Disentangling the Roles of Formant Proximity and Stimulus Prototypicality on Asymmetries in Vowel Perception

Ying Ying Liu

School of Communication Sciences and Disorders
Faculty of Medicine

McGill University, Montreal
August 2019

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Communication Sciences and Disorders

Table of Contents

Acknowledgements

I would like to first express gratitude toward my thesis supervisor Dr. Linda Polka, who has shown me so much dedication and patience through her consistent support and valuable feedback during this process. I honestly couldn't possibly have asked for a better mentor. Under her guidance, I have grown as much as a student and as an individual. In addition, I am thankful to Dr. Meghan Clayards and Dr. Susan Rvachew for their constructive comments during the preparation phase and final phase of this paper, as well as Dr. Matthew Masapollo, who has provided the stimuli necessary for the creation of the present experiments along with additional advice. I would also like to thank the research coordinator Paloma Noriega for helping with part of participant testing. Finally, this research was funded by the NSERC discovery grant (RGPIN 05823-2018) to L. Polka.

Contribution of Authors

This thesis was supervised by Dr. Linda Polka and have greatly benefitted from her extensive feedback. Moreover, Dr. Matthew Masapollo and Dr. Lucie Ménard have generated and courteously provided the original stimulus set from which the tokens for the current study were selected. I have created the experimental set-up under the guidance of Dr. Linda Polka, completed data-collection, the analysis as well as the research manuscript for this Master's thesis.

Abstract

Disentangling the Roles of Formant Proximity and Stimulus Prototypicality on Asymmetries in Vowel Perception

Vowel discrimination is often asymmetric such that discriminating the same vowel pair is easier in one direction compared to the opposite direction. The Natural Referent Vowel (NRV; Polka & Bohn, 2011) framework interprets these directional asymmetries as a universal bias favoring "focal" vowel (i.e., vowels with prominent spectral peaks formed by the convergence of adjacent formants). The Native Language Magnet (NLM; Kuhl, 1991) model interprets asymmetries in terms of a language-specific bias due to distortion of perceptual space around native language vowel prototypes instead. To test these views, Masapollo *et al*. (2017) compared English- and French-speaking adults' discrimination of synthetic /u/ variants; this was informative because English /u/ is naturally less focal than French /u/. Their findings revealed asymmetries to be predicted by focalization only; however, stimulus limitations may explain the lack of prototype effects. The current study was designed to address these potential stimulus limitations. To do so, we synthesized a more refined series of vowel stimuli systematically varying in smaller psychophysical steps around the English /u/ and French /u/ prototypes to amplify the measurement of focalization and prototype effects. Native English speakers completed a category goodness-rating task followed by an AX-discrimination task using these new variants. Results indicated effects of both focalization and prototype. Moreover, they also show that these effects depend on the acoustic distance between tokens along the stimulus series.

Résumé

Discerner les rôles de la proximité des formants et de la prototypicalité du stimulus sur les asymétries de la perception de voyelles

La discrimination des voyelles se fait souvent de façon asymétrique, de sorte que discriminer la même paire de voyelles est plus facile dans une direction comparée à la direction opposée. Le modèle des voyelles de référence naturelles (Natural Referent Vowel – NRV ; Polka & Bohn, 2011) interprète ces asymétries directionnelles comme un biais universel favorisant les voyelles «focales» (c'est-à-dire les voyelles ayant des pics spectraux proéminents formés par la convergence des formants adjacents). Le modèle de l’aimant linguistique de la langue maternelle (Native Language Magnet – NLM ; Kuhl, 1991) interprète plutôt les asymétries en fonction de biais linguistiques spécifiques en lien avec la distorsion de l'espace perceptuel entourant les prototypes de voyelles appartenant au répertoire de sons de la langue maternelle du locuteur. Pour vérifier ces points de vue, Masapollo et al. (2017) ont comparé la discrimination des variantes synthétiques de la voyelle /u/ chez des adultes anglophones et francophones ; cette recherche était instructive car le /u/ anglais est naturellement moins «focal» que le /u/ français. Leurs résultats ont démontré que la présence des asymétries était prédite uniquement par la focalisation. Cependant, des limitations liées aux stimuli utilisés peuvent potentiellement expliquer l'absence d'effets de prototype. Cette étude a été conçue pour adresser ces limitations potentielles en lien avec les stimuli. Pour ce faire, nous avons synthétisé une série plus raffinée de voyelles-stimuli variant systématiquement en termes de distance en pas psychophysiques, de chaque côté des prototypes du /u/ anglais et du /u/ français afin d’amplifier à la fois les effets de focalisation et de prototype. Les locuteurs natifs de la langue anglaise ont effectué une tâche d’évaluation de qualité linguistique catégorique suivie d’une tâche de discrimination AX ciblant ces nouvelles variantes. Les résultats ont révélé la présence d’effets de focalisation et de prototype, et démontrent également que ces effets dépendent de la taille des intervalles acoustiques le long de la série de stimuli.

Introduction

Past literature reveals that a language-general process dominates speech perception in young infants (Kuhl, 1991, 1994; Polka & Werker, 1994), allowing them to discriminate between nearly every speech sound contrast during the first few months of their life, regardless of whether or not the sounds are present in their native linguistic repertoire. In contrast, discrimination of foreign language speech sounds often proves to be a much more challenging task for adults, which contributes to the difficulty of acquiring a second language. For instance, /r/ and /l/ are very difficult to distinguish for native Japanese-speaking adults as these sounds are used contrastively in English to differentiate meaning (e.g. ramp vs. lamp) but not in Japanese. However, in early infancy, both Japanese-learning and English-learning infants have no trouble discriminating these consonants (Tsushima, Takizawa, Sasaki, Shiraki, Nishi, Kohno, Menyuk & Best, 1994; Kuhl et al., 2006).

As babies acquire more exposure to the language(s) around them, their perceptual abilities become attuned to the speech sound categories that exist in their specific linguistic environment. In consequence, by their first birthday, their discrimination of native sound contrasts become heightened while their discrimination of non-native contrasts shows a significant decrease (Kuhl, 1994; Werker & Tees, 1984). For instance, at 6-8 months of age, English-learning infants can distinguish between the retroflex /t/ sound (produced with the tongue curled upwards toward the hard palate) and the dental /t/ sound (produced with the tongue against the upper teeth) which are used contrastively in Hindi but not in English (Werker & Tees, 1984). By the age of 10-12 months, the discrimination of this non-native contrast by English-learning infants (but not Hindi-learning infants) is reduced and becomes similar to the low level of perception observed in adult English-speakers while Hindi-learning infants' ability to perceive this contrast remains intact

(Werker and Tees, 1984).

This shift from universal to language-specific speech perception in the first year of life is now well attested in the literature. Importantly, although young infants show some ability to discriminate speech sounds across all languages, research findings also reveal that not all discriminable speech sounds have equal perceptual salience in infant perception. Instead, babies seem to show some distinct speech perceptual biases. One form of evidence confirming the existence of these proclivities is that infant discrimination performance for some speech sounds is strongly affected by the order of presentation of these sounds. In other words, discrimination performance would be significantly better when a vowel pair is presented in one direction (E.g. /ɪ/ to /i/) compared to the opposite direction (E.g. /i/ to /ɪ/) (Polka and Werker, 1994; Polka and Bohn, 1996; Masapollo, Polka & Molnar, 2017). This finding is well established in infant vowel perception in both within-category (E.g. distinguishing two tokens of the sound "oo" /u/) and between-category vowel discrimination (E.g. distinguishing "ee" /i/ from "ih" /ɪ/) (Polka & Bohn, 2003; 2011). Although to a lesser degree, similar findings have been found in adult perception of non-native contrasts (Schwartz & Escudier, 1989; Masapollo, Polka & Molnar, 2017; Masapollo, Polka & Menard, 2017). For instance, Schwartz et al. (1989) observed asymmetries when they explored French adult's discrimination of a continuum of synthesized vowels that were within-category variants of /e/ and this pattern seem to persist through studies targetting other vowels such as /u/ (Polka & Molnar, 2017; Masapollo et al., 2017). In sum, studies have revealed that direction of change has an impact on how effectively infants and adults distinguish speech sounds from each other. These consistent directional asymmetries in both infants and adults mark the presence of a perceptual bias which suggests that certain vowels might have a higher level of perceptual salience than others.

There are currently two theoretical models aiming to account for these directional asymmetries in vowel discrimination. The Natural Referent Vowel (NRV) framework proposes that both generic, language-universal processes and language-specific experience influence vowel perception (Polka & Bohn, 2003; 2011). During speech, the movements and positions of vocal articulators produce an acoustic signal that is shaped by the resonant frequencies of the vocal tract, also known as formants. Vowels differ in the degree to which their spectrally adjacent formants converge in frequency. When formants become spectrally close to each other, acoustic energy heightens and concentrates into a narrow spectral range. This creates a prominent bump in the speech spectrum that increases the perceptual salience of the vowel sound (Polka & Bohn, 2003; 2011). This convergence of formants is termed focalization; the most extreme levels of focalization are observed for the corner vowels /i/ (‘ee’), /u/ (‘oo’) and /a/ (“aw”); these 3 vowels are used in every human language and form the corners of an acoustic and articulatory vowel space that encompasses all possible vowel sounds. Numerous infant and adult studies show that, within this space, discrimination from a less focal vowel to a more focal vowel is easier than discrimination from a more focal to a less focal vowel (Polka & Bohn, 2011; Masapollo, Polka & Menard, 2017). According to the NRV, this generic, language-universal bias plays an important role in developing and supporting vowel perception across the life span (Polka & Bohn, 2011). Infants show this bias in the first few months of life, perhaps as early as birth, regardless of language experience. Comparable directional asymmetries in terms of direction and strength are found whether infants discriminate native or non-native vowel sounds (Polka & Bohn, 2003). As infants gain more linguistic experience in their environment during the first year of their lives, they begin to tune their perception to align with the specific vowel categories in their native language(s). This will enhance or reduce the initial generic bias due to formant focalization (Polka

& Bohn, 2011, Masapollo & Polka, 2014), depending on the specific vowels which are part of their native language repertoire. Thus, in line with the NRV, the natural referent vowel bias provides an initial scaffold that supports the acquisition of a more detailed vowel system. Overall, according to this framework, both generic/language-universal and language-specific biases operate to shape vowel perception in mature, adult speech perceivers (Polka & Bohn, 2002; 2011, Masapollo, Polka, & Molnar 2014).

