs, L., Cools, L., Miller, H. C., ... & Van Diest, I. (2017). Mixed evidence for the potential of non-invasive transcutaneous vagal nerve stimulation to improve the extinction and retention of fear. *Behaviour Research and Therapy*, *97*, 64-74. https://doi.org/10.1016/j.brat.2017.07.005

Butt, M. F., Albusoda, A., Farmer, A. D., & Aziz, Q. (2020). The anatomical basis for transcutaneous auricular vagus nerve stimulation. *Journal of anatomy*, *236*(4), 588-611.

Calloway, R., Karuzis, V., Tseng, A., Martinez, D., & O'Rourke, P. (2020). Auricular Transcutaneous Vagus Nerve Stimulation (tVNS) Affects Mood and Anxiety during Second Language Learning. In *CogSci*.

Chen, Y., Lu, X., & Hu, L. (2023). Transcutaneous auricular vagus nerve stimulation facilitates cortical arousal and alertness. *International Journal of Environmental Research and Public Health*, *20*(2), 1402. https://doi.org/10.3390/ijerph20021402

Chen, L., Zhang, J., Wang, Z., Zhang, X., Zhang, L., Xu, M., ... & Ming, D. (2021). Effects of transcutaneous vagus nerve stimulation (tVNS) on action planning: A behavioural and EEG study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 1675-1683. https://doi.org/10.1109/TNSRE.2021.3131497

Christensen, R. H. B. (2022). ordinal - Regression Models for Ordinal Data. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal

Clark, K. B., Krahl, S. E., Smith, D. C., & Jensen, R. A. (1995). Post-training unilateral vagal stimulation enhances retention performance in the rat. *Neurobiology of learning and memory*, *63*(3), 213-216.

Clark, K. B., Naritoku, D. K., Smith, D. C., Browning, R. A., & Jensen, R. A. (1999). Enhanced recognition memory following vagus nerve stimulation in human subjects. *Nature neuroscience*, *2*(1), 94-98.

Coffey, E.B.J., Herholz, S.C., Scala, S. and Zatorre, R.J., 2011, June. Montreal Music History Questionnaire: a tool for the assessment of music-related experience in music cognition research. In *The Neurosciences and Music IV: Learning and Memory, Conference. Edinburgh, UK*.

Collins, L., Boddington, L., Steffan, P. J., & McCormick, D. (2021). Vagus nerve stimulation induces widespread cortical and behavioral activation. *Current Biology*, *31*(10), 2088-2098. https://doi.org/10.1016/j.cub.2021.02.049

De Ridder, D., Vanneste, S., Engineer, N. D., & Kilgard, M. P. (2014). Safety and efficacy of vagus nerve stimulation paired with tones for the treatment of tinnitus: a case series. *Neuromodulation: Technology at the Neural Interface*, *17*(2), 170-179. https://doi.org/10.1111/ner.12127

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680-689.

Dörnyei, Z. (2009). The L2 motivational self system. In Dörnyei, Z., & Ushioda, E. (Eds.), *Motivation, language identity and the L2 self*. Multilingual Matters.

Dörnyei, Z. (2001). New themes and approaches in second language motivation research. *Annual Review of Applied Linguistics*, *21*, 43–59. doi:10.1017/S0267190501000034

Elger, G., Hoppe, C., Falkai, P., Rush, A. J., & Elger, C. E. (2000). Vagus nerve stimulation is associated with mood improvements in epilepsy patients. *Epilepsy research*, *42*(2-3), 203-210. https://doi.org/10.1016/S0920-1211(00)00181-9

Engineer, C. T., Engineer, N. D., Riley, J. R., Seale, J. D., & Kilgard, M. P. (2015). Pairing speech sounds with vagus nerve stimulation drives stimulus-specific cortical plasticity. *Brain stimulation*, *8*(3), 637-644. https://doi.org/10.1016/j.brs.2015.01.408

Ferstl, M., Teckentrup, V., Lin, W. M., Kräutlein, F., Kühnel, A., Klaus, J., ... & Kroemer, N. B. (2022). Non-invasive vagus nerve stimulation boosts mood recovery after effort exertion. *Psychological Medicine*, *52*(14), 3029-3039. https://doi.org/10.1017/S0033291720005073

Frangos, E., Ellrich, J., & Komisaruk, B. R. (2015). Non-invasive access to the vagus nerve central projections via electrical stimulation of the external ear: fMRI evidence in humans. *Brain stimulation*, *8*(3), 624-636. https://doi.org/10.1016/j.brs.2014.11.018

Gardner, R. C. (2000). Correlation, causation, motivation, and second language acquisition. *Canadian Psychology/Psychologie Canadienne*, *41*(1), 10. https://doi.org/10.1037/h0086854

Ghacibeh, G. A., Shenker, J. I., Shenal, B., Uthman, B. M., & Heilman, K. M. (2006). The influence of vagus nerve stimulation on memory. *Cognitive and behavioral neurology*, *19*(3), 119-122. https://doi.org/10.1097/01.wnn.0000213908.34278.7d

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209. https://doi.org/10.7717/peerj.3209

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and individual differences*, *102*, 74-78. https://doi.org/10.1016/j.paid.2016.06.069

Giraudier, M., Ventura-Bort, C., & Weymar, M. (2020). Transcutaneous vagus nerve stimulation (tVNS) improves high-confidence recognition memory but not emotional word processing. *Frontiers in Psychology*, *11*, 1276. https://doi.org/10.3389/fpsyg.2020.01276

Gisev, N., Bell, J. S., & Chen, T. F. (2013). Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, *9*(3), 330-338. https://doi.org/10.1016/j.sapharm.2012.04.004

Golestani, N., & Zatorre, R. J. (2004). Learning new sounds of speech: reallocation of neural substrates. *Neuroimage*, *21*(2), 494-506. https://doi.org/10.1016/j.neuroimage.2003.09.071

Hallett, M. (2000). Transcranial magnetic stimulation and the human brain. *Nature*, *406*(6792), 147-150. https://doi.org/10.1038/35018000

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning*, *62*, 79-109. https://doi.org/10.1111/j.1467-9922.2012.00707.x

Honda, C. T., Clayards, M., & Baum, S. R. (2024). Exploring Individual Differences in Native Phonetic Perception and Their Link to Nonnative Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance.* Advance online publication. https://doi.org/10.1037/xhp0001191

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and language*, *76*(3), 227-252. https://doi.org/10.1006/brln.2000.2382

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278. https://doi.org/10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57. https://doi.org/10.1016/S0010-0277(02)00198-1

Jacobs, H. I., Riphagen, J. M., Razat, C. M., Wiese, S., & Sack, A. T. (2015). Transcutaneous vagus nerve stimulation boosts associative memory in older individuals. *Neurobiology of Aging*, *36*(5), 1860-1867. https://doi.org/10.1016/j.neurobiolaging.2015.02.023

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99.

Kaan, E., De Aguiar, I., Clarke, C., Lamb, D. G., Williamson, J. B., & Porges, E. C. (2021). A transcutaneous vagus nerve stimulation study on verbal order memory. *Journal of Neurolinguistics*, *59*, 100990. https://doi.org/10.1016/j.jneuroling.2021.100990

Kang, K. H., & Guion, S. G. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *The Journal of the Acoustical Society of America*, *119*(3), 1672-1683. https://doi.org/10.1121/1.2166607

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1-26. https://doi.org/10.18637/jss.v082.i13

Lai, J., & David, S. V. (2021). Short-term effects of vagus nerve stimulation on learning and evoked activity in auditory cortex. *eNeuro*, *8*(3). https://doi.org/10.1523/ENEURO.0522-20.2021

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159-174. https://doi.org/10.2307/2529310

Li, P., Zhang, F., Tsai, E., & Puls, B. (2013). Language history questionnaire (lhq 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition, 17,* 673-680. https://doi.org/10.1017/S1366728913000606

Liu, A., Rong, P., Gong, L., Song, L., Wang, X., Li, L., & Wang, Y. (2018). Efficacy and safety of treatment with transcutaneous vagus nerve stimulation in 17 patients with refractory epilepsy evaluated by electroencephalogram, seizure frequency, and quality of life. *Medical science monitor: international medical journal of experimental and clinical research*, *24*, 8439-8448. https://doi.org/10.12659/MSM.910689

Llanos, F., McHaney, J. R., Schuerman, W. L., Han, G. Y., Leonard, M. K., & Chandrasekaran, B. (2020). Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults. *NPJ science of learning, 5*(1), 1-11. https://doi.org/10.1038/s41539-020-0070-0

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language learning*, *53*(S1), 167-210. https://doi.org/10.1111/1467-9922.00227

Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X15000667

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on psychological science*, *6*(2), 114-133. https://doi.org/10.1177/1745691611400234

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101-B111. https://doi.org/10.1016/S0010-0277(01)00157-3

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279-297. https://doi.org/10.1017/S1366728909990022

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, *60*(1), 48-58. https://doi.org/10.1080/02701367.1989.10607413

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 89-108. https://doi.org/10.3758/cabn.2.2.89

McHaney, J. R., Schuerman, W. L., Leonard, M. K., & Chandrasekaran, B. (2023). Transcutaneous Auricular Vagus Nerve Stimulation Modulates Performance but Not Pupil Size During Nonnative Speech Category Learning. *Journal of Speech, Language, and Hearing Research*, 1-19. https://doi.org/10.1044/2023_JSLHR-22-00596

Moulton, W. G. (1962). *The sounds of English and German*. Chicago, IL: University of Chicago Press.

Munro, M., & Mann, V. (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics*, *26*(3), 311-341. https://doi.org/10.1017/S0142716405050198

Neuser, M. P., Teckentrup, V., Kühnel, A., Hallschmid, M., Walter, M., & Kroemer, N. B. (2020). Vagus nerve stimulation boosts the drive to work for rewards. *Nature communications*, *11*(1), 3555. https://doi.org/10.1038/s41467-020-17344-9

Pandža, N. B., Phillips, I., Karuzis, V. P., O'Rourke, P., & Kuchinsky, S. E. (2020). Neurostimulation and pupillometry: New directions for learning and research in applied linguistics. *Annual Review of Applied Linguistics*, *40*, 56-77. https://doi.org/10.1017/S0267190520000069

Perfors, A., & Ong, J. (2012). Musicians are better at learning non-native sound contrasts even in non-tonal languages. Proceedings of the 34th Annual Meeting of the Cognitive Science Society, held in Sapporo, Japan, 1-4 August, 2012.

