Brain Networks for Singing and Cello Playing: Integration of Sensory-Motor Information

Melanie Segado

Montreal Neurological Institute
McGill University, Montreal
December 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Contents

List of Figures

Acknowledgements

The research presented here represents the combined efforts of many people over many years, none more so than my co-supervisors Robert Zatorre and Virginia Penhune, without whom none of the research presented here would have been possible. I am fortunate to have had the opportunity to spend this time learning from both of them and I know that I am a better scientist because of it. I would also like to gratefully acknowledge the feedback and advice given by the members of my advisory committee, Caroline Palmer and Marcelo Wanderley. This research would also not have been possible without Avrum Hollinger, who designed and built the MR-Compatible cello used throughout this thesis, and Joseph Thibodeau who lent his technical expertise to every stage of experiment setup, data collection, and data analysis. Boris Kleber provided a great deal of feedback on the fMRI experimental design and on the interpretation of the subsequent results. Alexandre Lehmann was instrumental to designing the EEG experiment, and Phillipe Albouy and Isabelle Arseneau-Bruneau provided much needed hands on training during the initial stages of EEG data collection. Isabelle also translated the abstract of this thesis into French. Yavar Korkian provided invaluable assistance with the behavioural and EEG analyses, and endless moral support (and espresso) during the writing process. My undergraduate research assistants (Luke Zawadiuk, Han Byul Moon, Vanessa Mukli, and Yilin Zhang) diligently helped recruit participants, and helped collect and clean data. Yves Methot at the Center for Interdisciplinary Research on Music Media Technology (CIRMMT) helped troubleshoot numerous iterations of the experimental setup, the Montreal Neurological Institute McConnell Brain Imaging Center staff (Ron Lopez, Louise Marcotte, and David Costa) helped craft innovative solutions to the unique challenges posed by in-scanner cello performances, and Marc Bouffard and Patrick Bermudez fixed countless software and server issues. I would also like to extend my gratitude my lab mates (past and present) for years of feedback on presentations, funding applications, and abstracts. I am lucky to have such fantastic scientists as colleagues and friends. I want to acknowledge my brother for patiently teaching me about electronics, and python, and music, and machine learning, and so many cool things that helped me get to this point. Finally, I want to thank my parents for supporting me through many (many) years of school. I know you worked extremely hard, and I am very grateful for the opportunities I’ve had as a result.

Auditory-motor integration underlies our ability to speak, to sing, and to play musical instruments. Singing and playing continuous pitch musical instruments, like the cello, are specifically contingent on a highly developed pitch regulatory system. The brain mechanisms contributing to vocal pitch regulation have been studied extensively for singing, and for vocalization more generally. However, singing and instrument playing rely on completely independent motor effectors to generate pitched sounds. Moreover, the auditory-motor association used for musical instrument playing is much more arbitrary than that used for singing given that neither the sounds nor the movements carry any significance for vocal communication. Thus, comparing pitched sound production in musical instrument playing to singing is instructive to better understand the relevant neural mechanisms.

To date, no study has directly compared the neural activity patterns associated with singing to a matched musical instrument within the same individuals. In this thesis we present three experiments that test a central hypothesis that playing a musical instrument (in this case the cello), which is a phylogentically new cultural task, makes use of the phylogenetically old singing network in order to regulate pitch. To do so we take advantage of a unique custom-built instrument: the fMRI-compatible opto-acoustical cello. In Experiment 1 we use fMRI to demonstrate that despite relying on completely discrete motor effectors and having very different evolutionary relevance, the brain areas recruited for singing and cello playing directly overlap within the same individuals in areas within the singing network. The singing network comprises auditory (HG, pSTG), motor (SMA, M1, ACC), and auditory-motor integration areas (SMG, IPS). In Experiment 2 we replicate and expand on this finding by showing that this overlap in recruited brain regions measured with fMRI is specifically tied to the auditory-motor integration that is necessary for pitch regulation by introducing a pitch feedback perturbation and asking participants to either ignore the introduced perturbation or to compensate for it. This manipulation selectively activates many of the dorsal-stream auditory-motor integration regions. Finally, in Experiment 3 we show preliminary EEG evidence that the temporal dynamics of auditory-motor integration in singing are also similar during cello playing and singing using a similar pitch perturbation protocol. Taken together these findings extend our current understanding of the auditory-motor integration system

- specifically the auditory-vocal system and the singing network - by showing that brain networks in place for vocal pitch regulation can be reused by tasks with learned, arbitrary auditory-motor associations like those required to play the cello.