Alternatively, the Native Language Magnet (NLM) framework posits for an exclusively language-experience focused explanation of vowel discrimination asymmetries. This model suggests that directional asymmetries are due to an experience-dependent bias shaped by native language (NL) “prototype” vowels (Kuhl, 1993). These NL prototypes are language-specific and consist of the highest-rated tokens in terms of goodness of each native vowel category as perceived by a native speaker. Prototypic vowels are formed through statistical learning as infants gain exposure to the language(s) in their environment, aiding their acquisition of native vowel categories (Iverson & Kuhl, 1995; Kuhl, 1991; Kuhl *et al.*, 1992; Iverson & Kuhl, 1995, 2000). Kuhl (1994) proposes that exposure to NL prototypes distorts and shrinks the perceptual space around them as if they were “perceptual magnets”. This causes vowel tokens in the area immediately surrounding a prototype to be perceptually closer to the prototype than they are physically. In consequence, it becomes more difficult to discriminate exemplars surrounding a prototype than exemplars surrounding a non-prototype. This is referred to as the Native Language Perceptual Magnet Effect. According to this theory, discrimination asymmetries arise because performance is enhanced when discriminating a change toward the prototype compared to a change in the opposite direction. However, the prototypic vowel used in Kuhl’s study (1991) also has greater formant convergence than the non-prototypic vowel, thus it is more focal. Therefore, it is unclear whether perceptual asymmetries in vowel perception reported by Kuhl (1991) are the result of intrinsic language-general focalization effects, are shaped by experience with language-specific prototypes (perceptual magnet effects), or involve both processes.

A first attempt to address this issue was made in a study conducted by Masapollo et al (2017) using an array of synthesized vowels recognized as “oo” /u/ by both Canadian-English and Canadian-French native speakers. They focused on French /u/ and English /u/ because the acoustic and production patterns for this particular vowel in these languages are distinct from each other. French /u/ is produced with greater lip-rounding and protrusion than English /u/; as a result, French /u is also more focal than English /u/ (Escudero & Polka, 2003; Martin, 2002; MacLeod et al. 2009; Noiray et al., 2011). These natural acoustic and articulatory differences provide a perfect context to assess the role of focalization and native language prototypes and to test the predictions of the NRV and NLM models. The study was composed of an identification task and a discrimination task. During the identification task, Canadian-English and Canadian-French monolingual adults first identified vowel categories (/u/,/i/,/o/, /ə/) and provided category goodness ratings (1-5) for each of the synthesized vowels varying in equal psychophysical steps along an F2 continuum across two F1 values (F1 = 275Hz & 300Hz) in the vowel space (See Figure 1; used in Masapollo et al., 2017). Among these tokens, 22 back vowels were consistency identified by members of both language groups as /u/ (85%+). Then, composite goodness score was compiled for each of this smaller array of vowels. Using the tokens that received the highest ratings according to the English listeners and French listeners respectively, two subsets as identified in Figure 1 (less-focal/English-prototypic group: stimuli u1, u2, u3 & more-focal/French-prototypic group: stimuli u5, u6, u7) were selected to test the predictions of the NRV and NLM frameworks in a subsequent AX discrimination task. Their results from both

Candian-English and Canadian-French participants revealed an asymmetry as predicted based on a focalization effect while no evidence of a perceptual magnet effect was found (Masapollo, Polka, Molnar, & Ménard, 2017).

More recently, Zhao, Masapollo, Polka, Ménard & Kuhl (2019) used similar synthesized /u/ vowel tokens to investigate directional asymmetries by assessing their neurophysiological correlates using auditory frequency-following response (FFR) measures. FFR may be tracked using EEG electrodes and originates from brain signals generated by cortical and subcortical structures in the auditory pathway (See Zhao et al., 2019). Monolingual American English-speaking adults listened to repetitions of the prototypical English /u/ and the prototypical French /u/ from the set of synthesized vowels developed in the previous study (Masapollo et al., 2017) in either an oddball or a reversed-oddball paradigm while their EEG signals were recorded. In the oddball paradigm, the French /u/ prototype (more focal token) served as the "oddball" while in the reversed-oddball paradigm, the English /u/ prototype (less focal token) served as the "oddball". They found similar results as reported in Masapollo et al. (2017); evidence of focalization effects were observed but no effects of a native language prototype were apparent. However, French adults were not tested so these results must be interpreted with caution.

Although neither Masapollo et al. (2017) or Zhao et al. (2019) provided evidence to support the native language magnet effect, this may be due to stimulus limitations. In both studies, the set of synthesized vowels used may have been unsuited for measuring prototype effects because the stimuli did not vary in small enough differences along the relevant formant dimensions to capture the prototype effect. This possibility was suggested by prior work. In Kuhl (1991), the strongest evidence of the perceptual magnet effect was found when adults discriminated small within-category differences in a narrow acoustic region very close to the prototype. In fact. no prototype effect was not observed when discriminating vowels that were more acoustically distant from the prototype. Thus, a more refined set of stimuli that vary in smaller acoustic steps around each prototype may be required to achieve a deeper understanding of the influence of innate perceptual biases and/or of language-specific experience on the perception of vowels. Therefore, new stimuli were created for use in the present study which aims to optimize the ability to measure both potential focalization and prototype magnet effects.

The goal of the current study aimed to examine whether both generic, universal biases and language-specific experience influence the perception of vowels. Precisely, we were interested in whether perceptual asymmetries in vowel perception, as observed in adults, are influenced by focalization as predicted by the Natural Referent Vowel framework (Polka & Bohn, 2004, 2011) and/or by native language experience as predicted by the Native Language Magnet model (Kuhl, 1993). In order to achieve this goal, we used a new set of /u/ vowel stimuli synthesized to include small variations in vowel quality that can allow us to measure and identify both potential effects on vowel perception. We then presented these stimuli to adult monolingual speakers of Canadian English in two consecutive tasks. First, they rated the category goodness of the stimuli to confirm the presence of a native language prototype, as was observed in previous work by Masapollo, Polka & Molnar (2017). In second task, they were asked to discriminate select pairs of /u/ variants from this new stimulus set to determine whether the emerging directional asymmetries could be predicted based on formant convergence or on language-specific prototypicality. For this thesis, we focused on testing monolingual English adults for the practical reasons that the English monolingual population is relatively more accessible compared to the French monolingual population in Montreal.

As described above, limitations of stimuli used in prior work have motivated the present study. Thus, in order to outline specific predictions that allow us to assess the role of generic focalization and language-experience effects, we need to first describe the new stimuli that were used in this study. Taking the locations of the English /u/ prototype and the French /u/ prototype observed in Masapollo et al. 2017 as a starting point, a large set of 80 new vowel stimuli systematically varying in smaller acoustic distances was created. As shown in Figure 2, this new set of synthesized vowels was created to form five orbitals surrounding good exemplars of Canadian-English /u/ and Canadian-French /u/ based on data reported in Masapollo, Polka & Molnar (2017). These target prototypes are V6 (for English) and V11 (for French). As was the case in the Masapollo et al. (2017) study, the new stimuli were created using the Variable Linear Ariculatory Model (VLAM) (Maeda, 1979, 1990; Boë, 1999; Ménard et al., 2004) These vowels around the prototypes varied in equal psychophysical steps on the Bark scale (Zwicker and Ternhardt, 1980) and differed systematically in terms of the proximity of their formants (focality) as well as their position in relation to French and English /u/ prototypes. As shown in Figure 2, the orbitals overlapped on a common vector containing the prototypic English /u/ (V6) and prototypic French /u/ (V11) tokens identified in Masapollo et al (2017). Twelve tokens along this vector (as identified on Figure 2 from V3 to V14), were selected for this study. These variants were selected to include both the French (more focal) and English (less focal) /u/ prototypes as well as three additional variants on each side of the prototypes. Thus, in the selected final set, ranging from V3 to V14 (See Figure 3), the /u/ vowels along this vector become more focal as token number increases. Prior research (Masapollo, Polka & Molnar, 2017) confirmed that all of the selected stimuli were within the region that was consistently identified as /u/ by monolingual English and by monolingual French adults.

Hypotheses

As mentioned previously, past work has shown an effect of focalization in the discrimination of /u/ vowels in adult monolingual speakers of English and French while no indication of the perceptual magnet effect was found (Masapollo et al,, 2017; Zhao et al., 2019). With the select stimuli from the new corpus we should now be able to assess the potential effects proposed by both theories. Overall, we hypothesize that discriminating larger within- and between-category differences will reveal asymmetries due to focalization effects while discriminating smaller within-category differences in the immediate area surrounding each prototype will reveal asymmetries shaped by experience with language-specific categories.

More specifically, for English adults, we predict that if focalization is the only factor influencing vowel perception, then discriminating a change from a less focal vowel to a more focal vowel will be easier than the reverse direction at any point along the series (from V3 to V11). In fact, this asymmetry would be observed regardless of the native language of the participants.

Alternatively, if both focalization and native-language magnet effects play a role, then, for English adults, directional asymmetries based on focalization will be observed when discriminating pairs across the series (higher performance when direction of change is from less-focal to more-focal), but the direction of the asymmetry will reverse in the region proximal to the native language prototype (higher performance when the direction of change is from more-focal to less-focal). Thus focalization-based asymmetries will be observed above and below the prototype region and prototype-based asymmetries will be observed for tokens close to the prototype. For example, if the English prototype is near V6, better discrimination may be observed for changes from a more-focal vowel to a less-focal vowel in that region (E.g. discriminating from V9 to V6 will be easier than from V6 to V9) and focalization based asymmetries will be observed for stimulus pairs below V6 (E.g. V3 to V6 easier than V6 to V3) and above V9 (E.g. V9 to V11 easier than V11 to V9.)

Importantly, discrimination performance will also be modulated by the acoustic distance between vowels within a pair, i.e., pairs that are 3 psychophysical steps apart on the continuum will be easier to discriminate than pairs that are 1 psychophysical step apart. Moreover, acoustic distances that are too small may be too difficult and result in floor effects and acoustic distances that are too large may be too easy and result in ceiling effects. Either extreme would make it difficult to detect focalization or prototype effects. For this reason, we will vary the number of steps between vowels within discrimination pairs to assess a range of intervals along the continuum. This will increase the opportunity to isolate both focalization and prototype magnet effects.

Methods

Participants

Twenty adult monolingual North-American English speakers were included in the study. The mean (SD) age was 20 (1.5) years and included 4 males. All participants were recruited through social media (i.e. Facebook), online advertising websites (i.e. McGill classified ads) and direct advertising in local English universities (i.e. McGill University, Concordia University) in the Greater Montreal Area. Three additional participants were tested but excluded from the final sample analysis due to failure to meet language inclusion criteria as described below (2) and failure to follow task directions (1). A short online language screening questionnaire was sent to potential participants to determine their eligibility before they were invited to take part in the study. A more extensive language questionnaire, the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumfield & Kaushanskaya, 2007) was also administered on the testing day to verify their language background. Only data from subjects who meet the following language criteria according to both questionnaires were included in the analysis:

i) The participant is a native speaker of English with minimal to no experience with other languages.

ii) Both of the participant's parents are native speakers of English.

iii) The participant has not received linguistics or phonetics training.

iv) The participant does not have a history of speech or hearing disorders.

v) The participant has only received education at monolingual schools in English.

vi) The participant did not receive formal instruction of a second language prior to 10 years of age.

vii) The participant isn't actively learning any languages currently.

viii) The participant does not use a second language on a regular basis.

Stimuli

Twelve vowel stimuli were selected from the larger set of synthesized tokens described above. Each stimulus was 400ms in duration (see introduction for more details).