Phillips, I., Calloway, R. C., Karuzis, V. P., Pandža, N. B., O'Rourke, P., & Kuchinsky, S. E. (2021). Transcutaneous auricular vagus nerve stimulation strengthens semantic representations of foreign language tone words during initial stages of learning. *Journal of cognitive neuroscience*, *34*(1), 127-152. https://doi.org/10.1162/jocn_a_01783

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language learning*, *64*(4), 878-912. https://doi.org/10.1111/lang.12079

Pulvermüller, F., & Schumann, J. H. (1994). Neurobiological mechanisms of language acquisition. *Language learning*, *44*(4), 681-734. https://doi.org/10.1111/j.1467-1770.1994.tb00635.x

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology*, *28*(9), 1419-1427. https://doi.org/10.1016/j.cub.2018.03.026

Rizzo, P., Beelke, M., De Carli, F., Canovaro, P., Nobili, L., Robert, A., ... & Ferrillo, F. (2003). Chronic vagus nerve stimulation improves alertness and reduces rapid eye movement sleep in patients affected by refractory epilepsy. *Sleep*, *26*(5), 607-611. https://doi.org/10.1093/sleep/26.5.607

Rufener, K. S., Geyer, U., Janitzky, K., Heinze, H. J., & Zaehle, T. (2018). Modulating auditory selective attention by non-invasive brain stimulation: Differential effects of transcutaneous vagal nerve stimulation and transcranial random noise stimulation. *European Journal of Neuroscience*, *48*(6), 2301-2309. https://doi.org/10.1111/ejn.14128

Rufener, K. S., Wienke, C., Salanje, A., Haghikia, A., & Zaehle, T. (2023). Effects of transcutaneous auricular vagus nerve stimulation paired with tones on electrophysiological markers of auditory perception. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2023.06.006

Saito, H. (2021). *The effects of frequency and type of feedback on L2 phoneme learning* (Doctoral dissertation, McGill University).

Saito, K., Dewaele, J. M., Abe, M., & In'nami, Y. (2018). Motivation, emotion, learning experience, and second language comprehensibility development in classroom settings: A cross-sectional and longitudinal study. *Language Learning*, *68*(3), 709-743. https://doi.org/10.1111/lang.12297

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187-224. https://doi.org/10.1017/S0142716417000418

Sawchenko, P. E. (1983). Central connections of the sensory and motor nuclei of the vagus nerve. *Journal of the autonomic nervous system*, *9*(1), 13-26. https://doi.org/10.1016/0165-1838(83)90129-7

Schuerman, W. L., Chandrasekaran, B., & Leonard, M. K. (2022). Arousal states as a key source of variability in speech perception and learning. *Languages*, *7*(1), 19. https://doi.org/10.3390/languages7010019

Sellaro, R., van Leusden, J. W., Tona, K. D., Verkuil, B., Nieuwenhuis, S., & Colzato, L. S. (2015). Transcutaneous vagus nerve stimulation enhances post-error slowing. *Journal of cognitive neuroscience*, *27*(11), 2126-2132. https://doi.org/10.1162/jocn_a_00851

Sharon, O., Fahoum, F., & Nir, Y. (2021). Transcutaneous vagus nerve stimulation in humans induces pupil dilation and attenuates alpha oscillations. *Journal of Neuroscience*, *41*(2), 320-330. https://doi.org/10.1523/JNEUROSCI.1361-20.2020

Shetake, J. A., Engineer, N. D., Vrana, W. A., Wolf, J. T., & Kilgard, M. P. (2012). Pairing tone trains with vagus nerve stimulation induces temporal plasticity in auditory cortex. *Experimental neurology*, *233*(1), 342-349. https://doi.org/10.1016/j.expneurol.2011.10.026

Shim, H. J., Kwak, M. Y., An, Y. H., Kim, D. H., Kim, Y. J., & Kim, H. J. (2015). Feasibility and safety of transcutaneous vagus nerve stimulation paired with notched music therapy for the treatment of chronic tinnitus. *Journal of Audiology & Otology*, *19*(3), 159. https://doi.org/10.7874/jao.2015.19.3.159

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter?. *Psychological science*, *17*(8), 675-681. https://doi.org/10.1111/j.1467-9280.2006.01765.x

So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195-221.

Strange, W., Bohn, O. S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *The Journal of the Acoustical Society of America*, *115*(4), 1791-1807.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & psychophysics*, *36*(2), 131-145. https://doi.org/10.3758/bf03202673

Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, *43*(4). https://doi.org/10.1515/iral.2005.43.4.329

Sun, J. B., Cheng, C., Tian, Q. Q., Yuan, H., Yang, X. J., Deng, H., ... & Qin, W. (2021). Transcutaneous auricular vagus nerve stimulation improves spatial working memory in healthy young adults. *Frontiers in Neuroscience*, *15*, 790793. https://doi.org/10.3389/fnins.2021.790793

Sun, L., Peräkylä, J., Holm, K., Haapasalo, J., Lehtimäki, K., Ogawa, K. H., ... & Hartikainen, K. M. (2017). Vagus nerve stimulation improves working memory performance. *Journal of Clinical and Experimental Neuropsychology*, *39*(10), 954-964. https://doi.org/10.1080/13803395.2017.1285869

Thair, H., Holloway, A. L., Newport, R., & Smith, A. D. (2017). Transcranial direct current stimulation (tDCS): a beginner's guide for design and implementation. *Frontiers in neuroscience*, *11*, 276151.

Thakkar, V. J., Engelhart, A. S., Khodaparast, N., Abadzi, H., & Centanni, T. M. (2020). Transcutaneous auricular vagus nerve stimulation enhances learning of novel letter-sound relationships in adults. *Brain Stimulation*, *13*(6), 1813-1820. https://doi.org/10.1016/j.brs.2020.10.012

Thakkar, V. J., Richardson, Z. A., Dang, A., & Centanni, T. M. (2023). The effect of non-invasive vagus nerve stimulation on memory recall in reading: A pilot study. *Behavioural Brain Research*, *438*, 114164. https://doi.org/10.1016/j.bbr.2022.114164

Van Leusden, J. W., Sellaro, R., & Colzato, L. S. (2015). Transcutaneous Vagal Nerve Stimulation (tVNS): a new neuromodulation tool in healthy humans?. *Frontiers in psychology*, *6*, 102. https://doi.org/10.3389/fpsyg.2015.00102

Ventura-Bort, C., Wirkner, J., Genheimer, H., Wendt, J., Hamm, A. O., & Weymar, M. (2018). Effects of transcutaneous vagus nerve stimulation (tVNS) on the P300 and alpha-amylase level: a pilot study. *Frontiers in human neuroscience*, *12*, 202. https://doi.org/10.3389/fnhum.2018.00202

Villani, V., Tsakiris, M., & Azevedo, R. T. (2019). Transcutaneous vagus nerve stimulation improves interoceptive accuracy. *Neuropsychologia*, *134*, 107201. https://doi.org/10.1016/j.neuropsychologia.2019.107201

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of cognitive neuroscience*, *8*(3), 231-256. https://doi.org/10.1162/jocn.1996.8.3.231

White, E. J., Hutka, S. A., Williams, L. J., & Moreno, S. (2013). Learning, neural plasticity and sensitive periods: implications for language acquisition, music training and transfer across the lifespan. *Frontiers in systems neuroscience*, *7*, 90. https://doi.org/10.3389/fnsys.2013.00090

Williams, M., Burden, R., Poulet, G., & Maun, I. (2004). Learners' perceptions of their successes and failures in foreign language learning. *Language Learning Journal*, *30*(1), 19-29. https://doi.org/10.1080/09571730485200191

Yakunina, N., Kim, S. S., & Nam, E. C. (2017). Optimization of transcutaneous vagus nerve stimulation using functional MRI. *Neuromodulation: technology at the neural interface*, *20*(3), 290-300. https://doi.org/10.1111/ner.12541

Yuan, H., & Silberstein, S. D. (2016a). Vagus nerve and vagus nerve stimulation, a comprehensive review: part I. *Headache: The Journal of Head and Face Pain*, *56*(1), 71-78. https://doi.org/10.1111/head.12647

Yuan, H., & Silberstein, S. D. (2016b). Vagus nerve and vagus nerve stimulation, a comprehensive review: part II. *Headache: The Journal of Head and Face Pain*, *56*(2), 259-266. https://doi.org/10.1111/head.12650

Zuo, Y., Smith, D. C., & Jensen, R. A. (2007). Vagus nerve stimulation potentiates hippocampal LTP in freely-moving rats. *Physiology & behavior*, *90*(4), 583-589. https://doi.org/10.1016/j.physbeh.2006.11.009

## Supplemental Materials

### Demographic Information

As described in the main text, participants completed a questionnaire about demographics, language history and proficiency, and musical experience. Supplemental Table 1 displays a summary of some demographic information of interest for each group.

**Supplemental Table 1**

*Summary of language and music experience for each group*

| Measure | Control (Mean ± SD) | taVNS-vowel (Mean ± SD) | taVNS-fricative (Mean ± SD) | One-way ANOVA output |
|---|---|---|---|---|
| Age of first L2 exposure | 11.30 ± 4.59 | 10.1 ± 4.50 | 7.88 ± 3.18 | $F_{2,34} = 1.62$, $p = 0.21$ |
| Years of second language exposure | 8.47 ± 4.93 | 8.93 ± 6.11 | 11.50 ± 5.68 | $F_{2,34} = 0.83$, $p = 0.45$ |
| Self-rated L2 listening ability (1 = very poor, 7 = native like) | 3.13 ± 1.36 | 3.00 ± 1.36 | 3.25 ± 1.67 | $F_{2,34} = 0.08$, $p = 0.92$ |
| Self-rated L2 speaking ability (1 = very poor, 7 = native like) | 2.60 ± 1.40 | 2.36 ± 1.34 | 2.38 ± 0.52 | $F_{2,34} = 0.16$, $p = 0.85$ |
| Years of musical training | 5.25 ± 5.85 | 3.60 ± 4.72 | 1.86 ± 4.42 | $F_{2,42} = 1.68$, $p = 0.20$ |

*Note*. L2 = second language.

### Amplitude Analyses

As described in the main text, a mixed-effects logistic regression model was fit with the goal of determining whether participants' accuracy on the training task depended on the amplitude of taVNS received. Supplemental Table 2 displays the output of this model and Supplemental Figure 1 displays accuracy over time for stimulated compared to unstimulated trials. No significant effects related to trial type or stimulation amplitude were found.