Resumé

Parler, chanter, ou jouer d'un instrument de musique s'appuie sur des fonctions telles que l'intégration auditive-motrice. En chant et au violoncelle, l'ajustement de l'intonation s'effectue sur un continuum, ce qui sollicite un système régulateur de la hauteur (tonale) des sons des plus sophistiqués. L'activité cérébrale qui contribue aux ajustements de la hauteur de la voix a largement été étudiée en chant, et plus généralement, pour la production vocale. Cependant, jouer un son déterminé, sur un instrument comme le violoncelle, diffère en s'appuyant sur des effecteurs moteurs complètement indépendants comparativement au chant. De plus, l'association auditive-motrice présente lors du jeu instrumental demeure plus arbitraire, car les sons et mouvements n'ont pas les mêmes significations que ceux utilisés couramment durant la communication verbale. Ainsi, comparer la production sonore en chant à celle du violoncelle pourrait permettre de mieux comprendre les mécanismes neuronaux sous-jacents à l'intégration auditive-motrice.

À ce jour, aucune étude n'a directement comparé, chez de mêmes individus, les modèles d'activité neuronale qui sont associés au chant et au jeu instrumental. Cette thèse présente trois expériences afin de tester l'hypothèse centrale suivante : que le jeu instrumental (au violoncelle), une tâche culturelle et phylogénétiquement nouvelle, ait recours à un réseau phylogénétiquement déjà établi par la production vocale pour ajuster la hauteur des sons. Pour ce faire, nous profitons d'un instrument unique, fabriqué spécifiquement sur mesure : un violoncelle opto-acoustique compatible avec l'imagerie par résonnance magnétique fonctionnelle (IRMf).

La première expérience emploie l'IRMf pour démontrer qu'au sein de mêmes individus, les régions du cerveau utilisées pour chanter et jouer du violoncelle se chevauchent directement dans le réseau neuronal gouvernant le chant, malgré le recours à des effecteurs moteurs complètement indépendants et la pertinence évolutive très différente du jeu instrumental. Le réseau associé au chant comprend les systèmes auditif (HG, pSTG), moteur (SMA, M1, ACC), et les régions d'intégration auditive-motrice (SMG, IPS).

La deuxième expérience reproduit et approfondit ces résultats en montrant, avec l'IRMf et l'introduction d'une perturbation de l'intonation des sons générés que les participants doivent ignorer ou tâcher de compenser, que les régions de chevauchement observées sont spécifiquement

reliées à la régulation de la hauteur des sons. En effet, cette manipulation active de façon sélective plusieurs régions du circuit/boucle dorsal de l'intégration auditive-motrice.

Enfin, la troisième expérience offre de premières données probantes émanant de l'électro-encéphalographie (EEG), à l'aide d'un protocole de perturbation de la hauteur des sons similaire à la deuxième expérience. Ces résultats préliminaires, constitués de mesures des complexes d'ERP N1-P1-P2, indiquent que la dynamique temporelle de l'intégration auditive-motrice du chant serait semblable à celle du jeu au violoncelle. Mis ensemble, ces résultats permettent d'étendre la compréhension actuelle du système d'intégration auditif-moteur — particulièrement en ce qui a trait au système auditif-vocal et au réseau neuronal gouvernant le chant — en montrant que les réseaux cérébraux en place pour la régulation de la hauteur des sons peuvent être réutilisés lors de tâches requérant des associations auditives-motrices apprises et arbitraires, comme le jeu au violoncelle.

Chapter 1

This general introduction was written by Melanie Segado with the supervision of Dr. Robert J. Zatorre and Dr. Virginia B. Penhune.

Chapter 2

This manuscript was written by Melanie Segado with the supervision of Dr. Robert J. Zatorre and Dr. Virginia B. Penhune. The section about the design and construction of the MR-Compatible cello was written by Dr. Avrum Hollinger, and the section about the pitch extraction and analysis GUI was written by Joseph Thibodeau, both of whom were instrumental in successfully collecting the data.

Chapter 3

This manuscript was written by Melanie Segado with the supervision of Dr. Robert J. Zatorre and Dr. Virginia B. Penhune.

Chapter 4

This manuscript was written by Melanie Segado with the supervision of Dr. Robert J. Zatorre and Dr. Virginia B. Penhune. Dr. Alexandre Lehmann helped design the experiment, but has not yet been shown the results.

Chapter 5

The general discussion was written by Melanie Segado with the supervision of Dr. Robert J. Zatorre and Dr. Virginia B. Penhune.