Experimental design and procedure

Each participant was tested individually in a quiet testing laboratory at the McGill School of Communication Sciences and Disorders. The experiment lasted approximately 1 hour in duration and the participants were compensated for their time. After a verbal overview of the study given by the experimenter, participants were invited to fill out the consent form as well as the

LEAP-Q (Marian *et al.*, 2007) to confirm their language background. The experimental portion of the study included two perceptual tasks completed in succession during a single session. A category goodness rating task was carried out first to verify that the items varied as expected with respect to perceived goodness of fit to the listener's native-language /u/ vowel category as well as to verify the approximate position of their native vowel prototype. Following the rating task, the same participants completed an AX (same/different) discrimination task (Iverson & Kuhl, 1995). The data were collected using SuperLab (Version 5; Cedrus Corporation; San Pedro, CA). Instructions were both given verbally at the beginning of the study and displayed at the screen during the task. All auditory stimuli were presented through headphones connected to the testing computer at a pre-set volume deemed comfortable by the participant prior to the beginning of the trials. Each task is described in more details in the sections below. *Note: for simplicity we will describe the discrimination task first followed by the identification task. However, please keep in mind that the subjects actually performed the task in the reverse order (i.e. The category goodness rating task was completed first followed by the discrimination task). The results section will present each task following the order that they were completed by the subject.*

Discrimination task

In the discrimination task, the participants completed 288 self-paced AX discrimination trials. The trials were organized in two experimental language blocks of 144 trials each preceded by 6 practice trials. An interstimulus interval of 1500ms was used. The two blocks of AX items were created based on the locations of the English /u/ prototype (V6) and French /u/ prototype (V11) reported in Masapollo et al. (2017); for simplicity we will refer to them as the English Prototype block and the French Prototype block.

The English Prototype block included “different” pairs in which the English prototype (V6) was paired with each nearby variant located one to three steps away on either side of it for a total of 6 pairings ranging from V3 to V9 — V3-V6, V4-V6, V5-V6, V6-V7, V6-V8, V6-V9 (See Figures 2 & 3). In addition, “same” pairs, built from every variant used in the “different” pairs from the English Prototype block paired with itself (V3-V3, V6-V6, V8-V8 etc.), were also included.

The French Prototype block was created following the same structure. It included “different” pairs in which the French prototype (V11) was paired with each nearby variant located one to three steps away on either side of it for a total of 6 pairings ranging from V8 to V14 — V8-V11, V9-V11, V10-11, V12-V11, V13-V11 V14-V11 (See Figures 2 & 3). “Same” pairs were also included. These were built from every variant used in the “different” pairs from the French Prototype block paired with itself (V9-V9, V10-V10, V11-11 etc.).

Prior to each block, there were 6 practice trials to familiarize the participants with the task as well as the range of stimuli. The practice trials included 3 same trials and 3 different trials. Participants did not receive feedback regarding their performance at any point during the task; short breaks were granted at specified progress points between trials if needed. On each trial, subjects heard a pair of vowels presented through headphones. The interval between the first and second vowel within each pair was 1500 ms. This relatively long ISI was used to create memory demands that tap into phonetic processing (Werker and Logan, 1985; Cowan and Morse, 1986; Repp and Crowder, 1990; Masapollo, Polka & Molnar, 2017).

Across language blocks, an equal number of “different” and “same” pairs were presented to the participants during trials. Each “different” pair (e.g.V3-V6) was presented 12 times per block, 6 times with the tokens in one order (e.g V3-V6) and 6 with tokens in the reverse order (e.g. V6-

V3). The “same” pairs were composed of the any token within the array repeated twice while the different pairs were composed of either the English /u/ prototype or the French /u/ prototype as observed in Masapollo et al. (2017) and another token in the surrounding area at a distance of 1, 2 or 3 steps away as outlined above. Each “same” pair (e.g. V3-V3) was presented 12 times per block with the exceptions of V8-V8 and V9-V9. The V8 and V9 tokens formed 2-steps and 3-steps pairs with each of the prototype tokens (e.g. V6-V8, V6-V9, V8-V11, V9-V11) and thus belonged to both the English and French blocks. To ensure that the “same” pairs formed by these tokens were not overrepresentated compared to other pairs, they were only shown 6 times per block (for a total of 12 times across language blocks).

Participants were instructed to listen to the pair of vowels on each trial and indicate, via keypress, whether they perceive the two vowel tokens to be the same or different. No feedback was provided about their performance at any point during the task. The orders of presentation for trials as well as for Prototype blocks (English; French) were randomized across participants. Overall, the AX pairs were created to test discrimination asymmetries in four experimental regions (A, B, C, D) along the vector on which they were situated (See Figure 4). Each region was comprised of four variants including either the English or the French /u/ prototype ; within each region, the stimuli were used to form AX pairs that included the target prototype (English V6 or French V11) and each of the 6 variants in close proximity to it (3 on the left and 3 on the right).

Category goodness rating task

During the goodness rating task, the participants heard the exact same vowel pairs that they were presented in the AX discrimination task (i.e. in two experimental blocks of 144 trials each preceded by 6 practice trials). Across both testing blocks, participants were exposed to an equal number of “same” and “different” vowel pairs (ISI: 1500ms). Depending on the condition to which the participant was assigned, the subject was instructed to respond only to the first or only to the second vowel in each pair throughout the task. Both the order of presentation for the blocks as well as the assigned target vowel to be rated (1st or 2nd) were counterbalanced across participants. On each trial, the participants were asked to judge whether the specified target vowel they heard corresponds to the sound “oo” /u/ as in the word “boo” and instructed to provide a category goodness rating for the designated vowel (by pressing the appropriate number) on a scale of 1 to 7, where 1 is a very poor exemplar of the target vowel and 7 is a very good exemplar of the target vowel, as produced by a native speaker. If the participants judged the vowel to be a sound other than /u/, they were asked to indicate this by pressing the “N" key.

This paradigm, adapted from Lotto, Kluender and Holt (1998), was chosen over the typical vowel identification and rating tasks where vowels are presented in isolation, as it is well known that contextual sounds may influence the perception of vowels (Eimas, 1963; Nearey, 1989). Since the discrimination AX paradigm requires presentations of vowel pairs, using the same format in the rating task provides consistency across the rating and discrimination tasks. In fact, the presentation format used in the identification task also fully matches the format used in the discrimination task as described above. Essentially, subjects heard the same vowel tokens presented the same number of times in the same token pair format in both tasks. This was done to minimize effects of stimulus context across the two tasks.

Results

Category goodness rating task

Participants’ ratings of the selected 12 /u/ tokens (V3-V14) were analysed to assess the position of our predicted English /u/ prototype. As stated above, subjects rated either the first or the second token of pairs of vowels presented in succession on a goodness scale of 1 to 7. If a vowel was not perceived as /u/, they were instructed to press the "N" key[1]. Each stimulus was rated at least 18 times, the target prototypes (V6, V11) as well as the tokens which were part of both language blocks (V8, V9) were rated more frequently as they appeared in a higher number of trials. For each subject, a mean category goodness rating was calculated for each vowel variant. Trials where the stimulus was identified as other (by selecting "N") received a score of 0. The group median of the average rating scores from each subject was determined. Figure 4 shows a plot of each stimulus in an F1 x F2 vowel space. Each circle corresponds to a vowel stimulus and the size of the circle is scaled to represent the magnitude of the median rating score. The median rating score is indicated on each circle and the number below the median rating shows the number of subjects (out of 20) who gave that stimulus their highest average goodness rating.

Although the V6 stimulus received the highest goodness rating in previous research by Masapollo et al. (2017), the location of the best exemplar of /u/ in the current study was shifted in the direction of focality. The highest median group rating scores were found closer to the center of our stimulus array. The English prototypical /u/ appears to be in the area around V8 (median rating score: 4.9) instead of V6 (Median rating score: 4.6) as it was previously found. In addition, 8 participants out of 20 designated the middle variant V8 (F1: 289.75Hz; F2: 881.20Hz) to be the best exemplar of English /u/. Nevertheless, although the "prototype" identified in these ratings is more focal than the original English prototype V6, it is still less focal than the V11 token that corresponded to the French prototype in Masapollo et al (2017).

Recall that participants rated the /u/ variants within different types of AX pairs. The ratings were analyzed separately for trials where the rated stimuli were presented within "same" pairs (E.g. V7-V7) and when they were presented within "different" pairs (E.g. V6-V8). Our results show

[1] This option was used by participants on less than 0.06% of all test trials, confirming that all selected tokens were consistently judged to be within the correct vowel category /u/ according to monolingual Canadian-English speakers.

similar group median score patterns across these two contexts (See Figures 4 & 5), confirming that the shifted position of the English /u/ prototype is robust. However, it is worth noting that there seems to be more variability in the goodness ratings when the rated stimuli were presented in "different" pairs. This is not surprising given that phonetic context plays a role in vowel perception (Lotto et al., 1998).

Discrimination task

For each different stimulus pair, we computed an A' score to index the accuracy of each subject in discriminating that vowel pair. A' scores were used to ensure that discrimination patterns were not the result of a response bias to select either "Same" or "Different". Individual scores were converted to *A'* scores using the following formula: $A'=0.5+(H-FA)(1+H-FA)/[(4H1-FA)]$ where H is the proportion of hits and *FA*, the proportion of false alarms[2].

Recall that we predicted that discrimination performance will change across the series showing asymmetries determined by focality when discriminating stimuli near the ends of the series and asymmetries determined by prototype effects for pairs near the perceiver's native language prototype. To test these effects, accuracy was examined across four regions of the series identified in Figure 3; Region A corresponds to pairs involving stimuli V3, V4 ,V5 and V6; region B corresponds to pairs involving V6, V7, V8, and V9; Region C corresponds to pairs involving V8, V9, V10 and V11 and Region D corresponds to pairs involving V11, V12, V13, and V14 (See Figure 3).

All statistical analyses for the discrimination task were performed using the *lme4/car* libraries in R (Bates, 2010) and the data plots were made using the *plyr* package. In order to assess the effects of focalization and prototype, A' scores were submitted to four ANOVAs with two within-subjects' factors: Region (A, B, C, D) and Direction of change (Toward Focality/Against Focality). Recall that for regions A and B, each stimulus was paired with V6 to form pairs that were 1 step away, 2 steps away and 3 steps away from V6. Likewise, each stimulus within Regions C & D was paired with V11 to form pairs that were 1 step away, 2 steps away and 3 steps away from V11. The first ANOVA assessed accuracy only on the 1-step pairs in each region. The 2nd

[2] Similar patterns of results were observed when raw percent correct was used as the dependent measure instead of A' scores.

ANOVA assessed accuracy only on 2-step pairs in each region; the 3rd ANOVA assessed accuracy only on 3-step pairs in each region. The 4th ANOVA assessed overall accuracy on all pairs (1-,2- and 3-step combined) within each region. Analyses were conducted with different step sizes to determine if effects were masked by ceiling or floor effects related to overall acoustic distance between pairs.