**Supplemental Table 2**

*Summary of the mixed-effects logistic regression model predicting accuracy on the training task, with trial type and stimulation amplitude as fixed effects*

**Fixed effects**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $z$ | $p$ |
|---|---|---|---|---|
| Intercept | 1.866 | 0.191 | 9.748 | <.001 |
| Trial Type | -0.093 | 0.220 | -0.423 | 0.672 |
| Trial Number | 0.684 | 0.097 | 7.065 | <.001 |
| Amplitude | 0.382 | 0.285 | 1.341 | 0.180 |
| Trial Type:Trial Number | -0.093 | 0.138 | -0.677 | 0.499 |
| Trial Type:Amplitude | 0.632 | 0.445 | 1.420 | 0.155 |
| Trial Number:Amplitude | 0.158 | 0.174 | 0.906 | 0.365 |
| Trial Type:Trial Number:Amplitude | -0.148 | 0.284 | -0.520 | 0.603 |

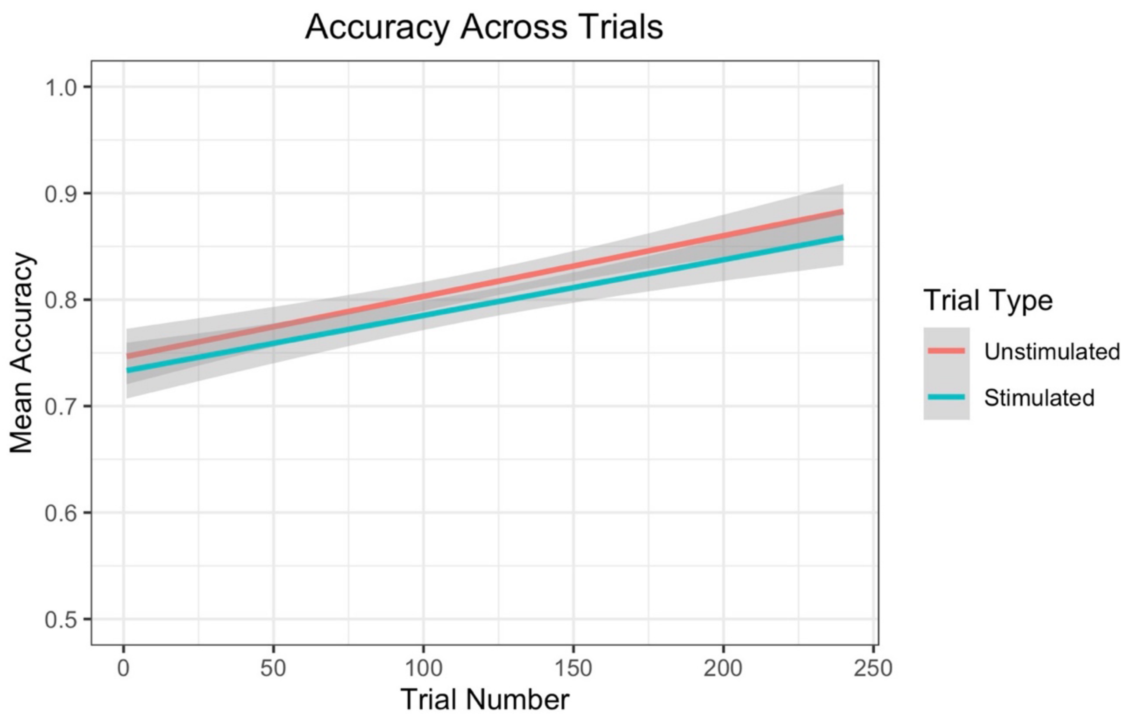
**Random effects**

| Group | Term | Variance | SD |
|---|---|---|---|
| Stimulus | Intercept | 0.692 | 0.832 |
| | Amplitude | 0.172 | 0.415 |
| | Trial Number | 0.053 | 0.231 |
| Subject | Intercept | 0.523 | 0.723 |
| | Trial Number | 0.076 | 0.276 |
| | Trial Type | 1.279 | 1.131 |

*Note.* Number of observations: 7200, groups: stimulus (40), subject (30). *p*-values calculated using the Laplace approximation. Model equation: Accuracy ~ Trial Type * Trial Number * Amplitude + (1 + Trial Type + Trial Number || Subject) + (1 + Trial Number + Amplitude || Stimulus).

**Supplemental Figure 1**

*Average accuracy as trials progressed on the training task*




*Note.* Lines of best fit with 95% confidence intervals are shown.

**Retention of Correct Stimulus-Response Associations**

As described in the main text, a linear mixed-effects regression model was fit predicting the retention of correct stimulus-response associations across blocks, with the goal of determining whether retention might be improved by stimulation. Supplemental Table 3 displays the output of this model and Supplemental Figure 2 displays retention over time for each group. No significant group effects were found.

**Supplemental Table 3**

*Summary of the linear mixed-effects regression model predicting retention of correct stimulus-to-response associations on the training task*

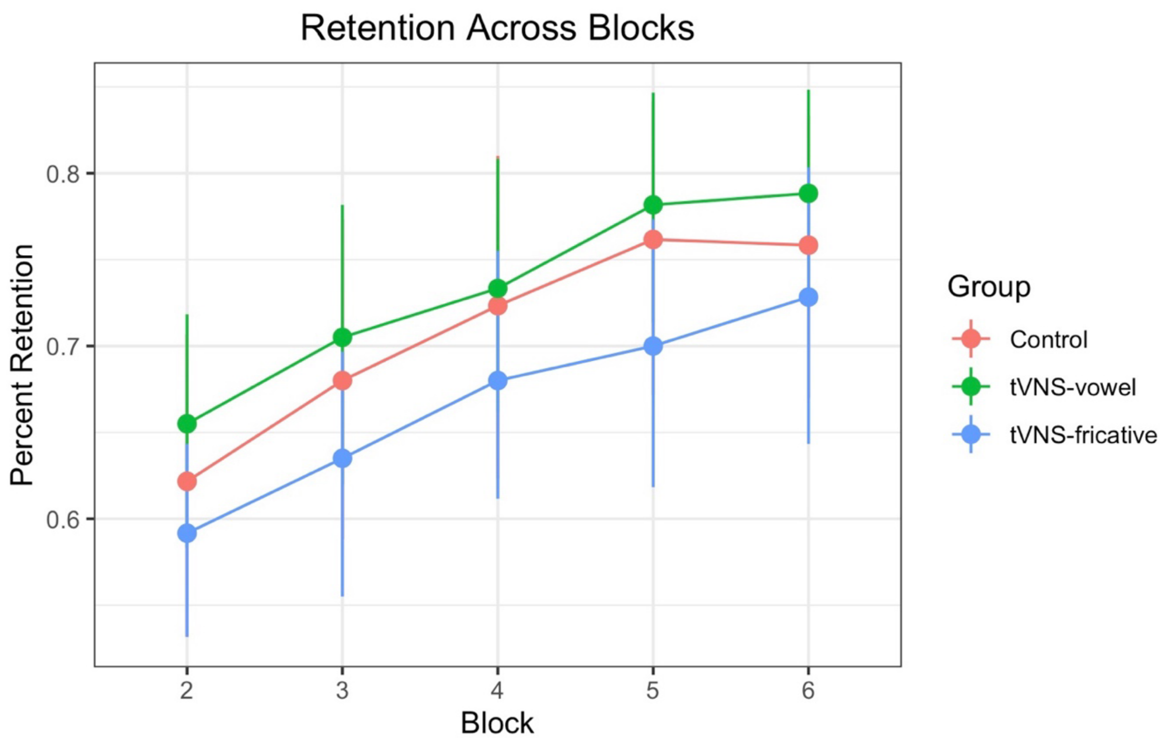
| **Fixed effects** | | | | | |
|---|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *df* | *p* |
| Intercept | 0.622 | 0.041 | 15.113 | 54.577 | <.001 |
| Group taVNS-vowel | 0.033 | 0.058 | 0.573 | 54.577 | 0.569 |
| Group taVNS-fricative | -0.030 | 0.058 | -0.516 | 54.577 | 0.608 |
| Block 2-3 retention | 0.058 | 0.023 | 2.549 | 336 | 0.011 |
| Block 3-4 retention | 0.102 | 0.023 | 4.442 | 336 | <.001 |
| Block 4-5 retention | 0.140 | 0.023 | 6.117 | 336 | <.001 |
| Block 5-6 retention | 0.137 | 0.023 | 5.971 | 336 | <.001 |
| Contrast | 0.010 | 0.058 | 0.173 | 73.858 | 0.863 |
| Group taVNS-vowel:Block 2-3 | -0.008 | 0.032 | -0.257 | 336 | 0.797 |
| Group taVNS-fricative:Block 2-3 | -0.015 | 0.032 | -0.463 | 336 | 0.643 |
| Group taVNS-vowel:Block 3-4 | -0.023 | 0.032 | -0.721 | 336 | 0.472 |
| Group taVNS-fricative:Block 3-4 | -0.013 | 0.032 | -0.412 | 336 | 0.681 |
| Group taVNS-vowel:Block 4-5 | -0.013 | 0.032 | -0.412 | 336 | 0.681 |
| Group taVNS-fricative:Block 4-5 | -0.032 | 0.032 | -0.978 | 336 | 0.329 |
| Group taVNS-vowel:Block 5-6 | -0.003 | 0.032 | -0.103 | 336 | 0.918 |
| Group taVNS-fricative:Block 5-6 | 0.000 | 0.032 | 0.000 | 336 | 1.000 |
| Group taVNS-vowel:Contrast | -0.073 | 0.082 | -0.898 | 73.858 | 0.372 |
| Group taVNS-fricative:Contrast | -0.113 | 0.082 | -1.387 | 73.858 | 0.170 |
| Block 2-3:Contrast | -0.037 | 0.046 | -0.801 | 336 | 0.424 |
| Block 3-4:Contrast | -0.083 | 0.046 | -1.82 | 336 | 0.070 |
| Block 4-5:Contrast | -0.047 | 0.046 | -1.019 | 336 | 0.309 |
| Block 5-6:Contrast | -0.020 | 0.046 | -0.437 | 336 | 0.662 |
| Group taVNS-vowel:Block 2-3:Contrast | -0.017 | 0.065 | -0.257 | 336 | 0.797 |
| Group taVNS-fricative:Block 2-3:Contrast | -0.003 | 0.065 | -0.051 | 336 | 0.959 |
| Group taVNS-vowel:Block 3-4:Contrast | -0.033 | 0.065 | -0.515 | 336 | 0.607 |
| Group taVNS-fricative:Block 3-4:Contrast | 0.013 | 0.065 | 0.206 | 336 | 0.837 |
| Group taVNS-vowel:Block 4-5:Contrast | 0.007 | 0.065 | 0.103 | 336 | 0.918 |
| Group taVNS-fricative:Block 4-5:Contrast | -0.057 | 0.065 | -0.875 | 336 | 0.382 |
| Group taVNS-vowel:Block 5-6:Contrast | -0.033 | 0.065 | -0.515 | 336 | 0.607 |
| Group taVNS-fricative:Block 5-6:Contrast | -0.060 | 0.065 | -0.927 | 336 | 0.355 |

| **Random effects** | | | |
|---|---|---|---|
| Group | Term | Variance | SD |
| Subject | Intercept | 0.021 | 0.146 |
| | Contrast | 0.034 | 0.185 |

*Note.* Number of observations: 450, groups: subject (45). *p*-values/*df* calculated using the Satterthwaite approximation. Model equation: Percent Retention ~ Group * Block * Contrast + (1 + Contrast || Subject).