Chapter 1 - Auditory-Motor Integration in Singing and Cello Playing

1.1 Overview

Auditory-motor integration is integral to everyday tasks like speaking, singing, and playing musical instruments. While the specific task demands can vary widely – requiring the orofacial articulators in the case of speaking and singing, and bimanual coordination in the case of string instrument performance – the underlying set of requirements is very similar. Movements produce a set of ordered, pitched sounds. If there is a discrepancy between the produced sounds and the intended sounds, then the movements need to be adjusted accordingly. Several models of auditory-motor integration have been put forward to explain how each stage of this process is executed and where in the brain the necessary processing is happening. The earliest and best characterized models are based on vocalizations, speaking, and singing. These models, notably the DIVA model of speech (Tourville and Guenther 2011) and the Singing Network (Kleber and Zarate 2014), are based on a control system account of auditory-motor integration, comprising feedforward and feedback loops thought to regulate various features of vocal production, including vocal pitch. In studies of vocalization and singing, some key brain structures have emerged as critical for the auditory, motor, and auditory-motor integration requirements of pitch regulation in both contexts. However, the network for vocalization may have a unique representation in the brain because it is phylogenetically old, and because of the primacy of both speech and non-speech vocalizations. Or, it may be the case that this system contributes to auditory-motor integration in other contexts. Thus, in order to extend our understanding of the auditory-motor integration system it is necessary to look beyond vocalization.

Previous work done on musical instrument performance has helped to characterize the neural correlates of auditory-motor integration outside a vocal context, and has shown that a number of the same regions engaged in vocalization are engaged in playing. However, none of this research has been done on instruments that have a continuous pitch mapping. This is significant because it means that no work to date has directly addressed the neural correlates of pitch regulation in an instrument that is more comparable to singing. In the work we are presenting in this thesis we aim to address three specific questions. First, whether parts of the brain network recruited for singing directly overlap with those recruited for cello playing. Second, whether similar regions contribute to pitch regulation in both instruments, and third, how these regions work together during an auditory-motor integration task.

1.2 Background

1.2.1 Voluntary Pitch Control

Speaking and singing are similar in many regards. They both require precise control over the vocal articulators and continuous monitoring of auditory feedback. Several studies have directly compared speaking and singing and found that they share several neural substrates (Ozdemir, Norton, and Schlaug 2006; Hickok et al. 2003; Jeffries, Fritz, and Braun 2003). A study that directly compared sung to spoken vowels (Ozdemir, Norton, and Schlaug 2006) found overlapping brain activity for singing and speaking in inferior pre- and post-central gyrus, superior temporal gyrus (STG), and superior temporal sulcus (STS) but also found that singing recruited right STG, primary sensory (S1) and motor cortex (M1), inferior central operculum, and inferior frontal gyrus (IFG) to a greater extent than speaking. The authors interpreted this to mean that there is a general network for vocal production, with a right lateralization for singing(Ozdemir, Norton, and Schlaug 2006). Because of these global similarities between speech and singing, it is helpful to look at speech-based models of auditory-vocal integration to interpret how the system may be working together to accomplish the task.

One such model is the DIVA model put forward by Tourville and Guenther (Tourville and Guenther 2011). This model comprises a feedforward component and an error driven feedback loop, and depends on the existence of an ‘internal model’ of speech to which auditory and somatosensory feedback can be compared. When the speaker produces an utterance using the speech articulators, this output is perceived by the auditory system and the somatosensory system. The input to these systems is compared to an “auditory error map” and a “somatosensory error map," which in turn send signals forward to the "feedback control map" that modify the position and trajectory of the articulators for the next utterance. This model also puts forth brain regions that may be contributing to each stage of the process. Noteworthy among them are the posterior STG (pSTG) which is hypothesized to be the location of the auditory target and error maps, and the supramarginal (SMG), which is hypothesized to be the location of the somatosensory target and error maps.