Recall that if focality alone has an effect, we expect to find a direction of change effect showing better performance when direction of change is towards focality (i.e. discriminating toward a stimulus with a higher number in the series [E.g. V3-V6]) across all regions. However, if prototype effects also occur, the direction of change should reverse at the specific region close to the prototype. Assuming V6 is the prototype, we expected the reverse to occur at region B; however, given that the prototype appears to be closer to V8 in our rating data, the reverse would be more likely to occur in region C. Note that "toward focality" or "against focality" indications in the results section simply pertains to the general Direction of Change in each Region and are not necessarily reflecting pure focalization effects as changes toward the prototype (V8) are also in the direction of focality in Regions A and B. More explanation differentiating the effects of focalization and prototype will be provided in the discussion section.

1-step pairs

The ANOVA for 1-step pairs (See Figure 7a) showed a significant main effect of Region [$F(3,57)=8.824$, $p<0.001$ ***] and a significant interaction between Region and Direction of Change [$F(3,57)=9.843$, $p<0.001$ *** $d_{unb} = 0.254$]. No significant effect of Direction of Change was found [$F(1,19)=2.747$, $p=0.114$ NS].

Simple main effect analyses of direction were a performed for each region to assess asymmetries across the regions. A significant effect of direction was found in regions C and D (A:

[$F(1,19)=0.818$, $p=0.377$ NS $d_{unb} = 0.291$]; B: [$F(1,19)=2.374$, $p=0.140$ NS $d_{unb} = 0.742$]; C: [$F(1,19)=4.886$, $p=0.040$ $d_{unb} = 0.538$]; D: [$F(1,19)=6.821$, $p=0.017$ $d=0.616$ $d_{unb} = 0.592$]). As shown in Figure 7b, participants were more accurate at discriminating changes in the direction of focality in regions A and B but these effects were not significant. In regions C And D, there were significant directional asymmetries in the opposite direction, showing better discrimination for changes toward our prototype (against focality) (See Table 1).

2-step pairs

The analysis of 2-step pairs (See Figure 8a) also showed main effects of Region [$F(3,57)=8.102$, $p<0.001$ ***] and Direction of Change [$F(1,19)=6.565$, $p=0.019$ *] as well as a significant interaction between Region and Direction of Change toward focality [$F(3,57)=5.258$, $p=0.003$ ** $d_{unb} = 0.222$].

Simple main effects analyses of Direction revealed significant differences in Regions B and D and results approaching significance in Regions A and C (A: [$F(1,19)=3.270$, $p=0.086$. $d_{unb} = 0.372$]; B: [$F(1,19)=14.670$, $p=0.001$ ** $d_{unb} = 0.589$]; C: [$F(1,19)=3.443$, $p=0.079$. $d_{unb} = 0.338$]; D: [$F(1,19)=4.999$, $p=0.038$ * $d_{unb} = 0.547$]). As shown in Figure 8b, listeners were more accurate when differentiating in the direction of focality in Regions A, B and D. In Region C, accuracy was higher when listeners discriminated in the opposite direction (See Table 1).

3-step pairs

Note the discrimination was higher (approaching ceiling) for 3-step pairs. Nevertheless, for 3-step (See Figure 9a) data, the ANOVA showed main effects of Region [$F(3,57)=4.226$, $p=0.009$ **] and Direction of Change toward focality [$F(1,19)=5.296$, $p=0.033$ *] as well as a significant interaction between Region and Direction of Change [$F(3,57)=4.787$, $p=0.005$ ** $d_{unb} = 0.296$].

Simple effects analyses of Direction revealed that differences in Regions B and D were significant (A: [F(1,19)=1.314, p=0.266 NS d_{unb} = 0.336]; B: [F(1,19)=10.420, p=0.004 ** d_{unb} = 0.654]; C: [F(1,19)=1.789, p=0.197 NS d_{unb} = 0.308]; D: [F(1,19)=8.06, p=0.011 * d_{unb} = 0.807]). As shown in Figure 9b, higher discrimination accuracies were found when participants discriminated from a less focal vowel to a more focal vowel in Regions B and D (See Table 1).

All steps combined

Finally, an overall analysis performed for all psychophysical steps collapsed revealed significant a main effect of Region [F(3,57)=12.450, p<0.001 ***]. As expected, there was also a significant interaction between Region and Direction of Change [F(3,57)=5.009, p=0.004 ** d_{unb} = 0.373]. No significant effect of Direction of Change was found [F(1,19)=0.673, p=0.422 NS].

Simple effects analyses of Direction showed significant differences in two out of four regions and differences approaching significance in one region (A: [F(1,19)=4.202, p=0.054 . d_{unb} = 0.462]; B: [F(1,19)=11.34, p=0.003 ** d_{unb} = 0.950]; C: [F(1,19)=7.428, p=0.013 * d_{unb} = 0.724]; D: [F(1,19)=0.184, p=0.673 NS d_{unb} = 0.618]). Compared to the results of the stepwise analyses, similar patterns were found when the distances are collapsed. As shown in Figure 10, listeners were more accurate when discriminating in the direction of focality in Region B and showed a trend approaching significance towards focality for region A, while discriminating in the opposite direction is easier in Region C (See Table 1).

Discussion

Category Goodness Rating Task

As expected, the full range of selected synthesized variants was reliably identified as the vowel /u/ by our participants, but participants showed variability in terms of the median goodness score assigned for each variant, suggesting the presence of an internal language experience dependent "ranking system". This data pattern is consistent with previous findings from Masapollo et al. (2017) as well as the original study supporting the NLM model by Kuhl (1991). In fact, the ability to distinguish these subtle differences in category goodness is now well established in prior work (Kuhl, 1991; 1993) and clearly plays a role in supporting the efficient perception and production we display in our native language.

The position of our English prototype did not correspond to V6 (Median Rating Score: 4.6) as we have previously observed in results from Masapollo et al. (2017). Instead, it shifted to the center of the array, between the original English /u/ prototype V6 and the original French /u/ prototype V11 reported in Masapollo et al. (2017). In fact, the highest median goodness scores were found for V8 (Median Rating Score: 4.9) as well as the variants immediately surrounding it (Median Rating Score for V7 = 4.8; Median Rating Score for V9: 4.7). Moreover, 8 participants out of 20 rated the middle variant V8 (F1: 289.75Hz; F2: 881.20Hz) as the best exemplar of English /u/, for this stimulus array. Sub-analyses made on median goodness scores of "Same" and "Different" trials also revealed similar findings (See Figures 4 & 5).

There are multiple reasons why the prototype suggested by our data did not fully correspond to the one found in the Masapollo et al. (2017) study. First, the effects of the immediate acoustic context in vowel perception are well established in the literature (Lotto et al., 1998; Werker and Logan, 1985; Cowan and Morse, 1986; Repp and Crowder, 1990; Masapollo, Polka & Molnar, 2017). As our participants were instructed to make category goodness judgements for target vowels presented within vowel pairs, their ratings might be influenced by their exposure to non-target vowels as well. Second, dialect differences between our English participants and the English participants in Masapollo et al. (2017) may also play a role. In the previous study, the English adult participants were people from Ontario receiving university education in Montreal. They were exposed to French during their university years while living in Montreal, but had limited direct and/or regular exposure to French during their childhood. In the present study, 11 out of 20 of our participants were Montreal-natives who grew up in monolingual anglophone families and received education through the English school system. However, Montreal is a predominantly French city. In consequence, they were inevitably exposed to the French language to a certain extent in their day-to-day lives even though they did not actively use it. As the perceptual magnet effect is subject to dialectal variation (Frieda, Walley, Flege & Sloane, 1999), the prototype shift toward focality might be due to differences in the regional variation in English. Production data from Boberg (2008) suggest that the mean F1 values across linguistic contexts for /u/ are relatively lower for English speakers from Quebec (F1 = 567Hz), the Maritimes (F1 = 541Hz) as well as the Prairies (F1 = 564Hz) compared to other regions of Canada (E.g. British Columbia: F1 = 619Hz). Not surprisingly, these areas included provinces with considerable amount of French presence (Statistics Canada, 2016). This might play into our participants' preference of a more focal English /u/ token as it also has a higher F1 value compared to the original /u/ found in Ontario-natives in Masapollo et al. (2017). In fact, past literature show that Canadian-French /u/s are generally produced with both lower F1 and F2 values compared to English vowels. Because of this higher level of formant convergence, the French /u/ is also more focal (Masapollo et al., 2017; Escudero & Polka, 2003; MacLeod et al., 2009; Noiray et al., 2001). Thus, it is possible that the English /u/ prototype has shifted to a more focal position from the exposure to French /u/s in the participants' environment. Though, while the English /u/ prototype observed in this study is more focal than the one observed in Masapollo et al. (2017), it is still significantly less focal than the French prototype they report.

Discrimination Task

As detailed in the section above, the NRV model proposes that both universal, language-general biases shape directional biases in vowel perception while the NLM model proposes an exclusively language experience-based approach to explaining these perceptual asymmetries. According to the NRV model (Polka & Bohn, 2011), listeners will have more ease at discriminating from a less-focal to a more-focal vowel. In contrast, according to the NLM model (Kuhl, 1991), listeners would perform better at discriminating from a less-prototypic to a more-prototypic vowel. Contrary to the findings of Masapollo et al. (2017) and Zhao et al. (2019) which revealed the effects of focalization alone, our discrimination findings show that both prototype and focalization effects shape directional asymmetries in vowel perception.

Recall that if only vowel formant proximity played a role in producing perceptual biases, regardless of the Region in which the pair is situated, participants would perform better if they heard a vowel token with less formant convergence followed by a vowel with more formant convergence (from a less-focal to a more focal vowel) compared to if the direction of change was reversed. Significant effects of focalization were apparent in all levels of analysis (see Figures 6-9). However, the 1-step and overall analyses revealed a significant reversal in the discrimination asymmetry in the Region C [$F(1,19)=4.886$, $p=0.040$] where the current English /u/ prototype (V8) was located, suggesting that prototype effects were at also at play in this particular area of the synthesized vowel distribution. In addition, the 2-steps analysis for Region C [$F(1,19)=3.443$, p=0.079 .] also shows results approaching significance in support of this reversal effect. Specifically, 1-step analysis for Region C pertained to the discrimination performance for the V10-V11 pair. The V10 token is very close to the best rated exemplars V8 (Median Rating Score: 4.9) and V9 (Median Rating Score: 4.7). Although none of the pairs in this study were purposefully designed to target the V8-V9 region due to the original assumption that the English prototype would be located around V6, Region C provided a favorable environment for the prototype effect to manifest. Going from V10-V11 would be comparable to going from a prototypic vowel to a non-prototypic vowel and the reverse direction (V11 to V10) would be comparable to going from a non-prototypic vowel toward a prototypic vowel. Consistent with the predictions of the NLM framework (Kuhl, 1991), discriminating from V11 to V10 yielded better performance compared to the opposite direction. This trend remained significant [F(1,19)=7.428, p=0.013 *] when data from all stepwise comparisons in the Region C were collapsed in the overall analysis, providing further evidence for the presence of native language magnet effects. As for the 3-step data, very high accuracy rates (nearing 90% correct) could be observed for pairs in the first 3 Regions of analysis (A, B, C). The presence of this strong ceiling effect might suggest that the pairs are quite easily discriminated and thus, there is a possibility that the prototype effect could not be accurately measured.