**Supplemental Figure 2**

*Average percentage of correct responses retained across blocks*




*Note.* Retention is defined as correctly responding to a stimulus on block *n* and block *n*-1. Vertical lines denote 95% confidence intervals for group retention on each block.

## Reaction Time and Post-Error Slowing Analyses

As described in the main text, a mixed-effects linear regression model was fit predicting reaction times on the training task, with the goal of determining whether stimulation might speed up reaction times. Supplemental Table 4 displays the output of this model and Supplemental Figure 3 displays reaction times as trials progressed for each group. No significant group effects were found. A linear mixed-effects regression model was also fit predicting post-error slowing values in order to determine whether stimulation might increase reaction times on trials following errors. Supplemental Table 5 displays the output of this model and Supplemental Figure 4 displays post-error slowing values for each group. Again, no significant group effects were found.

**Supplemental Table 4**

*Summary of the linear mixed-effects regression model predicting reaction times on the training task*

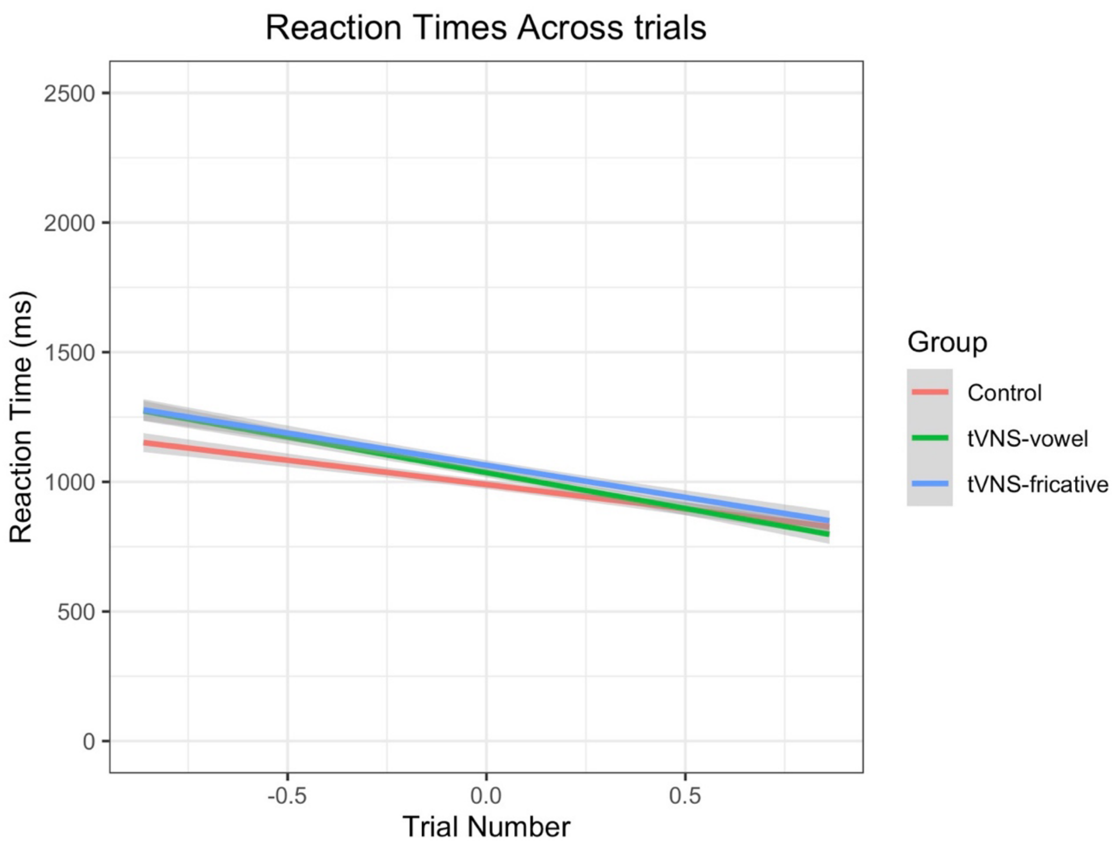
| **Fixed effects** | | | | | |
|---|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | *t* | *df* | *p* |
| Intercept | 6.808 | 0.074 | 91.473 | 44.506 | <.001 |
| Group taVNS-vowel | 0.095 | 0.104 | 0.913 | 41.982 | 0.366 |
| Group taVNS-fricative | 0.135 | 0.104 | 1.300 | 42.025 | 0.201 |
| Trial Number | -0.238 | 0.033 | -7.105 | 41.707 | <.001 |
| Contrast | 0.077 | 0.054 | 1.442 | 62.052 | 0.154 |
| Group taVNS-vowel:Trial Number | -0.081 | 0.047 | -1.717 | 41.261 | 0.093 |
| Group taVNS-fricative:Trial Number | -0.058 | 0.047 | -1.222 | 42.094 | 0.229 |
| Group taVNS-vowel:Contrast | 0.003 | 0.067 | 0.052 | 41.874 | 0.958 |
| Group taVNS-fricative:Contrast | 0.119 | 0.067 | 1.773 | 42.294 | 0.083 |
| Trial Number:Contrast | -0.013 | 0.041 | -0.325 | 41.387 | 0.747 |
| Group taVNS-vowel:Trial Number:Contrast | 0.092 | 0.057 | 1.602 | 40.608 | 0.117 |
| Group taVNS-fricative:Trial Number:Contrast | 0.007 | 0.058 | 0.125 | 42.931 | 0.901 |

| **Random effects** | | | |
|---|---|---|---|
| Group | Term | Variance | SD |
| Subject | Intercept | 0.079 | 0.282 |
| | Trial Number | 0.012 | 0.11 |
| | Contrast | 0.029 | 0.17 |
| | Trial Number:Contrast | 0.006 | 0.078 |
| Stimulus | Intercept | 0.214 | 0.08 |

*Note.* Number of observations: 8495, groups: subject (45), stimulus (40). *p*-values/*df* calculated using the Satterthwaite approximation. Model equation: RT ~ Group * Trial Number * Contrast + (1 + Trial Number * Contrast || Subject) + (1 | Stimulus).

**Supplemental Figure 3**

*Average reaction times as trials progressed on the training task*




*Note.* Lines of best fit with 95% confidence intervals are shown.

**Supplemental Table 5**

*Summary of the linear mixed-effects regression model predicting post-error slowing on the training task*

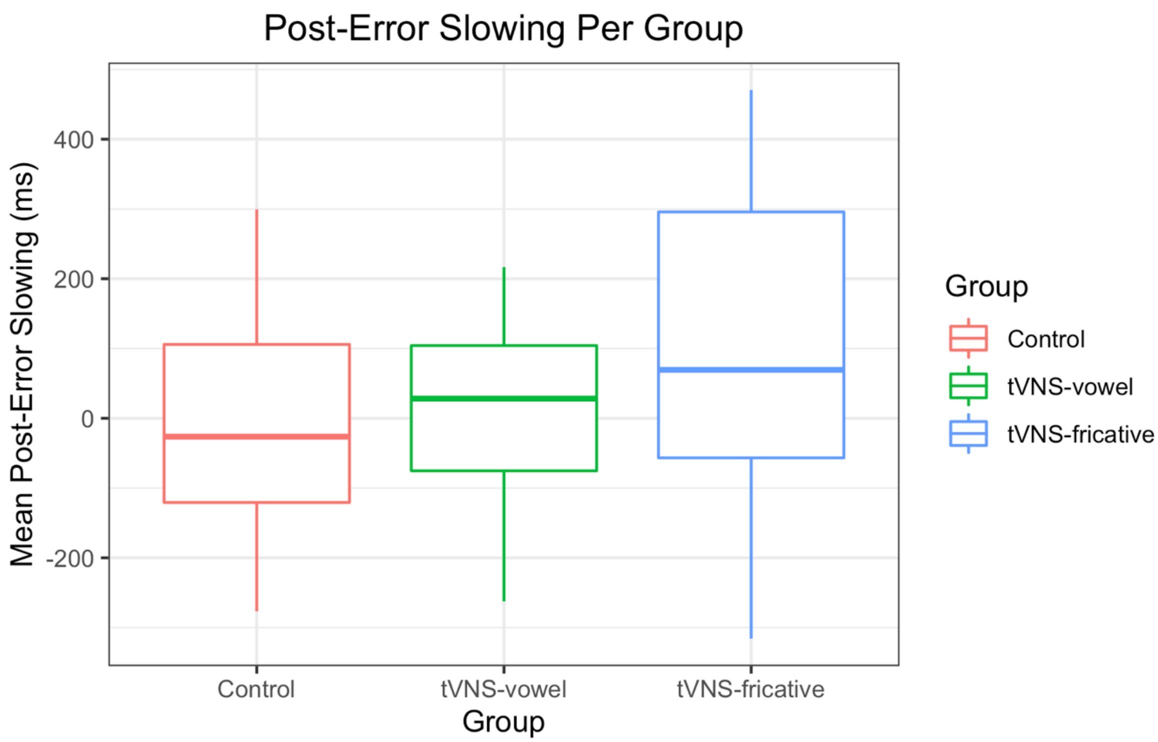
| **Fixed effects** | | | | | |
|---|---|---|---|---|---|
| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $t$ | $df$ | $p$ |
| Intercept | 6.215 | 62.395 | 0.100 | 224.994 | 0.921 |
| Group taVNS-vowel | -14.452 | 87.595 | -0.165 | 1103.234 | 0.869 |
| Group taVNS-fricative | 114.617 | 82.515 | 1.389 | 1096.867 | 0.165 |

| **Random effects** | | | |
|---|---|---|---|
| Group | Term | Variance | SD |
| Stimulus | Intercept | 6400 | 80.00 |

*Note.* Number of observations: 1107, groups: stimulus (40). *p*-values/*df* calculated using the Satterthwaite approximation. Model equation: PES ~ Group + (1 | Stimulus).