This model is itself based on work done largely in non-human primates that characterized the brain structures involved in vocalizations (Uwe Jürgens 2009; Uwe Jürgens and Pratt 1979a, 1979b). As was summarized in Jürgens (2009), converging evidence from anatomical studies, single unit recording studies, and microstimulation studies identified a few key cortical (M1, anterior cingulate cortex [ACC]) and subcortical (periaqueductal grey [PAG]) regions that were involved in vocalization. The PAG was found to directly innervate phonotory motor neurons (Dujardin and Jürgens 2005), which is likely why stimulation of the PAG resulted in vocalization (Uwe Jürgens and Ploog 1970; Larson, Charles R. 1985), and lesions of the PAG resulted in mutism (Uwe Jürgens and Pratt 1979b). As in non-human primates, lesions of human PAG also result in mutism (Esposito et al. 1999), and tasks like speaking and singing that require vocalization consistently recruit the PAG (Zarate and Zatorre 2005; Perry et al. 1999; Kleber and Zarate 2014). Stimulation of the ACC also elicits vocalization despite not directly innervating laryngeal musculature (Eliades and Wang 2012), and likely governs the voluntary initiation of vocalization. As was noted in Jurgen's 2002 review of mamalian vocalization (Uwe Jürgens 2002), lesions of the ACC cause a decrease in voluntary initiation of vocalizations in both non-human (Sutton, Trachy, and Lindeman 1981; Trachy, Sutton, and Lindeman 1981) and human primates (Rubens 1975; Uwe Jürgens and von Cramon 1982). Research has also shown that the ACC is involved in auditory-motor integration. Stimulating the ACC of marmosets has been shown to decrease activity in auditory cortex prior to vocalization (Eliades and Wang 2003) and the extent to which auditory cortex activity was attenuated by the ACC varied depending on the volume of auditory feeback (Eliades and Wang 2012). In bats, the ACC has been shown to mediate the vocal pitch-shift reflex that is necessary for them to accurately echolocate during flight despite receiving auditory feedback that is subject to the doppler shift (Metzner 1996; Gooler and O'Neill 1987).

The ACC is a core component of the singing network (Kleber and Zarate 2014) and is hypothesized to mediate the initiation of vocal corrections (Zarate and Zatorre 2008; Zarate, Wood, and Zatorre 2010).For singing, to a much greater extent than for speaking, it is essential to voluntarily regulate pitch. As such, much of the research looking specific at pitch regulatory control has used singing, not speech. By studying brain activity during singing (Kleber et al. 2007; Perry et al. 1999; Zarate and Zatorre 2005), imagined singing (Kleber et al. 2007), and singing with altered auditory (Zarate and Zatorre 2008; Zarate, Wood, and Zatorre 2010; Kleber et al. 2017) and sensory (Kleber et al. 2013) feedback researchers have characterized the singing network, which is a subset of brain regions within the auditory-vocal network that are specifically involved in singing (Kleber and Zarate 2014). These include regions that govern feed forward control over the vocal articulators (M1, ACC, PAG), regions that process auditory feedback (STG), regions that compare feedback to the expected output (intraparietal sulcus [IPS], anterior insula [aINS], and update the forward model accordingly (supplementary motor area [SMA], dorsal pre-motor cortex [dPMC]). Many of these same regions are also recruited for musical instrument performance (R. M. Brown, Zatorre, and Penhune 2015; Zatorre, Chen, and Penhune 2007).

A model of auditory-motor integration incorporating findings from both the singing and musical instrument literature (Zatorre, Chen, and Penhune 2007) describes a network of feedforward and feedback interactions that occur between regions that are hypothesized to govern motor planning and execution, auditory and sensory feedback, and auditory-motor and sensorimotor transformations. Studies of musical instrument playing have largely relied on piano and keyboard playing (Haslinger et al. 2004; Baumann et al. 2005; Parsons et al. 2005; Bangert et al. 2006; R. M. Brown, Zatorre, and Penhune 2015), but also include simulated violin performance (Lotze et al. 2003, 1999), simulated guitar playing (Buccino et al. 2004; Vogt, Higuchi, and Roberts 2007; Higuchi et al. 2012), and playing a trumpet (Gebel et al. 2013). These studies have highlighted the brain structures associated with sound-to-action mapping, and shown that musicians jointly activate auditory and motor structures even if they are just listening to sounds (J. L. Chen, Penhune, and Zatorre 2008; Lahav et al. 2005), just moving their hands (Lotze et al. 2003, 1999), or even imaging performance in the absence of any sound or movement (Lotze et al. 2003, 1999). Moreover, studies on musical instrument learning have shown that the association between sound and action can develop with very little training (Wollman et al. 2018; J. L. Chen, Rae, and

Watkins 2012; Lega et al. 2016), and that this association is likely encoded in the dPMC (J. L. Chen, Rae, and Watkins 2012). This finding was recently replicated using the cello instead of the keyboard (Wollman et al. 2018).

1.2.2 Compensatory reflexes

The pitch-shift reflex was first characterized in speech. Research has long showed that perturbing pitch feedback during phonation results in a compensatory response (Elman 1981). Elman showed that increasing or decreasing the pitch that participants were hearing themselves produce would cause them to change their produced pitch to try to ‘correct’ for the perturbation. This paradigm was refined by Houde and Jordan (1998) and has since been used extensively to study the auditory-motor integration system as it provides direct evidence that auditory feedback is being used to influence movements.