To continue, 2-steps and 3-steps data also seemed to show a reversal directional pattern back to favoring focalization in Region D, providing support for the assumption that language magnet effects might only be detectable in the region immediately surrounding the prototype (Kuhl, 1991). While a significant effect in the reverse direction (toward the prototype) was present in the same Region for the 1-step data [F(1,19)=6.821, p=0.017], A’ scores suggest possible presence of a floor effect as the participants' performance levels were at chance (See Figure 7b). Thus, these results might not accurately reflect the predictions of the theories as the task seems too difficult.

Moreover, it is also interesting to note that consistently more robust effects in the direction of focality were observed in Region B for all levels of analysis except 1-step compared to the other regions. This might have been caused by the fact that this particular region is affected by both the NRV and NLM frameworks in the same direction. In fact, considering the proximal position of the vowel tokens in Region B in relation to the English prototype V8, the predictions of both frameworks would suggest higher performance when discriminating pairs in the direction of focality. Precisely, Region B includes V6-V7, V6-V8 and V6-V9 pairs, which means that when participants are discriminating in the direction of the prototype V8, they are also discriminating in the direction of focality and vice-versa. Consistent with this prediction, data from 2-step, 3-step and overall analyses in Region B all showed strong effects favoring performance the direction of focality, providing further support for the presence of prototype effects.

General Discussion

The present study investigated the effects of language-general focalization described by the NRV framework (Polka & Bohn, 2011) and language-experience based prototypes described by the NLM framework (Kuhl & Iverson, 1995; Kuhl et al., 2008; Kuhl, 1991) on directional asymmetries. Specifically, we examined the effects of formant proximity and focality on the within-category speech sound discrimination abilities of monolingual Canadian speakers of English. In contrast to the previous experiments conducted by Masapollo et al. (2017) and Zhao et al. (2019) investigating these perceptual biases using an expanded array of synthesized /u/ stimuli. Our select stimuli feature a set of tokens systematically varying in smaller psychophysical steps along a vector containing both the English and French prototypes. As native language magnet effects are often strongest in the area immediately surrounding the prototype (Kuhl, 1991), we believed that using a finer grained stimulus array would allow us to better detect the presence of these effects which were absent in previous work (Masapollo et al., 2017; Zhao et al., 2019).

In the first experimental task, monolingual English listeners rated (from 1-7) the full array of /u/ variants in terms of category goodness presented inside paired vowel trials. All selected /u/ tokens were rated as being members of the /u/ vowel category fairly consistently. In terms of category goodness, the token with the highest rating for was found to be more central (V8; Median Goodness Score: 4.8) in the distribution compared to the original English /u/ prototype (V6; Median Goodness Score: 4.6) from Masapollo et al. (2017). In addition, our goodness ratings were also not as contrastive when comparing each variant with one another, we expected this due to the fact that we are comparing a range of stimuli that are much closer to each other in acoustic and perceptual space compared to similar prior work. In sum, the results from the Category Goodness Rating Task provided support for the suitability of the selected array of /u/ tokens for showing the effects of both focalization and prototype.

In the second experiment, the same participants took part in an AX discrimination task composed of the same pairs of vowels rated in the previous task. This allowed the participants to complete both tasks in the exact same linguistic contexts. Results show consistent prototype biases across all levels of analysis, specifically in the immediate regions (B, C) surrounding the prototype in specific stepwise comparisons in addition to the focalization effects observed in 2-steps, 3-steps and overall analyses. When discriminating between a non-prototypic variant and a prototypic variant (V8) which has a higher level of formant convergence (more focal) compared to the non-prototype, the directional asymmetry favoring change from a less-focal to a more-focal token is simply enhanced as both prototype and focalization effects dictate change toward the same direction. In contrast, when discriminating between non-prototypic variant and a prototypic variant (V8) which has a lower level of formant convergence (less focal) compared to the non-prototype, prototype effects will dominate the effects of focalization and reverse the directional asymmetry, favoring a change from a less-prototypic to a more prototypic vowel instead. In brief, this experiment provides a first behavioral account of interactional effects between language-general focalization effects and language-experience dependent prototype effects in explaining directional asymmetries in vowel perception.

The discrimination findings also reveal that it is challenging to isolate focalization and prototype effects within the same stimulus array. Focalization effects appear to be more apparent and easier to isolate compared to prototype effects in most regions of analysis while prototype effects can only be detected in a very restricted region immediately surrounding the prototype. It seems that to measure the prototype effect we need to know the precise location of the native language prototype for a particular stimulus array and testing context. As well, it is optimal to measure these effects within the same subjects as we have done in the present study. These findings may help explain why the NLM effect has been difficult to replicate across different studies in the past.

One limitation of the present study was the decision to design our discrimination task based on the English and French prototypes (V6, V11) found in Masapollo et al. (2017), thus the pairs were designed to include directional changes going toward and against these prototypes rather than the prototype identified from our category goodness rating data (V8). In consequence, there were no specific pairings designated to target directional changes directly from and toward the V8 prototype. Although evidence of the NLM theory was found, the effects of certain areas surrounding the prototype remain unknown (E.g. V8-V9; V8-V10) due to this limitation and would require more investigation. In addition, we were unfortunately not able to include data from monolingual French populations in this paper due to practical constraints, however French data would be useful in providing more insight on the interaction between prototypicality and focality on vowel discrimination performance as the French prototype is located in a region of high focality, thus prototype effects might be less strong.

Finally, the present research contributes to the understanding of existing theoretical models of speech perception specifically pertaining to the role of language-general and language-specific biases in shaping directional asymmetries. Aside from contributing further evidence for the NRV framework (Polka & Bohn, 2011) stipulating that the perceptual salience of focal vowels play a role in speech perception (Zhao et al., 2019; Masapollo et al., 2017; Schwartz, Abry, Boë, Ménard & Vallée, 2005 etc.), our current results also provided novel evidence in support of the experience-dependent predictions of the NLM framework, suggesting that both theories outline factors that influence vowel perceptual biases. Precisely, our data show that performance of Canadian English listeners during vowel discrimination tasks typically increases when the direction of change is from a less-focal to a more-focal token. However, when these tokens are also in close proximity to the English /u/ prototype, the native language magnet effect either overrides focalization processes when it acts against the direction of focality or enhance them to create more robust focalization effects when it acts in the same direction.

References:

Boberg, C. (2008). Regional Phonetic Differentiation in Standard Canadian English. *Journal of English Linguistics,36*(2), 129-154. doi:10.1177/0075424208316648

Cowan, N., & Morse, P. A. (1986). The use of auditory and phonetic memory in vowel discrimination. *The Journal of the Acoustical Society of America,79*(2), 500-507. doi:10.1121/1.393537

Eimas, P. D. (1963). The Relation between Identification and Discrimination along Speech and Non-Speech Continua. *Language and Speech,6*(4), 206-217. doi:10.1177/002383096300600403

Escudero, P., & Polka, L. (2003). A cross-language study of vowel categorization and vowel acoustics. In Sole, M.J., D. Recansens and J. Romero (eds.) *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics,61*(3), 561-577. doi:10.3758/bf03211973

Government of Canada, Statistics Canada. (2019, February 20). Language Highlight Tables, 2016 Census. Retrieved from https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/lang/Table.cfm?Lang=E&T=11&Geo=00

Iverson, P., & Kuhl, P.K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97, 553–562.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. Perception & Psychophysics, 50, 93–107.

Kuhl, P. K. (1994). Learning and representation in speech and language. Current Opinion in Neurobiology, 4, 812–822.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science,9*(2). doi:10.1111/j.1467-7687.2006.00468.x

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *The Journal of the Acoustical Society of America,103*(6), 3648-3655. doi:10.1121/1.423087

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guide. New York, NY, US: Cambridge University Press.

Marian, V., Blumfeld, H. K., and Kaushanskaya, M. (2007). “Language Experience and Proficiency Questionnaire (LEAP-Q),” Journal of Speech, Language & Hearing Research 50(4), 940–967 (2007).

Masapollo, M., & Polka, L. (2014). Asymmetries in vowel perception: Do they arise from focalization, perceptual magnets, or both? *The Journal of the Acoustical Society of America,135*(4), 2224-2224. doi:10.1121/1.4877267

Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *The Journal of the Acoustical Society of America,141*(4), 2857-2869. doi:10.1121/1.4981006

Ménard, L., Schwartz, J., & Boë, L. (2004). Role of Vocal Tract Morphology in Speech Development. *Journal of Speech, Language, and Hearing Research,47*(5), 1059-1080. doi:10.1044/1092-4388(2004/079)

Nearey, T.M. (1989): 'Static, dynamic, and relational properties in vowel perception', Journal of the Acoustical Society of America 85, 2088-2 1 13.

Polka, L., and Bohn, O. (1996). “A cross-language comparison of vowel perception in English-learning and German-learning infants,” Journal of the Acoustic Society of America, 100, 577–592.

Polka, L. & Bohn, O. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221-231

Polka, L., & Bohn, O. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39, 467-478.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance,20*(2), 421-435. doi:10.1037/0096-1523.20.2.421

Repp, B. H., & Crowder, R. G. (1990). Stimulus order effects in vowel discrimination. *The Journal of the Acoustical Society of America,88*(5), 2080-2090. doi:10.1121/1.400105

Schwartz, J., Abry, C., Boë, L., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion–Focalisation Theory. *Speech Communication,45*(4), 425-434. doi:10.1016/j.specom.2004.12.001

Schwartz, J., & Escudier, P. (1989). A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception. *Speech Communication,8*(3), 235-259. doi:10.1016/0167-6393(89)90004-6

Tsushima, T., Takizawa, O., Sasaki, M., Shiraki, S., Nishi, K., Kohno, M., Menyuk, P., and Best, C. (1994). “Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months:

Language-specific developmental changes in speech perception abilities," in *Proceedings of the International Conference on Spoken Language Processing*, Acoustical Society of Japan, Vol. 4, pp. 1695–1698.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development*, 7(1), 49-63. http://dx.doi.org/10.1016/S0163-6383(84)80022-3

Zhao, T. C., Masapollo, M., Polka, L., Ménard, L., & Kuhl, P. K. (2019). Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response. *Brain and Language,194*, 77-83. doi:10.1016/j.bandl.2019.05.002

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America,68*(5), 1523-1525. doi:10.1121/1.385079

Appendix: Figures and tables

Figure 1. Figure adapted from Masapollo et al. (2017); The figure at the left illustrate the range of stimuli originally tested in the identification experiment, all tokens were at equal psychophysical steps from each other. The figure at the right presents the 22 tokens which were consistently identified as /u/ by both Canadian-English and Canadian-French listeners. It also shows the position of the two sets of 3 stimuli used in the discrimination task selected from the highest rated tokens for each language group. The less-focal/English-prototypic set included tokens u1-u3 and the more-focal/French-prototypic included tokens u4-u6 (See more explanation in text; For additional details, see Masapollo et al. (2017).