**Supplemental Figure 4**

*Boxplots of post-error slowing values for each group*




**Motivation Analyses**

As described in the main text, an ordinal mixed-effects model was fit predicting scores across the three subscales of the motivation questionnaire, with the goal of determining whether stimulation might increase motivation post-training. Supplemental Table 6 displays the output of this model. No significant group effects were found.

**Supplemental Table 6**

*Summary of the ordinal mixed-effects regression model and post-hoc pairwise comparisons predicting participants' motivation ratings across the three subscales*

**Fixed effects**

| Coefficient | $\hat{\beta}$ | $SE(\hat{\beta})$ | $z$ | $p$ |
|---|---|---|---|---|
| Intercept 1\|2 | -3.440 | 0.557 | -6.174 | <.001 |
| Intercept 2\|3 | -1.897 | 0.548 | -3.459 | 0.001 |
| Intercept 3\|4 | -0.516 | 0.546 | -0.946 | 0.344 |
| Intercept 4\|5 | 0.678 | 0.546 | 1.242 | 0.214 |
| Intercept 5\|6 | 2.120 | 0.550 | 3.854 | <.001 |
| Intercept 6\|7 | 3.836 | 0.563 | 6.812 | <.001 |
| Time Post | 0.091 | 0.207 | 0.440 | 0.660 |
| Group taVNS-vowel | -0.419 | 0.537 | -0.780 | 0.435 |
| Group taVNS-fricative | -0.719 | 0.539 | -1.335 | 0.182 |
| Time Post:Group taVNS-vowel | 0.425 | 0.293 | 1.450 | 0.147 |
| Time Post:Group taVNS-fricative | 0.317 | 0.294 | 1.077 | 0.282 |

**Random effects**

| Group | Term | Variance | SD | Correlation |
|---|---|---|---|---|
| Subject | Intercept | 1.888 | 1.3741 | |
| | Time Post | 0.098 | 0.314 | -0.661 |
| Item | Intercept | 1.803 | 1.343 | |

**Post-hoc pairwise comparisons**

| Contrast | $\hat{\beta}$ | $SE(\hat{\beta})$ | $z$ ratio | $p$ |
|---|---|---|---|---|
| Pre – Post (Control group) | -0.048 | 0.232 | -0.205 | 0.837 |
| Pre – Post (taVNS-vowel group) | -0.846 | 0.253 | -3.343 | <.001 |
| Pre – Post (taVNS-fricative group) | -0.472 | 0.252 | -1.875 | 0.061 |

*Note.* Number of observations: 1078, groups: subject (45), item (12). *p*-values calculated using the Laplace approximation. Model equation: Rating ~ Time * Group + (1 + Time | Subject) + (1 | Item).

## General Discussion

### Summary of Results

In Chapter 1, we ascertained that healthy young adults show extensive individual differences in performance on native and non-native phonetic perception tasks. By relating these differences to one another, we clarified that slopes on 2AFC and VAS tasks do not measure the same construct, but that the two tasks are related through the consistency of listeners' responses; listeners with more consistent VAS responses showed steeper 2AFC slopes. We further found that listeners with more consistent responses on the native VAS task tended to have better non-native perception outcomes. This work elucidated the relationships among measures of individual differences in native and non-native speech perception.

In Chapter 2, we recorded EEG with the goal of determining, at the neural level, what might be contributing to the individual differences observed in Chapter 1. We anticipated that the FFR, as a neural marker of auditory processing, might relate to differences in native and non-native speech perception. Our results did not reveal conclusive evidence of relationships between differences in the FFR and differences in our measures of speech perception, though exploratory analyses pointed to a potential link between more consistent FFRs and more consistent native perception.

Finally, in Chapter 3, we asked whether taVNS could facilitate the acquisition of unfamiliar non-native speech sounds and improve language learning motivation. Two experimental groups of participants received stimulation during non-native perception training, while a control group did not receive any stimulation. We compared measures of non-native perception, production, and language learning motivation across groups. Overall, we did not find clear benefits of taVNS across our measures. One of the experimental groups did show significant decreases in feelings of pressure and tension associated with language learning (one of the three subscales of motivation measured) from pre- to post-training, which was not the case for the other groups, so there is weak evidence for a potential role of taVNS in improving some aspects of motivation during language learning.

**Individual Differences**

Across all three chapters of this thesis, one clear finding was that healthy young adults show extensive individual differences in both native and non-native speech sound processing. These differences are evident as measured neurally as well as behaviourally. This result highlights the necessity of accounting for individual differences when conducting speech perception research and neurolinguistic research more broadly. In countless studies, participants are grouped together despite there being wide-ranging variability in performance within a group, and analyses are only conducted at the group level. And yet, as demonstrated in this thesis, there is undoubtedly valuable information to be gained from considering individual differences and from seeking to uncover the correlates of such differences.

Given that adults vary so widely in their speech sound processing, it is also of key importance to develop personalized approaches to language education. No two learners are exactly the same, and learners can no doubt benefit from different interventions depending on their individual profiles. Future work will need to characterize different learner profiles and to ascertain how particular profiles might best be matched to particular interventions, as a step toward improving the efficacy of language learning programs.

**Challenging the Use of Categorical Tasks and the Theory of Categorical Perception**

In comparing individuals' responses across two different native speech perception tasks (2AFC and VAS), Chapters 1 and 2 clearly point to the importance of task choice when designing speech perception experiments. 2AFC tasks, which have been highly prevalent in past research, turn out to have several limitations: they are more prone to bias from adjacent tasks (Munson et al., 2017); they require a categorical response which limits participants' ability to describe their percept in a nuanced way; and as found in Chapters 1 and 2, their slopes seem to reflect consistency of responses, which may not always be the construct of central interest in a given study. 2AFC tasks may still be useful in circumstances where the slope of the identification function is not of primary interest, such as when studying shifts in category boundaries during perceptual learning paradigms. However, for studies investigating the gradient nature of speech perception, we can conclude that VAS tasks are a much more suitable option.

The current thesis also illustrates how task demands have contributed to the persistence of categorical perception as a theory, and it sheds light on the theory's limitations. According to the theory of categorical perception, which has been hugely influential not only in speech perception research but also in other fields such as vision (e.g., Franklin et al., 2005), our perception is warped based on our top-down knowledge so that continuous stimuli are encoded as distinct categories rather than as continua

(Goldstone & Hendrickson, 2010; Liberman et al., 1957). In the context of speech perception, successful listeners were therefore thought to be those that showed "categorical perception" as reflected by steep slopes on 2AFC tasks coupled with peaks at the crossover boundary on discrimination tasks; and researchers often relied solely on 2AFC slopes as a measure of categorical perception (e.g., Chiappe et al., 2001; Manis et al., 1997; Joanisse et al., 2000). Challenging this view, the first two chapters of this thesis show that when listeners are provided with the opportunity to respond in a continuous way to speech stimuli, many of them who would have been labelled as "categorical" listeners based on 2AFC performance in fact show gradient, within-category sensitivity. That is, some listeners with steep slopes on the 2AFC task showed shallow slopes on the VAS task, presumably because they adapted to task demands; they provided categorical responses when required to fit their percepts into one of two categories on the 2AFC task, and gradient responses when encouraged to indicate the within-category nuances of their percepts on the VAS task. Chapters 1 and 2 thus add to mounting evidence that the theory and interpretations of categorical perception do not hold up to scrutiny (see McMurray, 2022, for a review).

Given the increasingly strong case against categorical perception, alternative theories must be considered. Many years ago, Pisoni and Tash (1974) proposed a model under which between-category discrimination can be better than within-category discrimination (as has often been observed and used as evidence for categorical perception), even when the underlying auditory perception is gradient. This outcome is possible because when discriminating between-category sounds, the listener can harness both lower-level acoustic information from the continuous acoustic space and higher-level phonemic information from their category knowledge; but when discriminating within-category sounds, the listener can harness only lower-level acoustic information (Pisoni & Tash, 1974). Another model of speech perception, C-CuRE (computing cues relative to expectations), posits that phonetic cues are encoded continuously but are then recoded relative to expectations derived from the context (McMurray & Jongman, 2011). For instance, as a cue to voicing, a listener would initially encode a talker's true fundamental frequency (F0) at the onset of a vowel and would then recode the cue as the difference between its true value and its expected value based on the average F0 of the given talker and vowel (McMurray & Jongman, 2011). Our findings of within-category sensitivity (as measured by the VAS task) are in line with such models that allow for fundamentally gradient phonetic perception. We encourage the continued development and use of theoretical frameworks such as these, which account for both the within- and between-category sensitivity that have been documented in many speech perception studies.

Interestingly, there have been parallel discussions in the field of syntax regarding the need to move beyond categorical tasks and theories. Grammatical knowledge has often been considered categorical—it was thought that either a sentence is grammatically acceptable, or it is not. However, more recent work argues that grammatical acceptability judgments are intrinsically gradient (Lau et al., 2017), or that even though there may be a categorical distinction between grammatical and ungrammatical sentences, it is beneficial to measure constructs using a variety of paradigms in order to arrive at more robust conclusions (Sprouse, 2007). For example, in addition to traditional yes/no tasks that elicit categorical responses about grammatical knowledge, there are options like magnitude estimation, which is a technique that allows responses along an infinite continuum of values (Sprouse, 2007). It will likely be fruitful for the field of language research, and perhaps for cognitive science more broadly, to transition towards the use of such nuanced tasks in order to paint a more detailed picture of individuals' linguistic and cognitive abilities.