The finding that participants compensate for incorrect pitch feedback is stable across the literature for phonation (Burnett et al. 1998; Liu and Larson 2007; Toyomura et al. 2007) and for singing (Zarate, Wood, and Zatorre 2010; Zarate and Zatorre 2008). These studies have shown that the IPS and inferior parietal lobule ([IPL], specifically the SMG) along with STG, M1, SMA, and the ACC are recruited when an auditory-to-motor transformation is required, as is the case when attempting to stabilize pitch (Toyomura et al. 2007; Kleber and Zarate 2014). The pitch shift reflex has also been studied using EEG (Liu and Larson 2007). Research has shown that the N100 and P200 components are closely matched to the behavioural compensatory response and, as such, can be interpreted as relating to the detection of incorrect auditory feedback and selection of a compensatory motor response (Liu and Larson 2007). In a follow up study using time-frequency analyses, researchers found that phase synchronized fronto-central theta (5-8Hz) activation was enhanced in musicians relative to non-musicians, especially if musicians had absolute pitch (C. Luo et al. 2012). Frontal activation within the Theta band is thought to derive from activity in the ACC and the SMA (Behroozmand et al. 2015; Debener et al. 2005), both of which are commonly recruited in fMRI studies of auditory-motor integration. The authors interpreted this finding to mean that the evoked theta activation reflected the degree to which auditory feedback was successfully used to compensate. They also found non-phase synchronized delta (1-4Hz) activation was more prominent for non-musicians than for musicians, and that the magnitude of this activation was positively correlated with the magnitude of the compensatory response. As such, they proposed that the delta activation was likely related to the selection of the compensatory movement.

Another prominent example of an auditory-vocal reflex is the Lombard effect, which is when vocal intensity increases in amplitude when feedback is attenuated (J. Luo, Hage, and Moss 2018). This effect has been observed in humans during both speaking (J. Luo, Hage, and Moss 2018) and singing (Tonkinson 1994) and is remarkably well preserved across species (Smotherman 2007). This effect is seen in cats (Watanabe and Katsuki 1974), birds , bats, even some fish (J. Luo, Hage, and Moss 2018). Research on the Lombard effect has shown that the compensatory intensity change is mediated by a network of feedback (Auditory Cortex) and feed forward (M1, ACC, PAG) regions (J. Luo, Hage, and Moss 2018). Many of these regions have also been shown to mediate compensatory pitch changes in speaking (Tourville and Guenther 2011) and singing (Kleber and Zarate 2014). Another relevant similarity between these two reflexes is that the Lombard effect, like the pitch-shift reflex, is experience dependent. Research has shown that choral singers can learn to suppress the Lombard effect (Tonkinson 1994) and that expert singers, in contrast to non-singers, can suppress the pitch shift reflex (Zarate, Wood, and Zatorre 2010). Together these findings suggest that the pitch-shift reflex observed in singing is not, in fact, a product of expert musical training but rather that it is a true involuntary reflex that is phylogenetically old and part of a larger auditory-vocal system in place for maintaining vocal stability.

1.2.3 Poor Pitch Singing and Amusia

Insight on pitch regulatory control has also come from work done on people with poor pitch singing skills (Pfordresher and Mantell 2014; Pfordresher and Brown 2017, 2007; Berkowska and Dalla Bella 2009), and people with amusia (Peretz and Hyde 2003; Dalla Bella, Giguère, and Peretz 2009; S. Hutchins et al. 2010; Tremblay-Champoux et al. 2010). Characterizations of the singing accuracy range in the general population have found that most people can stay within 50 cents correct when they are singing a melody (Tremblay-Champoux et al. 2010; Pfordresher and Brown 2017). This is consistent with studies showing that untrained people can match target tones (Zarate and Zatorre 2005, 2008; Zarate, Wood, and Zatorre 2010; Kleber et al. 2017), and even compensate for large (200 cent) pitch feedback perturbations about as well as expert singers (Zarate and Zatorre 2008; Zarate, Wood, and Zatorre 2010). However, about 10-20% of the population are ‘poor pitch singers’ that tend to sing more than 100 cents from the correct pitch despite intact pitch perception (Welch 1979; Pfordresher and Brown 2007; Pfordresher and Mantell 2014).