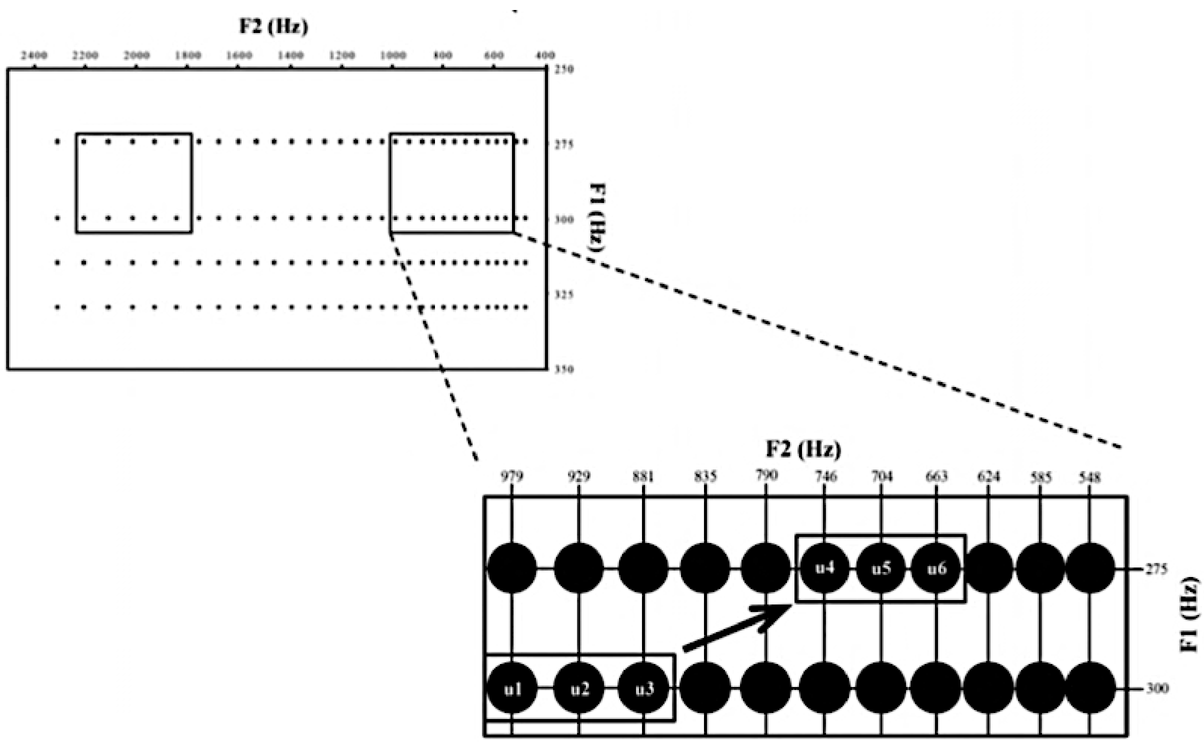

Figure 2. Selected vowel tokens on the common vector of synthesized /u/ stimuli around the English (V6) and French /u/ (V11) prototypes in a vowel space (mels) generated using the Variable Linear Articulatory Model.

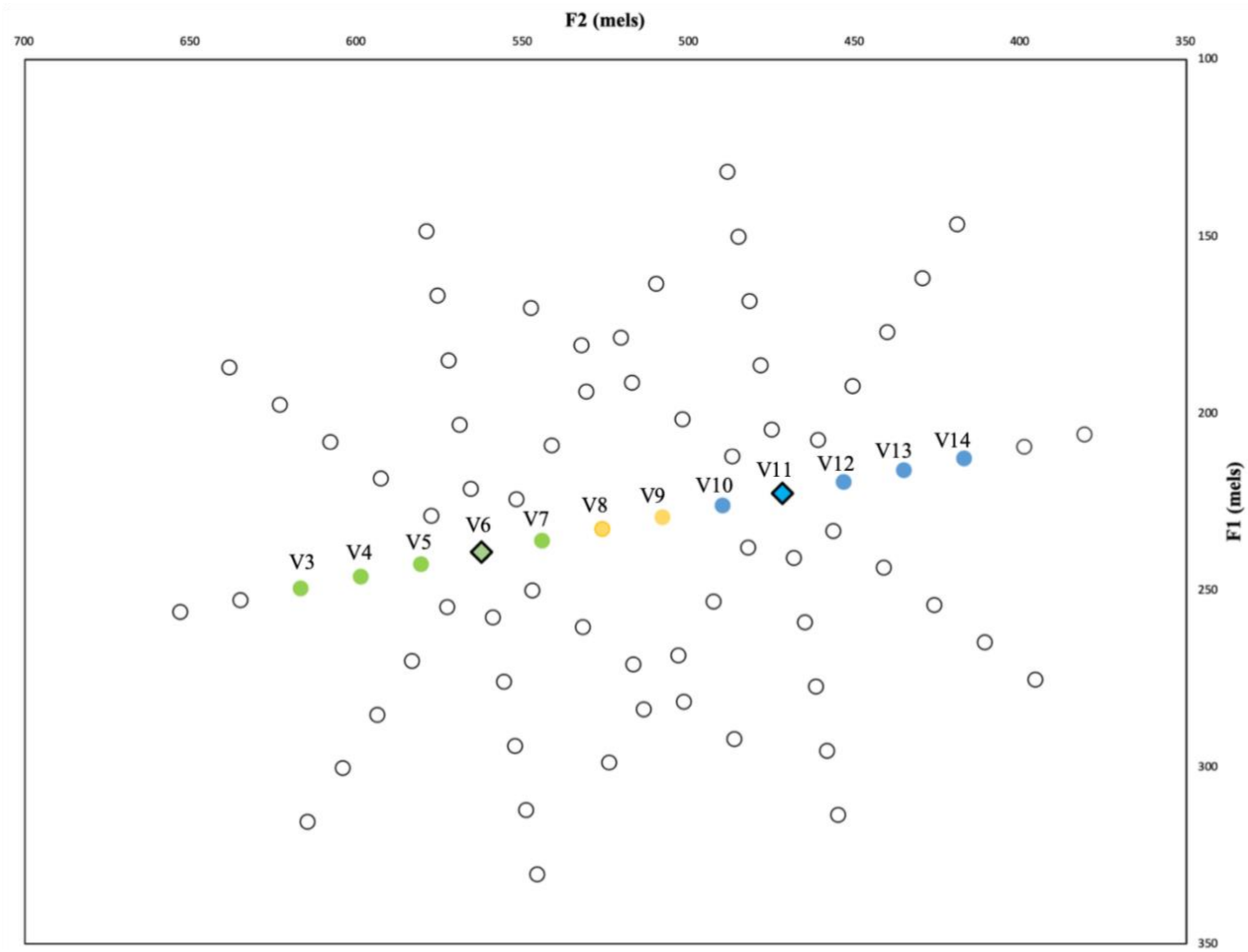

Figure 3. Visual representation of the Experimental Regions: Region A includes pairs involving stimuli V3, V4 ,V5 and V6 (English /u/ prototype); Region B includes pairs involving V6 (English /u/ prototype), V7, V8, and V9; Region C includes pairs involving V8, V9, V10 and V11 (French /u/ prototype) and Region D includes to pairs involving V11 (French /u/ prototype), V12, V13, and V14.

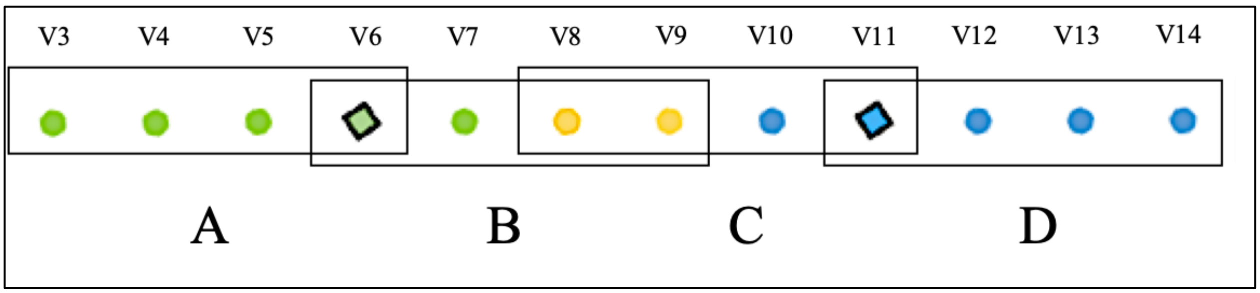

Figure 4. Overall median goodness scores for /u/ variants for monolingual Canadian English adult listeners. The upper number in each sphere indicates the median rating score for that variant (0-7) while lower number represents the number of subjects who have given their highest mean rating to that variant. The magnitude of the group median rating score is reflected in the size of the spheres. Although the graph is plotted in Hz, it is important to note that all tokens are at equal psychophysical distances from each other (Mels).

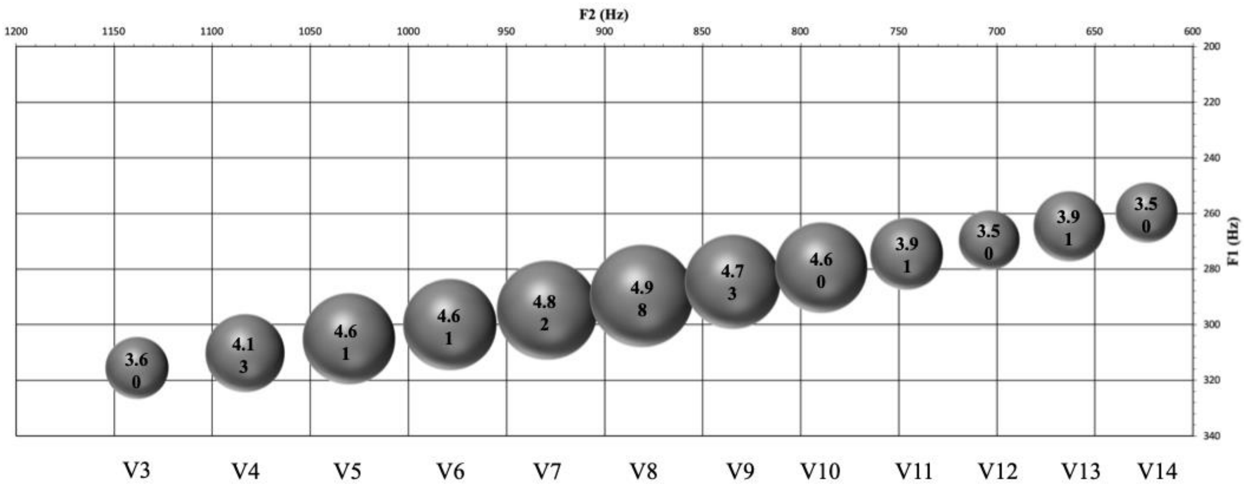

Figure 5. Median goodness scores for /u/ variants rated within “Same” trials for monolingual Canadian English adult listeners. The upper number in each sphere indicates the median rating score for that variant (0-7) while lower number represents the number of subjects who have given their highest mean rating to that variant. The magnitude of the group median rating score is reflected in the size of the spheres. Although the graph is plotted in Hz, it is important to note that all tokens are at equal psychophysical distances from each other (Mels).