**Neural and Behavioural Consistency of Responses to Speech Sounds**

Response consistency has been a relatively understudied behavioural measure of speech perception in comparison to other measures such as identification slope; and yet, this thesis and other current work are pointing to its importance (Fuhrmeister et al., 2023; Kapnoula & Samuel, 2023). Throughout Chapters 1 and 2 of this thesis, response consistency emerges as a noteworthy factor that relates to the successful perception of native and non-native speech sounds. Specifically, consistency seems to underlie optimal response patterns across the tasks that we administered: when participants responded more consistently to native speech sounds on a VAS task, they also tended to (1) respond more consistently to those sounds on a 2AFC task, (2) have steeper 2AFC slopes, indicating clear category boundaries, and (3) have shallower VAS slopes, indicating fine-tuned perception of subtle acoustic changes. Furthermore, Chapters 1 and 2 revealed a link between more consistent native perception and better identification of non-native speech sounds, which is corroborated by other recent research (Fuhrmeister et al., 2023; Kapnoula & Samuel, 2023). On the whole, these results suggest that listeners who can consistently map a speech sound to a behavioural response are at an advantage when it comes to successfully perceiving both native and non-native speech sounds.

Our novel findings about behavioural response consistency raise several questions. For instance, what are the functional neural correlates of this consistency? Chapter 2 investigated the FFR as one potential correlate, but the findings were not conclusive; a relationship between neural and behavioural response consistency emerged only when focussing on a sub-portion of the FFR in exploratory analyses. Despite these inconclusive results, the FFR remains a highly informative brainwave that shows interesting individual differences, and future work should study the behavioural correlates of these differences in depth. As for behavioural response consistency, perhaps its neural correlates emerge at a different level of sound processing. The FFR represents relatively early subcortical and cortical neural activity, reflecting the processing of a speech sound's spectral and temporal characteristics (Skoe & Kraus, 2010). It is possible that behavioural response consistency relates more to later processing stages involving speech comprehension or behavioural response planning. Work with functional neural measures that can capture higher-level speech processing and decision-making stages, such as the P3 event-related potential (which is sensitive to how long it takes to perceive and categorize a stimulus; Luck, 2014) or functional magnetic resonance imaging, might help to clarify which patterns of neural activity are associated with more consistent responses to speech sounds.

Another question that arises is how neural and/or behavioural response consistency might relate to other linguistic and cognitive measures. Perhaps a tendency to respond consistently at neural or behavioural levels is also related to other positive linguistic or cognitive outcomes, such as improved perception under more challenging listening conditions (e.g., perceiving speech with an unfamiliar accent or speech-in-noise) and improved performance on some measures of executive function, but this has yet to be determined. Additionally, perhaps consistency in speech perception is linked to consistency in speech production. Recall that in the present thesis, better non-native perception was predicted by more consistent native perception. In other work, better non-native perception has been predicted by more precise native production (i.e., more compact productions in acoustic space; Kartushina & Frauenfelder, 2014). This might suggest that consistent native perception and precise native production are themselves related. There have also been some more direct investigations of this question. One study found that children with more consistent behavioural responses to native sounds along a continuum from /ɹ/ to /w/ also tended to produce the /ɹ/ sound more accurately (McAllister Byun & Tiede, 2017). However, another study with adults failed to find a relationship between consistent behavioural responses to native vowels and precise production of those vowels (Cheng et al., 2021). In both cases, the researchers administered only a 2AFC task to measure native perception, and they used a different measure of response consistency that can be prone to ceiling effects (width of the category boundary, defined as the distance from the 25th to the 75th percentile of probability along a logistic function fitted to participants' 2AFC responses; McAllister Byun & Tiede, 2017). It would be interesting for future research to employ more sensitive tasks and measures (e.g., a consistency measure derived from VAS residuals, as used in this thesis and in Kapnoula et al. 2017) in order to investigate the question of potential relationships between consistency in speech sound perception and production. While the picture remains unclear as of yet, it may be the case that some individuals have more precisely defined categories in both perception and production, which promotes accurate performance in both modalities and facilitates the acquisition of novel non-native categories.

In light of the importance of response consistency as just described, another intriguing question is whether the consistency of neural and/or behavioural responses to sound can be improved through training. Using structural neuroimaging, Fuhrmeister and Myers (2021) found that differences in behavioural response consistency were predicted by differences in cortical gyrification, which is a neural measure thought to be more reflective of genetics than of experience. Consistency may therefore be partly innate, but this does not preclude the possibility that it could also be affected by experiential factors and training. Other prior work has compared the consistency of the FFR in bilingual versus monolingual participants, finding increased consistency for the bilingual group (Krizman et al., 2014). The authors concluded that bilingual experience increases the consistency of neural responses to sound (Krizman et al., 2014). Stronger evidence for such a conclusion could be provided by studies measuring neural response consistency within the same participants over time as those participants learn a new language. For instance, consistency could be compared before and after a non-native speech sound training intervention or an intensive language learning course. Note that, given the absence of conclusive evidence for a relationship between neural and behavioural response consistency, it is entirely conceivable that consistency at these two levels is differentially affected by training and experience, or that experience-related changes in consistency emerge on different timescales at each level. These possibilities have yet to be investigated, but previous research with non-native Mandarin tone training has found that training-related changes appear first in behavioural responses and later in the FFR, suggesting that there is indeed a distinction between processes at each level (Reetzke et al., 2018). In the future, there are many avenues to explore regarding the types of training and experiences that might be associated with more consistent responses to speech sounds at neural and behavioural levels.

Relatedly, it is also of interest to investigate to what extent the consistency of responses to speech sounds differs across clinical and non-clinical populations. At the neural level, reduced FFR response consistency has previously been associated with autism spectrum disorder (Otto-Meyer et al., 2018) and dyslexia (Hornickel & Kraus, 2013). Similarly, shallower 2AFC slopes on behavioural speech perception tasks (which reflect reduced response consistency as discussed in Chapters 1 and 2) have been related to dyslexia (Joanisse et al., 2000). By studying how response consistency differs across various populations, it may be possible to gain a deeper understanding of the mechanisms behind conditions such as dyslexia, and to eventually develop more targeted interventions to improve communication abilities in people with such conditions.

**Improving Speech Sound Learning**

As mentioned in the general introduction, learning non-native speech sounds in adulthood is often challenging, and this may be in part because the brain becomes less plastic and more specialized to process native speech sounds over time (see, for example, the concepts of sensitive periods and of native language neural commitment; Kuhl, 2004;

Werker & Hensch, 2015). Accordingly, it may be possible to improve speech sound learning in adulthood by improving neuroplasticity. In Chapter 3, we employed a neurostimulation method (taVNS) which has previously been shown to increase neuroplasticity (e.g., Engineer et al., 2015), with the aim of determining whether this method might enhance the acquisition of non-native speech sounds. We did not find evidence for effects of taVNS on speech sound learning; however, as discussed in Chapter 3, taVNS is just beginning to be investigated as a potential tool for enhancing language learning, and it may still hold promise for speech sound acquisition if a variety of stimulation parameters are tested.

Furthermore, as mentioned in the introduction, other neurostimulation methods exist that can increase neuroplasticity. These methods include transcranial magnetic stimulation (TMS), transcranial direct current stimulation (tDCS), and transcranial alternating current stimulation (tACS; Kricheldorff et al., 2022). Although such methods are generally less accessible and less capable of modulating widespread neural activity compared to taVNS (Frangos et al., 2015), they may have potential for improving speech sound learning in adulthood; yet very little work has tested this. In two of the few related studies, tDCS has been found to enhance performance on a non-native consonant cluster production task when it is administered prior to the task (Buchwald et al., 2019), and tDCS during a musical training task has been found to improve subsequent perception and production of non-native speech sounds (Borodkin et al., 2022). In addition, TMS has been shown to improve native speech production and fluency in people with aphasia (e.g., Finocchiaro et al., 2006; Szaflarski et al., 2011), presumably by modulating neuroplasticity.

The field of neurostimulation and language learning is evidently an exciting one, with many facets left to explore.

While neuroplasticity is generally reduced with age (Burke & Barnes, 2006), we now know that the brain does in fact remain plastic throughout life (Holman & de Villers-Sidani, 2014). Given this fact, and given the extensive evidence that many adults can actually learn unfamiliar non-native speech sounds (as demonstrated in this thesis and in myriad other studies, e.g., Iverson et al., 2005; Reetzke et al., 2018), we need not focus solely on neurostimulation as a means of enhancing speech sound acquisition. Behavioural training paradigms often lead to improvement in adults' non-native speech sound learning, and this is the case across different types of training (e.g., Giannakopoulou et al., 2017; Iverson et al., 2005). Therefore, it will be beneficial for subsequent studies to optimize existing non-native speech sound training paradigms and to compare the efficacy of various paradigms. As mentioned earlier, it may be especially effective to match training paradigms to individual learner profiles, since certain paradigms may be best suited to certain types of learners (Saito, 2023).

**Broader Relevance**

The current thesis focuses on the processing and acquisition of speech sounds. Beyond being of basic theoretical interest, non-native speech sound acquisition holds direct relevance for social integration, communication, and well-being. Because new phonemes are often difficult to acquire in adulthood, many adult learners have a non-standard accent when speaking a non-native language (Flege et al., 1995). Importantly, people who speak with a non-standard accent tend to have greater communication difficulties than native speakers, both based on their own subjective perceptions of their communication skills (Derwing & Rossiter, 2002) and based on objective measures of listener comprehension and processing (Adank et al., 2009; Munro & Derwing, 1995). Furthermore, non-standard accents are often stigmatized; for example, speakers with non-standard accents are consistently rated as being less intelligent, less successful, and less trustworthy than speakers with standard accents (Fuertes et al., 2012). It follows that people with non-standard accents tend to feel less of a sense of belonging in the country to which they have immigrated (Gluszek & Dovidio, 2010), and many of them believe that they would be respected more if they spoke with a standard accent (Derwing, 2003). Moving forward, it will be crucial to understand how to optimize speech sound acquisition so that non-native speakers can communicate effectively; how speech sounds spoken in a standard versus a non-standard accent are perceived differently, and how these differences might be mitigated; and how stigma around non-standard accents can be reduced with a view to improving equity and well-being.

In addition to their social significance, speech sounds are important as a building block of language learning more generally. As an illustration, successful native phonetic perception supports reading skills in childhood: children with steeper 2AFC slopes on native speech perception tasks (likely reflecting more consistent perception, as discussed in Chapters 1 and 2) also tend to show better reading performance (Chiappe et al., 2001; Joanisse et al., 2000). Moreover, speech discrimination training can improve reading skills in children with reading disabilities (González et al., 2002). In adulthood, non-native phonetic perception ability has been conceptualized as a “gatekeeper to the initial stages of foreign language learning" (Qi et al., 2019, p. 76). Supporting this view, participants who show better discrimination of non-native speech sounds are also more successful at learning new words containing those sounds (Chandrasekaran et al., 2010; Silbert et al., 2015). Similarly, participants with more efficient non-native speech sound processing (as measured using reaction times and EEG) also perform better on tests of non-native vocabulary knowledge and reading comprehension (Jakoby et al., 2011). Thus, phonetic perception seems to play a key role in supporting the acquisition of higher-order aspects of language, such as morphology, syntax, and semantics. By seeking to better understand and to improve phonetic perception, we can contribute to improving language learning more generally.