Research on poor pitch singing has shown that participants can sing along with recordings of themselves more accurately than they can sing along with recordings of other people, and certainly better than they can in the absence of a recording. The authors took this finding as evidence of a deficit in the ‘inverse model’ of vocal pitch regulation. In other words, that poor pitch singing results from an inability to take auditory feedback and match it with the appropriate corrective action (Pfordresher and Mantell 2014). Other research has suggested that poor pitch singing may result from deficits in motor control as opposed to errors with pitch perception (S. M. Hutchins and Peretz 2012) which is corroborated by the finding that people with amusia, who lack pitch perception, can still sing provided that it is done in the context of vocal imitation (S. Hutchins et al. 2010; S. Hutchins and Peretz 2012). Taken together, this research has shown that there may be multiple pathways for singing: one for vocal imitation that does not rely on auditory-to-motor transformations but rather relies exclusively on previously experienced sensorimotor associations, and another pathway that relies on monitoring auditory feedback and using it to guide corrective movements. The singing difficulties seen in congenital amusia likely result from a deficit in the latter given that vocal imitation is still largely intact. However, multiple studies have shown that this deficit does not stem from issues with primary auditory processing, and some have suggested that amusia more likely results from issues related to auditory-motor integration (Mandell, Schulze, and Schlaug 2007).

Structural (Hyde et al. 2007; Mandell, Schulze, and Schlaug 2007; Loui, Alsop, and Schlaug 2009) and functional (Hyde, Zatorre, and Peretz 2011) brain imaging studies comparing people with amusia to non-amusic controls have shown abnormal grey and white matter volumes in STG and IFG(Hyde et al. 2007; Mandell, Schulze, and Schlaug 2007), decreased anatomical connectivity between these two regions (Loui, Alsop, and Schlaug 2009), as well as decreased recruitment of IFG during a melody listening tasks with a concomitant decrease in functional connectivity between STG and IFG (Hyde, Zatorre, and Peretz 2011). The IFG is one of the structures identified in the DIVA model as a hypothesized location for the feedback control map that bridges the feedback and feedforward control systems. Further evidence for this comes from work done on ignoring incorrect pitch feedback, which has been shown to rely not only on auditory regions (pSTG), but also on the IFG (Zarate, Wood, and Zatorre 2010). Evidence supporting this hypothesis also comes from work done in EEG. Electrophysiological studies done on pitch processing in amusia have also shown that the N100 ERP, which is related to primary auditory processing, is comparable for amusic and non-amusic participants but that amusics lack the N200 and P300 components typically seen for auditory mismatch detection (Peretz, Brattico, and Tervaniemi 2005).

1.2.4 Singing Expertise

It is certainly true that compensating for incorrect pitch feedback is essential for executing auditory-motor sequences. However, auditory feedback is sometimes unavailable or not a reliable indicator of the movements made. This is common in choral or orchestral settings where musicians are frequently expected to maintain a stable pitch despite hearing the produced pitches of those around them much louder than their own voice or instrument. In these cases, it is necessary to ignore the incorrect pitch feedback and instead rely on somatosensory feedback and a strong forward model. Expert singers have been shown to sing on pitch even in the absence of auditory (Kleber et al. 2017) and sensory (Kleber et al. 2013) feedback, and when receiving incorrect auditory feedback (Zarate and Zatorre 2005, 2008; Zarate, Wood, and Zatorre 2010). This experience-dependent ability to rely exclusively on a feed-forward representation has been linked primarily to the right aINS (Kleber and Zarate 2014). Research comparing expert singers to non-singers under masked auditory feedback conditions found that expert singers, whose accuracy was not affected by the manipulation, recruited right aINS to a greater extent when singing without auditory feedback compared to singing with auditory feedback. Conversely non-singers, whose accuracy decreased without auditory feedback, recruited the aINS to an even lesser extent when singing without feedback than with it. Functional connectivity between aINS and IPL, IFG, and vocal areas of sensory cortex during masked auditory feedback conditions showed a similar pattern with increased connectivity for singers during masking and decreased connectivity for non-singers.

The opposite pattern of aINS recruitment was observed when sensory feedback was attenuated by anesthetizing the vocal cords of singers and non-singers despite the performance of singers being less affected by the manipulation (Kleber et al. 2013). In this case, singers activated aINS to a lesser extent when singing with anesthesia than without it while non-singers activated aINS more when singing with anesthesia. Taken together these results show that the aINS is a key substrate for switching from relying on auditory feedback to relying on somatosensory feedback in an experience-dependent manner (Kleber and Zarate 2014; Kleber et al. 2017). Insights on feedforward control mechanisms also come from work done with imagined performance which, by definition, lacks any feedback component. When opera singers were asked to imagine singing an aria they recruited brain regions related not only to preparatory motor control (SMA, PMC, ACC), but also some related to sensory-motor integration (IPL, aINS) and even some related to sensory feedback (S1) (Kleber et al. 2017). The finding that sensory and sensorimotor integration regions are recruited during imagined performance is further evidence that the vocal network as a whole is tuned not only for vocal production but also for auditory and sensory monitoring and adjustments.