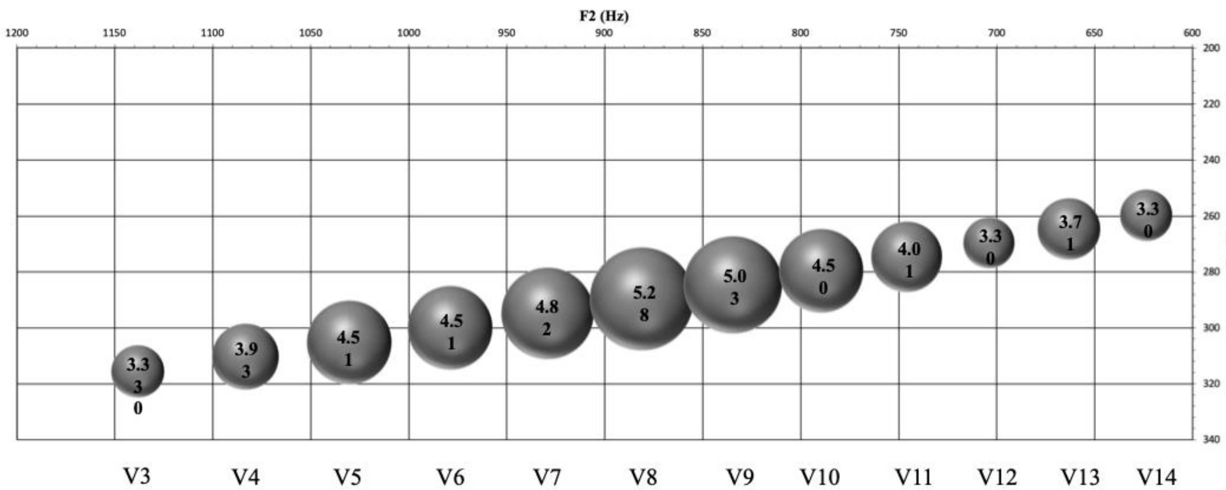

Figure 6. Median goodness scores for /u/ variants rated within "different" trials for monolingual Canadian English adult listeners. The upper number in each sphere indicates the median rating score for that variant (0-7) while lower number represents the number of subjects who have given their highest mean rating to that variant. The magnitude of the group median rating score is reflected in the size of the spheres. Although the graph is plotted in Hz, it is important to note that all tokens are at equal psychophysical distances from each other (Mels).

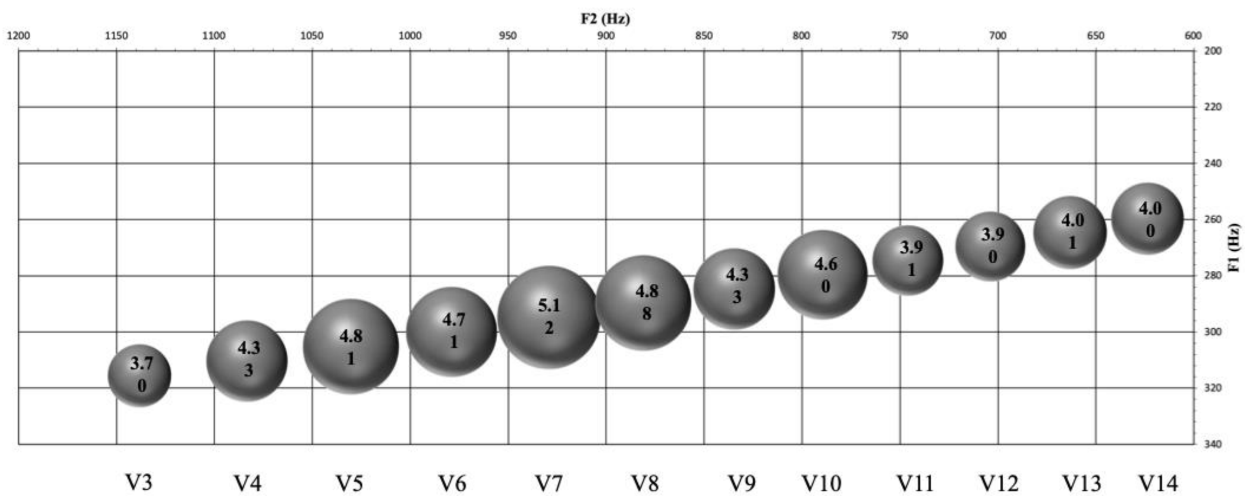

Figure 7. Accuracy of participant responses (A') by vowel region (A, B, C, D) and direction of change (Toward Focality, Against Focality) for trials involving vowel pairs located at 1 step from the original prototypes as show in 7a (top) . Native English listeners performed significantly better when discriminating from a more focal variant to a less focal variant Region C and D. *Significance levels: p = 0.00(***); p < 0.01 (**); p < 0.05 (*); p < 0.1 (.); p > 0.1 (NS) as shown in 7b (bottom)*

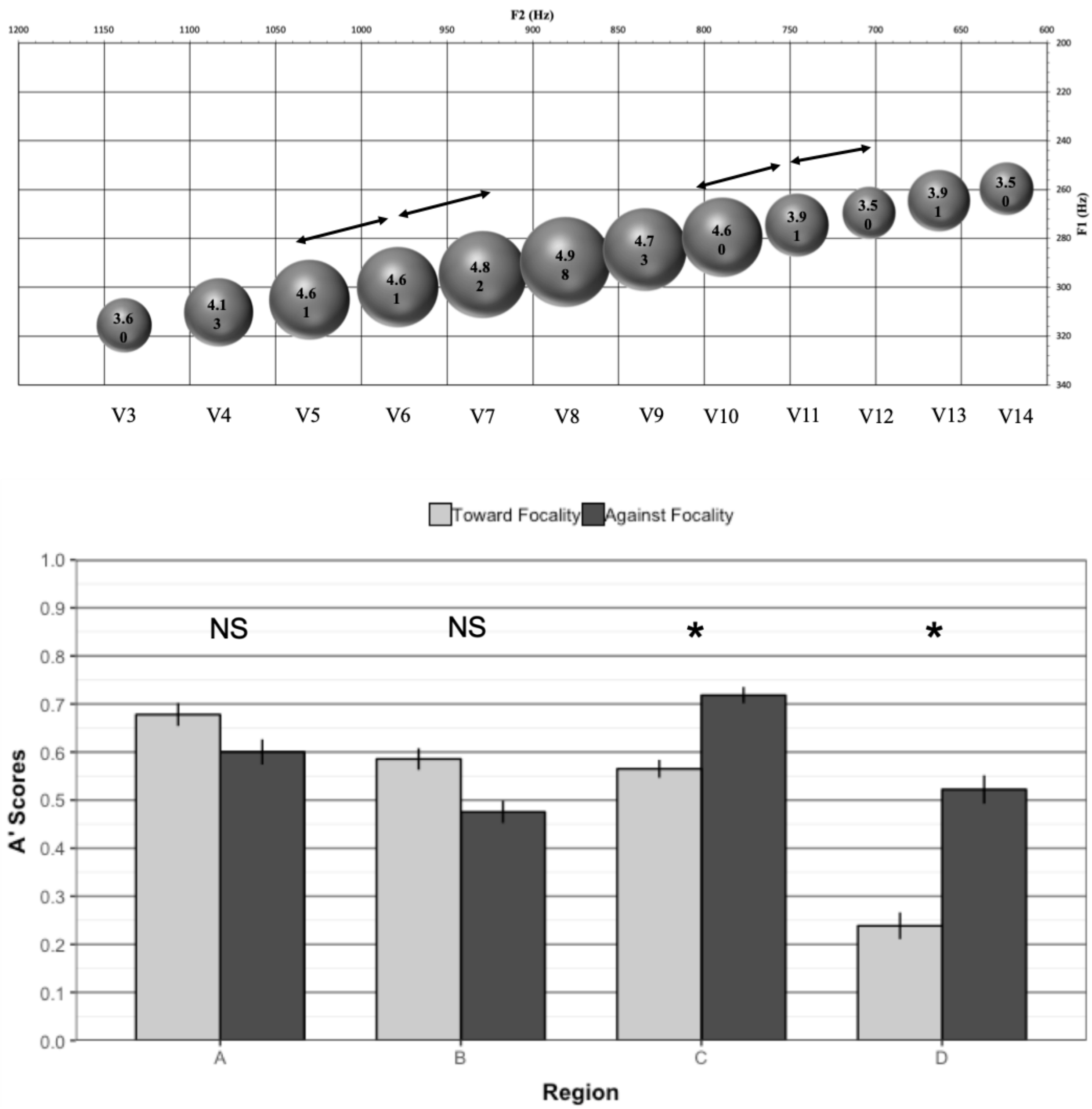

Figure 8. Accuracy of participant responses (A') by vowel region (A, B, C, D) and direction of change (Toward Focality, Against Focality) for trials involving vowel pairs located at 2 steps from the original prototypes (V6 & V11) as show in 8a (top). Native English listeners performed better when discriminating from a less focal to a more focal vowel in Regions A, B and D, while in the Region C, their discrimination rates were higher in the opposite direction. *Significance levels: p = 0.00(***); p < 0.01 (**); p < 0.05 (*); p < 0.1 (.); p > 0.1 (NS) as shown in 8b (bottom).*