As immigration increases and multilingualism becomes the norm rather than the exception (Romaine, 2006), it is now more crucial than ever to support language learning in adulthood. In 2022, over 436,000 new permanent residents and 604,000 temporary workers immigrated to Canada (Miller, 2023), and the number of new permanent residents is planned to rise to 500,000 in 2025 (Government of Canada, 2023). Many of these immigrants will need to learn English or French in order to integrate into their new community. Indeed, in Quebec alone, more than 45,000 adults registered to learn French through the government's francization program during the 2022-2023 year (CBC News, 2023). On a global level, the soaring interest in adult language learning is also clear. The language learning app Duolingo had 88.4 million monthly active users in the last quarter of 2023, which represents a 46% increase from the previous year (Duolingo, 2024). Evidently, our increasingly diverse and interconnected world has a strong need for more research and more tools that address effective language acquisition.

## General Conclusion

The current thesis sought to better understand native and non-native speech sound perception in adults, with a focus on individual differences and on improving the learning of non-native sounds. To this end, we employed a multidisciplinary approach drawing from neuroscience, psychology, linguistics, and cognitive science. Across the three studies presented here, it is possible to draw a few broad conclusions. First, there are clear individual differences in adults' native and non-native speech processing, both at neural and behavioural levels. Further work is needed to clarify how differences are related across these two levels and why they arise, but we provide tentative evidence that more consistent neural responses to speech sounds may also be predictive of more consistent behavioural responses. Response consistency seems to be a rich measure of neural and behavioural speech sound processing that predicts other perceptual outcomes; it will be fruitful for future research to investigate the construct of consistency in greater depth. Additionally, our work emphasizes that researchers must verify the reliability and validity of the tasks that they use, and that they should carefully select tasks based on the constructs that they wish to measure. Finally, this research points to a need for further investigation of how neurostimulation and behavioural training might be harnessed to render the language learning process more efficient and enjoyable for adults.

Overall, this thesis contributes to our understanding of how speech sounds are perceived and acquired differently by different people, and of how interventions (such as perceptual training and neurostimulation) may improve non-native language acquisition. By bringing together and building upon these main themes—that is, by better characterizing individual learner profiles and by developing specialized learning programs for different profiles—it may eventually be possible to optimize speech sound learning for a broad range of learners. Speech sounds are at the core of spoken communication, and communication is at the core of the human experience. Through improving speech sound acquisition, we can improve communication and human connection for the millions of adults who are, or who will be, learning new languages.

## General References


Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human perception and performance*, *35*(2), 520.

Austelle, C. W., O'Leary, G. H., Thompson, S., Gruber, E., Kahn, A., Manett, A. J., Short, B., & Badran, B. W. (2022). A comprehensive review of vagus nerve stimulation for depression. *Neuromodulation: Technology at the Neural Interface*, *25*(3), 309–315. https://doi.org/10.1111/ner.13528

Badran, B. W., Dowdle, L. T., Mithoefer, O. J., LaBate, N. T., Coatsworth, J., Brown, J. C., DeVries, W. H., Austelle, C. W., McTeague, L. M., & George, M. S. (2018). Neurophysiologic effects of transcutaneous auricular vagus nerve stimulation (tavns) via electrical stimulation of the tragus: a concurrent tavns/fmri study and review. *Brain Stimulation*, *11*(3), 492–500. https://doi.org/10.1016/j.brs.2017.12.009

Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002, November). The effect of perceived phonetic similarity on non-native sound learning by children and adults. In *Proceedings of the annual Boston university conference on language development* (Vol. 26, No. 1, pp. 36-47). Somerville: Cascadilla Press.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, *167*(224), 233-277.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775-794.

Boroda, E., Sponheim, S. R., Fiecas, M., & Lim, K. O. (2020). Transcranial direct current stimulation (tDCS) elicits stimulus-specific enhancement of cortical plasticity. *Neuroimage*, *211*, 116598.

Borodkin, K., Gassner, T., Ershaid, H., & Amir, N. (2022). tDCS modulates speech perception and production in second language learners. *Scientific Reports*, *12*(1), 16212.

Buchwald, A., Calhoun, H., Rimikis, S., Lowe, M. S., Wellner, R., & Edwards, D. J. (2019). Using tDCS to facilitate motor learning in speech production: The role of timing. *Cortex*, *111*, 274-285.

Burke, S. N., & Barnes, C. A. (2006). Neural plasticity in the ageing brain. *Nature reviews neuroscience*, *7*(1), 30-40.

CBC News. (2023, May 29). *Quebec launches new platform for signing up to learn French*. https://www.cbc.ca/news/canada/montreal/francisation-quebec-learning-french-immigrant-1.6858111

Chandrasekaran, B., Sampath, P. D., & Wong, P. (2010). Individual variability in cue-weighting and lexical tone learning. *The Journal of the Acoustical Society of America*, *128*(1), 456-465.

Chang, Y. H. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The journal of the acoustical society of America*, *141*(2), EL120-EL126.

Cheng, H. S., Niziolek, C. A., Buchwald, A., & McAllister, T. (2021). Examining the relationship between speech perception, production distinctness, and production variability. *Frontiers in Human Neuroscience*, *15*, 660948.

Chiappe, P., Chiappe, D. L., & Siegel, L. S. (2001). Speech perception, lexicality, and reading skill. *Journal of Experimental Child Psychology*, *80*(1), 58-74.

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., ... & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature communications*, *10*(1), 5036.

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Routledge.

Derwing, T. (2003). What do ESL students say about their accents?. *Canadian Modern Language Review*, *59*(4), 547-567.

Derwing, T. M., & Rossiter, M. J. (2002). ESL learners' perceptions of their pronunciation needs and strategies. *System*, *30*(2), 155-166.

Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences*, *105*(42), 16083-16088.

Duolingo, Inc. (2024, February 28). *Duolingo reports 65% dau growth, 45% revenue growth and record profitability in fourth quarter 2023*. https://investors.duolingo.com/newsreleases/news-release-details/duolingo-reports-65-dau-growth-45-revenue-growth-and-record

Engineer, C. T., Engineer, N. D., Riley, J. R., Seale, J. D., & Kilgard, M. P. (2015). Pairing speech sounds with vagus nerve stimulation drives stimulus-specific cortical plasticity. *Brain stimulation*, *8*(3), 637-644.

Engineer, N. D., Riley, J. R., Seale, J. D., Vrana, W. A., Shetake, J. A., Sudanagunta, S. P., ... & Kilgard, M. P. (2011). Reversing pathological neural activity using targeted plasticity. *Nature*, *470*(7332), 101-104.

Fernandez, E. M., & Cairns, H. S. (2010). The Hearer: Speech Perception and Lexical Access. In *Fundamentals of Psycholinguistics* (pp. 169-203). John Wiley & Sons.

Finocchiaro, C., Maimone, M., Brighina, F., Piccoli, T., Giglia, G., & Fierro, B. (2006). A case study of primary progressive aphasia: improvement on verbs after rTMS treatment. *Neurocase*, *12*(6), 317-321.

Fuhrmeister, P., & Myers, E. B. (2021). Structural neural correlates of individual differences in categorical perception. *Brain and Language*, *215*, 104919. https://doi.org/10.1016/j.bandl.2021.104919.

Fuhrmeister, P., Phillips, M. C., McCoach, D. B., & Myers, E. B. (2023). Relationships between native and non-native speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001213.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*, 233-277.

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). *Second language speech learning: Theoretical and empirical progress*, 3-83.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics*, *25*(4), 437-470.

Flege, J. E., & MacKay, I. R. (2004). Perceiving vowels in a second language. *Studies in second language acquisition*, *26*(1), 1-34.

Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, *97*(5), 3125-3134.

Frangos, E., Ellrich, J., & Komisaruk, B. R. (2015). Non-invasive access to the vagus nerve central projections via electrical stimulation of the external ear: fMRI evidence in humans. *Brain stimulation*, *8*(3), 624-636.

Franklin, A., Clifford, A., Williamson, E., & Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of experimental child psychology*, *90*(2), 114-141.

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120-133.

Galbraith, G. C., Arbagey, P. W., Branski, R., Comerci, N., & Rector, P. M. (1995). Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport*, *6*(17), 2363-2367.

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209.

Gluszek, A., & Dovidio, J. F. (2010). Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of language and social psychology*, *29*(2), 224-234.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., & Pallier, C. (2007). Brain structure predicts the learning of foreign speech sounds. *Cerebral cortex*, *17*(3), 575-582.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(1), 69-78. https://doi.org/10.1002/wcs.26.

González, M. D. R. O., García Espinel, A. I., & Guzmán Rosquete, R. (2002). Remedial interventions for children with reading disabilities: speech perception—an effective component in phonological training?. *Journal of Learning Disabilities*, *35*(4), 334-342.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds" L" and" R.". *Neuropsychologia*.

Government of Canada (2023, November 1). *Notice – Supplementary Information for the 2024-2026 Immigration Levels Plan*. https://www.canada.ca/en/immigration-refugees-citizenship/news/notices/supplementary-immigration-levels-2024-2026.html

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning*, *62*, 79-109.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: An investigation of English/r/-/l/perception and production by adult Japanese speakers. In *Second language studies: Acquisition, learning, education and technology*.

Hogan, M. K., Hamilton, G. F., & Horner, P. J. (2020). Neural stimulation and molecular mechanisms of plasticity and regeneration: a review. *Frontiers in Cellular Neuroscience*, *14*. https://doi.org/10.3389/fncel.2020.00271

Holman, C., & Villers-Sidani, E. D. (2014). Indestructible plastic: the neuroscience of the new aging brain. *Frontiers in Human Neuroscience*, *8*, 219.

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: a biological marker of dyslexia. *Journal of Neuroscience*, *33*(8), 3500-3504. https://doi.org/10.1523/JNEUROSCI.4205-12.2013.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America*, *118*(5), 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47-B57.