1.3 The present investigation

In this thesis we present three brain imaging experiments designed to test the central hypothesis that cello playing and singing recruit largely overlapping brain networks that work together in a similar way, but that there are also modality-specific differences. For the first two fMRI experiments, we compared singing and cello playing in both normal (Exp. 1) and pitch-shifted conditions (Exp. 2). We expected that singing would recruit motor regions consistent with the necessary orofacial articulators, including larynx area of motor cortex, and that cello playing would recruit areas of motor cortex corresponding to the hands and arms. Further, based on the literature reviewed above, we expected to see similar engagement of auditory-motor integration regions (specifically dPMC and IPS) when participants are required to adjust their performance in the face of altered auditory feedback. In the third experiment (Exp. 3), to assess the brain mechanisms underlying this compensatory pitch shift at a finer temporal grain, we conducted an EEG study using the same paradigm as in Experiment 2. In this experiment, we hypothesized that cellists would show similar P1-N1-P2 responses to perturbed pitch feedback when playing the cello and singing. In addition, we sought to test whether there was a reflexive pitch shift response for cello-playing as is seen in singing. We hypothesized that, if cello playing was making use of the same brain networks as singing for the same task, then it should follow that expert cellists would show the same compensatory pitch shift reflex that is observed in singing.

While this research focuses very specifically on singing and cello playing, these findings help address a more fundamental question about whether phylogenetically newer tasks like playing a musical instrument scaffold off of existing cortical infrastructure in place for vocalization to accomplish the same task. To answer our central research question required us to consider three main things: (1) selecting an appropriate musical instrument to match the continuous pitch control present in the vocal system, but (2) that did not engage vocal effectors, and (3) being able to manipulate auditory feedback in real time.

To fully study pitch control mechanisms, it is highly advantageous to make use of an instrument that has a continuous mapping of pitch to action. A wealth of work on auditory-motor integration has been done using the keyboard (R. M. Brown, Zatorre, and Penhune 2015; Parsons et al. 2005) and the guitar (Buccino et al. 2004; Vogt, Higuchi, and Roberts 2007; Vogt et al. 2004) some of which specifically relates to the use of pitch feedback to inform motor actions (R. M. Brown et al. 2012; Pfordresher et al. 2014). However, these instruments (keyed and fretted, respectively) do not allow for pitch to be modified in an ongoing manner, which is a core feature of vocal production. Finally, given that our goal was to determine the components of this system that are unique, or not, to vocalization, it is essential that our comparison instrument be fully independent of the vocal articulators.

The cello fits all of these criteria. Moreover, the positioning of the arms during cello playing makes it possible to adapt to the scanner environment unlike the violin/viola which require the instrument to be held between the chin and the shoulder. Finally, an enormous amount of information has been gleaned about the feedforward-feedback systems in vocal-motor control using paradigms where auditory feedback is perturbed or altered. The cello is also ideal for developing similar experiments where adjustments in auditory-motor integration are required in real time.

1.3.1 The MR-Compatible cello interface

The MR-Compatible cello interface that made the work in this thesis possible was designed and constructed by Dr. Avrum Hollinger at the Input Devices and Music Interaction Laboratory (IDMIL) of the Faculty of Music of McGill University (Hollinger and Wanderley 2013, 2015). This instrument presents a singular advantage because it allowed us to carry out real time performance tasks inside the scanner that could previously only be done with singing. However, its construction is also of general interest. The body and neck of the cello are made from a non-conductive fibreglass composite that is light, but still strong enough to sustain the tension of the strings. It has gut strings, which contain no metal and are ecologically valid since they are used by cellists to play baroque music. The fingerboard and bridge were 3D printed to allow for easy embedding of fibre-optic sensor/receiver pairs to capture left hand finger position and the vibration of the strings. As fingers are placed on the cello neck, light from the adjacent emitters is reflected into the corresponding sensors, which allows for analog finger position tracking. This also allows for capture of stylistic features like vibrato. The bridge sensors also allow for analog capture of string vibration, which was then re-converted to sound in real time. A miniature (30cm) metal-free cello bow was also specially designed to fit within the confines of the MR environment. A typical cello bow is approximately 72cm in length, which would not fit inside the 60cm diameter scanner bore. The bow also had optional sensor integration for capturing bowing gestures and bow pressure at the tip and frog. Bow pressure sensors measured the deflection of light as it passed through coiled optic fibres embedded in silicone rings (3.5cm diameter) held between the wood and hair of the bow. The 3-axis gesture capture sensor, which was mounted to the tip of the bow, measured the deflection of light as it passed through crystals whose properties are sensitive to the magnetic field inside the scanner environment.