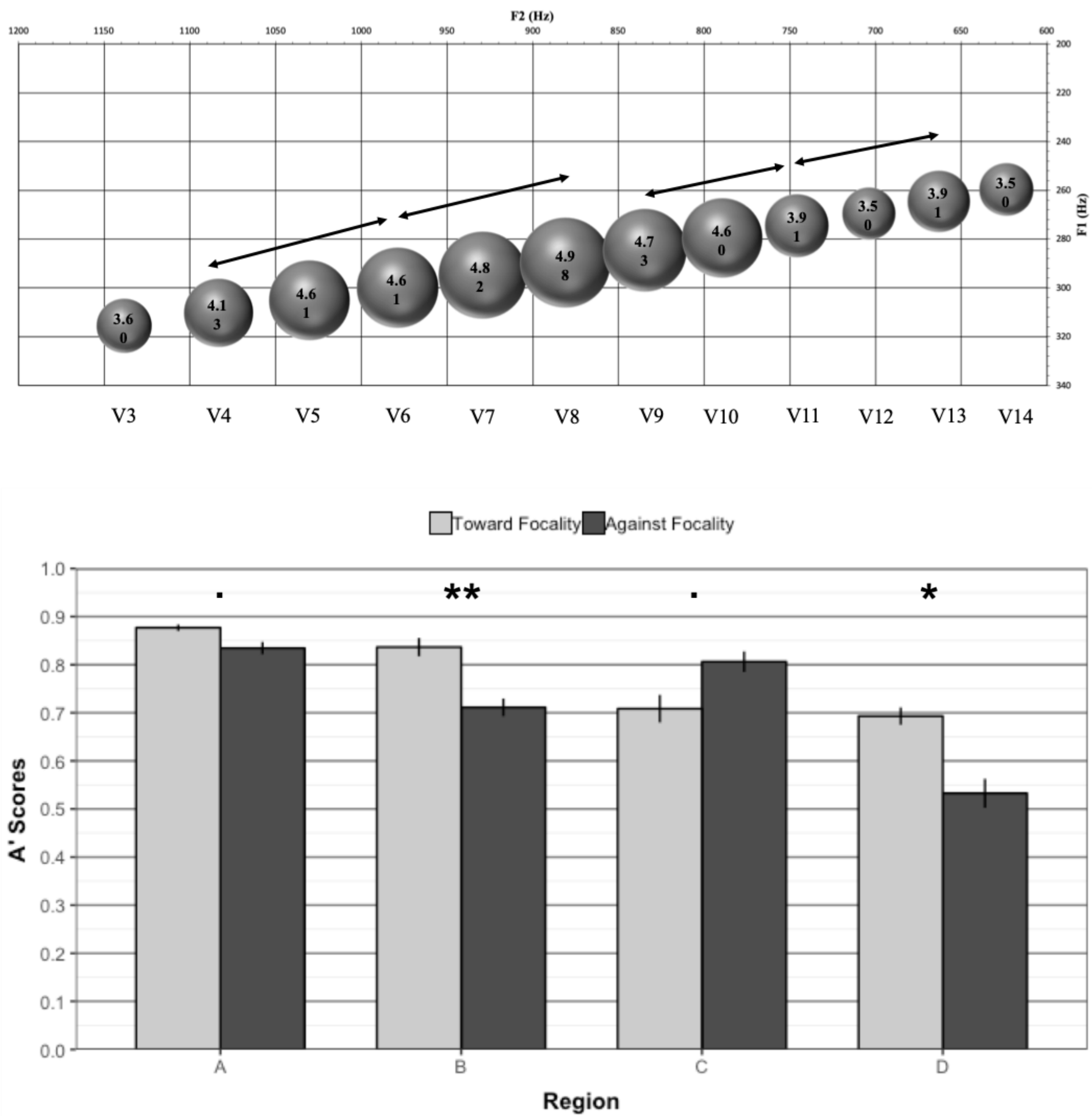

Figure 9. Accuracy of participant responses (A') by vowel region (A, B, C, D) and direction of change (Toward Focality, Against Focality) for trials involving vowel pairs located at 3 steps from the original prototypes (V6 & V11) as show in 9a (top). Native English listeners performed better when discriminating from a less focal to a more focal vowel in Regions A and D. *Significance levels: p = 0.00(***); p < 0.01 (**); p < 0.05 (*); p < 0.1 (.); p > 0.1 (NS) as shown in 9b (bottom).*

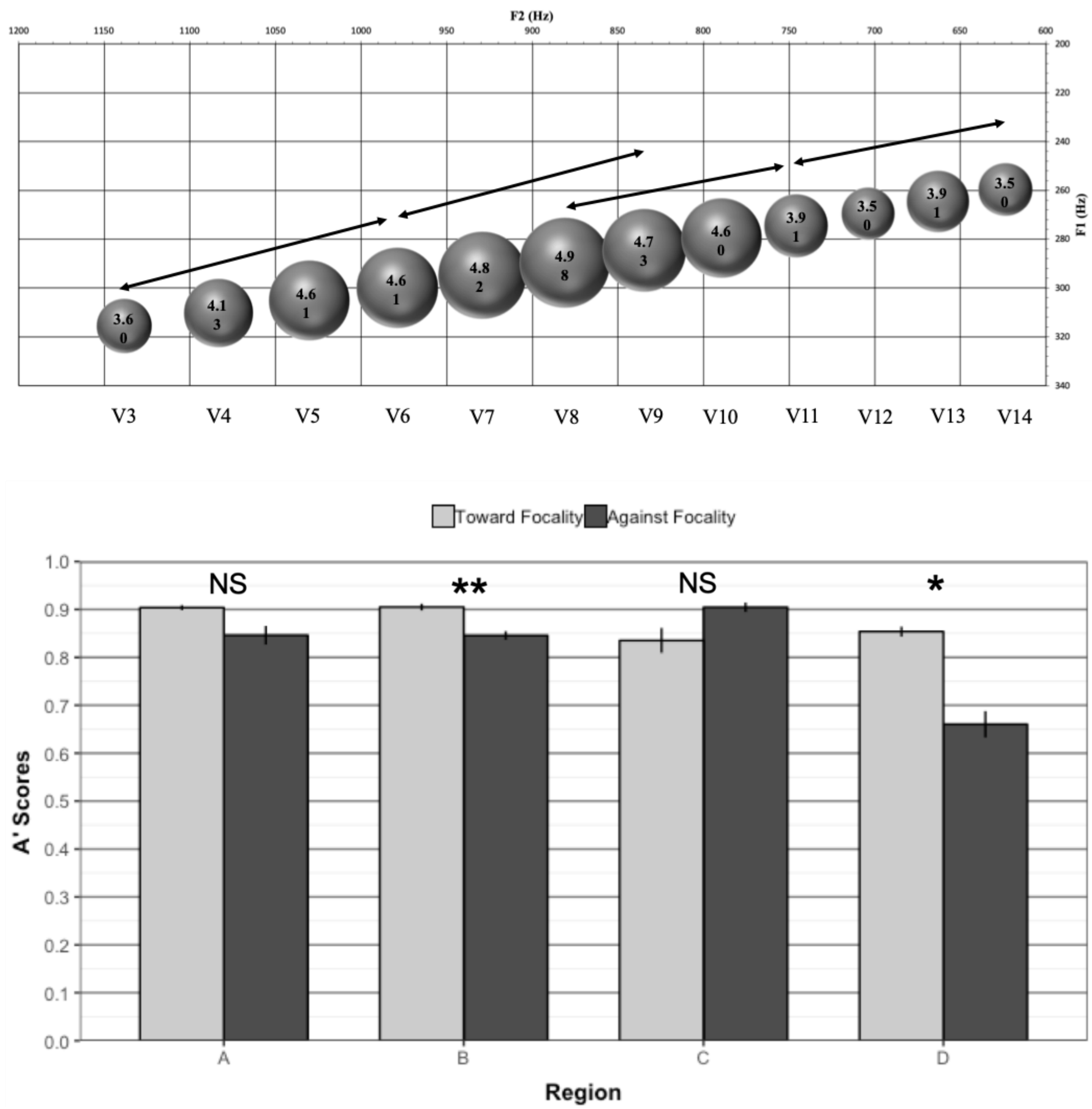

Figure 10. Accuracy of participant responses (A') by vowel region (A, B, C, D) and direction of change (Toward Focality, Against Focality) for all psychophysical steps collapsed. Native English listeners performed better when discriminating from a less focal to a more focal vowel in Regions A and B, while in the Region C, their discrimination rates were higher in the opposite direction. *Significance levels: p = 0.00(***); p < 0.01 (**); p < 0.05 (*); p < 0.1 (.); p > 0.1 (NS).*

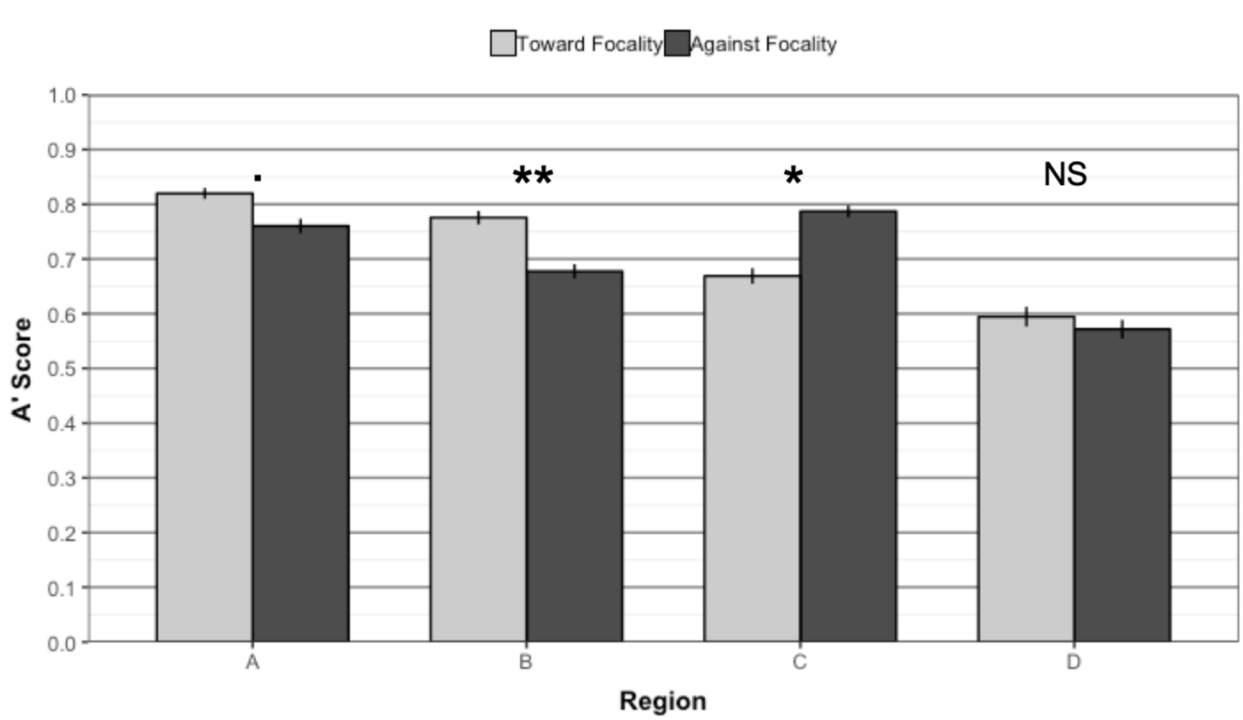

Table 1. Output table for all pairwise comparisons performed according to the number of psychophysical steps analysed, significant effects are in italic. *Significance levels: p = 0.00(***); p < 0.01 (**); p < 0.05 (*); p < 0.1 (.); p > 0.1 (NS).*

	1-step			**2-steps**		
	Direction of change	*F-Value*	*P-Value*	*Direction of change*	*F-Value*	*P-Value*
Region (A) x Direction of Change	Toward focality	0.818	0.377 NS	Toward focality	3.270	0.086 .
Region (B) x Direction of Change	Toward focality	2.374	0.140 NS	Toward focality	14.67	0.001 **
Region (C) x Direction of Change	Against focality	4.886	0.040 .	Against focality	3.443	0.079 .
Region (D) x Direction of Change	Against focality	6.821	0.017 *	Toward focality	4.999	0.038 *
	3-steps			**Collapsed across steps**		
	Direction of change	*F-Value*	*P-Value*	*Direction of change*	*F-Value*	*P-Value*
Region (A) x Direction of Change	Toward focality	1.314	0.266 NS	Toward focality	4.202	0.054 .
Region (B) x Direction of Change	Toward focality	10.42	0.004 **	Toward focality	11.34	0.003 **
Region (C) x Direction of Change	Against focality	1.789	0.197 NS	Against focality	7.428	0.013 *
Region (D) x Direction of Change	Toward focality	8.060	0.011 *	Toward focality	0.184	0.673 NS