Jakoby, H., Goldstein, A., & Faust, M. (2011). Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. *Psychophysiology*, *48*(11), 1517-1531.

Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of experimental child psychology*, *77*(1), 30-60. https://doi:10.1006/jecp.1999.2553.

Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. Brain and language, 192, 15-24. https://doi.org/10.1016/j.bandl.2019.02.004.

Kandel, E. R. (2013). *Principles of neural science* (5th ed.). McGraw-Hill.

Kapnoula, E. C., & McMurray, B. (2021). Idiosyncratic use of bottom-up and top-down information leads to differences in speech perception flexibility: Converging evidence from ERPs and eye-tracking. *Brain and Language*, *223*, 105031. https://doi.org/10.1016/j.bandl.2021.105031.

Kapnoula, E. C., & Samuel, A. G. (2023, May 31-June 2). *Examining the links between L1 phoneme categorization and non-native phonetic learning* [Poster presentation]. XVI International Symposium of Psycholinguistics, Vitoria-Gasteiz, Spain.

Kapnoula, E. C., Winn, M. B., Kong, E. J., Edwards, J., & McMurray, B. (2017). Evaluating the sources and functions of gradiency in phoneme categorization: An individual differences approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(9), 1594. https://doi.org/10.1037/xhp0000410.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences in L1 production on L2 pronunciation. *Frontiers in psychology*, *5*, 105122.

Kong, E. J., & Edwards, J. (2016). Individual differences in categorical perception of speech: Cue weighting and executive function. *Journal of Phonetics*, *59*, 40-57. https://doi.org/10.1016/j.wocn.2016.08.006.

Kricheldorff, J., Göke, K., Kiebs, M., Kasten, F. H., Herrmann, C. S., Witt, K., & Hurlemann, R. (2022). Evidence of neuroplastic changes after transcranial magnetic, electric, and deep brain stimulation. *Brain Sciences*, *12*(7), 929.

Krishnan, A., Bidelman, G. M., Smalt, C. J., Ananthakrishnan, S., & Gandour, J. T. (2012). Relationship between brainstem, cortical and behavioral measures relevant to pitch salience in humans. *Neuropsychologia*, *50*(12), 2849-2859. https://doi.org/10.1016/j.neuropsychologia.2012.08.013.

Krizman, J., & Kraus, N. (2019). Analyzing the FFR: A tutorial for decoding the richness of auditory function. *Hearing research*, *382*, 107779.

Krizman, J., Skoe, E., Marian, V., & Kraus, N. (2014). Bilingualism increases neural response consistency and attentional control: evidence for sensory and cognitive coupling. *Brain and Language, 128*, 34–40. https://doi.org/10.1016/j.bandl.2013.11.006.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. *Developmental neurocognition: Speech and face processing in the first year of life*, 259-274.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831-843.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 979-1000.

Lau, J. H., Clark, A., & Lappin, S. (2017). Grammaticality, acceptability, and probability: A probabilistic view of linguistic knowledge. *Cognitive science*, *41*(5), 1202-1241.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757-3768.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358.

Llanos, F., McHaney, J. R., Schuerman, W. L., Yi, H. G., Leonard, M. K., & Chandrasekaran, B. (2020). Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults. *Npj Science of Learning*, *5*(1). https://doi.org/10.1038/s41539-020-0070-0

Lorenz, I., Müller, N., Schlee, W., Langguth, B., & Weisz, N. (2010). Short-term effects of single repetitive TMS sessions on auditory evoked activity in patients with chronic tinnitus. *Journal of neurophysiology*, *104*(3), 1497-1505.

Luck, S. E. (2014). *An introduction to the event-related potential technique*. MIT press.

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia?. *Journal of experimental child psychology*, *66*(2), 211-235.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental science*, *11*(1), 122-134

McAllister Byun, T., & Tiede, M. (2017). Perception-production relations in later development of American English rhotics. *PloS one*, *12*(2), e0172022.

McHaney, J. R., Schuerman, W. L., Leonard, M. K., & Chandrasekaran, B. (2023). Transcutaneous auricular vagus nerve stimulation modulates performance but not pupil size during nonnative speech category learning. *Journal of Speech, Language, and Hearing Research : Jslhr*, *66*(10), 3825–3843. https://doi.org/10.1044/2023_JSLHR-22-00596

McMurray, B. (2022). The myth of categorical perception. *The Journal of the Acoustical Society of America*, *152*(6), 3819-3842. https://doi.org/10.1121/10.0016614.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological review*, *118*(2), 219.

Miller, M. (2023). *2023 Annual Report to Parliament on Immigration*. Immigration, Refugees and Citizenship Canada. https://www.canada.ca/content/dam/ircc/documents/pdf/english/corporate/publications-manuals/annual-report-2023-en.pdf

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics*, *18*(5), 331-340.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and speech*, *38*(3), 289-306.

Munson, B., Schellinger, S. K., & Edwards, J. (2017). Bias in the perception of phonetic detail in children's speech: A comparison of categorical and continuous rating scales. *Clinical Linguistics & Phonetics*, *31*(1), 56-79. https://doi.org/10.1080/02699206.2016.1233292.

O'Connor, J. D., Gerstman, L. J., Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1957). Acoustic cues for the perception of initial /w, j, r, l/I n English. *Word*, *13*(1), 24-43.

Omote, A., Jasmin, K., & Tierney, A. (2017). Successful non-native speech perception is linked to frequency following response phase consistency. Cortex, 93, 146-154. https://doi.org/10.1016/j.cortex.2017.05.005.

Otto-Meyer, S., Krizman, J., White-Schwoch, T., & Kraus, N. (2018). Children with autism spectrum disorder have unstable neural responses to sound. *Experimental Brain Research*, *236*, 733-743. https://doi.org/10.1007/s00221-017-5164-4.

Ou, J., & Yu, A. C. (2022). Neural correlates of individual differences in speech categorisation: evidence from subcortical, cortical, and behavioural measures. Language, Cognition and Neuroscience, 37(3), 269-284. https://doi.org/10.1080/23273798.2021.1980594.

Pandža, N. B., Phillips, I., Karuzis, V. P., O'Rourke, P., & Kuchinsky, S. E. (2020). Neurostimulation and pupillometry: new directions for learning and research in applied linguistics. *Annual Review of Applied Linguistics*, *40*, 56–77. https://doi.org/10.1017/S0267190520000069

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical–cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience*, *33*(3), 549-557. https://doi.org/10.1111/j.1460-9568.2010.07546.x.

Peña, M., Maki, A., Kovačić, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences*, *100*(20), 11702-11705.

Phillips, I., Calloway, R. C., Karuzis, V. P., Pandža, N. B., O'Rourke, P., & Kuchinsky, S. E. (2021). Transcutaneous auricular vagus nerve stimulation strengthens semantic representations of foreign language tone words during initial stages of learning. *Journal of Cognitive Neuroscience*, *34*(1), 127–152. https://doi.org/10.1162/jocn_a_01783

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & psychophysics*, *15*(2), 285-290.

Qi, Z., Han, M., Wang, Y., de Los Angeles, C., Liu, Q., Garel, K., ... & Perrachione, T. K. (2019). Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning. *NeuroImage*, *192*, 76-87

Reetzke, R., Xie, Z., Llanos, F., & Chandrasekaran, B. (2018). Tracing the trajectory of sensory plasticity across different stages of speech learning in adulthood. *Current Biology*, *28*(9), 1419-1427. https://doi.org/10.1016/j.cub.2018.03.026.

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: A normative dataset in healthy newborns. *Hearing Research*, *371*, 28-39.

Roach, P. (2009). *English phonetics and phonology paperback with audio CDs (2): A practical course*. 4th ed. Cambridge: Cambridge university press.

Romaine, S. (2006). Language Policy in Multilingual Educational Contexts. Encyclopedia of Language & Linguistics. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 584-596). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00646-5.

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing auditory precision hypothesis-L2. *Language Teaching*, *56*(4), 522-538.

Saito, K., Sun, H., & Tierney, A. (2018). Explicit and implicit aptitude effects on second language speech learning: scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. Bilingualism: Language and Cognition, 1-18. https://doi.org/10.1017/S1366728918000895.

Sebastián-Gallés, N., Soriano-Mas, C., Baus, C., Díaz, B., Ressel, V., Pallier, C., ... & Pujol, J. (2012). Neuroanatomical markers of individual differences in native and non-native vowel perception. *Journal of Neurolinguistics*, *25*(3), 150-162.

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., & Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *Journal of Phonetics*, *50*, 99-119.

Skoe, E., & Kraus, N. (2010). Auditory brain stem response to complex sounds: a tutorial. *Ear and hearing*, *31*(3), 302-324.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: does musical ability matter?. *Psychological science*, *17*(8), 675-681.

Song, J. H., Nicol, T., & Kraus, N. (2011a). Test–retest reliability of the speech-evoked auditory brainstem response. *Clinical Neurophysiology*, *122*(2), 346-355. https://doi.org/10.1016/j.clinph.2010.07.009.

Sprouse, J. (2007). Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics*, *1*, 123-134.

Stölten, K., Abrahamsson, N., & Hyltenstam, K. (2014). Effects of age of learning on voice onset time: Categorical perception of Swedish stops by near-native L2 speakers. *Language and Speech*, *57*(4), 425-450.

Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, *43*(4). https://doi.org/10.1515/iral.2005.43.4.329

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of/rl/by Japanese adults learning English. *Perception & psychophysics*, *36*(2), 131-145.

Strange, W., Hisagi, M., Akahane-Yamada, R., & Kubo, R. (2011). Cross-language perceptual similarity predicts categorial discrimination of American vowels by naïve Japanese listeners. *The Journal of the Acoustical Society of America*, *130*(4), EL226-EL231.

Szaflarski, J. P., Vannest, J., Wu, S. W., DiFrancesco, M. W., Banks, C., & Gilbert, D. L. (2011). Excitatory repetitive transcranial magnetic stimulation induces improvements in chronic post-stroke aphasia. *Medical science monitor: international medical journal of experimental and clinical research*, *17*(3), CR132.

Wang, B. S., Sarnaik, R., & Cang, J. (2010). Critical period plasticity matches binocular orientation preference in the visual cortex. *Neuron*, *65*(2), 246-256.

Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: new directions. *Annual Review of Psychology*, *66*, 173–96. https://doi.org/10.1146/annurev-psych-010814-015104

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America*, *75*(6), 1866-1878.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. I. (2005). Effects of language experience: neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703-720.