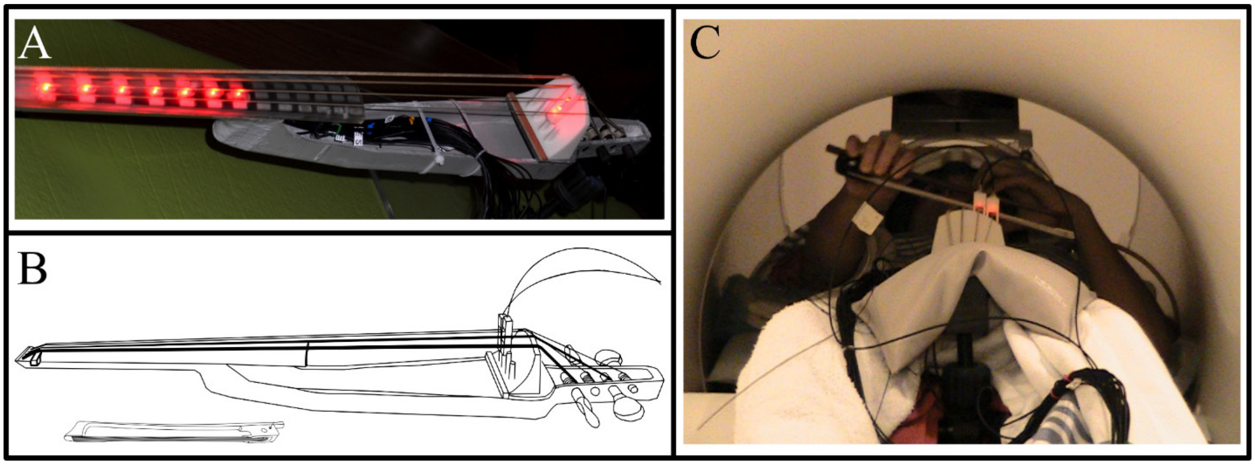

Figure 1-1 MR-Compatible Cello Interface | A) Optical sensors embedded in the fingerboard and bridge to capture finger position and string vibration, respectively. The optical sensor on the bridge of the cello provides real time, analog sound feedback. B) Line drawing of the MR-Compatible cello. The MR cello lacks a resonant body, but the length of the fingerboard and strings, and, by extension, the locations of where fingers would be placed to produce specific notes are the same as those found on a standard full-size cello. We also constructed a miniaturized bow to fully leverage the continuous pitch nature of the cello

1.3.2 Experiment 1: Shared brain network for singing and cello playing

The first experiment was designed to test three hypotheses. The first was that cellists would be able to play individual target tones inside the scanner on the MR compatible cello interface and also be able to sing them accurately. The second was that playing and singing would recruit a partially over-lapping network of regions, including in auditory (STG), motor (M1, SMA, ACC), and auditory-motor integration regions (IPS, aINS) that have been identified in previous studies of vocal production. Finally, we hypothesized that singing and cello playing would also recruit effector-specific regions of primary motor cortex, i.e., the larynx and hands. To test these hypotheses, we had expert cellists listen to and subsequently produce individual target tones both on the cello (cello playing) and with their voice (singing) while in the fMRI scanner. A sparse sampling fMRI design was used (Belin et al. 1999) so that the performance task could be done in the silent period between brain image acquisitions. The methodology of this experiment mirrors that of a previous fMRI study of singing conducted in our lab (Zarate and Zatorre 2005, 2008). To test the first hypothesis that cellists could sing and play the cello accurately, we analyzed the average difference between the target and produced pitch in each condition. We then examined regions active for singing and cello playing compared to rest and then carried out a statistical conjunction to assess the degree of overlap between the two networks. Further, we regressed individual pitch accuracy scores against BOLD signal changes to assess whether activity in these regions was related to task performance. Finally, we looked at task-based functional connectivity to determine whether or not these brain regions were working together to accomplish the task, and whether the degree of functional connectivity was correlated with good performance on the task.

1.3.3 Experiment 2: Pitch regulation in singing and cello playing

In Experiment 2 we tested a second sample of cellists singing and playing inside the fMRI scanner. To test whether regions thought to be involved in pitch feedback control were similar in the two modalities we used a pitch perturbation paradigm based on previous studies (Zarate and Zatorre 2005, 2008; Zarate, Wood, and Zatorre 2010). In this task, participants heard a series of target pitches that they were required to sing or play back. On 66% trials, a 100-Cent sudden-onset pitch perturbation was introduced (up/down). Participants were required to either compensate for the shift or to ignore it. The compensate condition was introduced to see which areas were specifically recruited for auditory-motor feedback control. The ignore condition was introduced to address two questions. The first was whether cellists would exhibit an automatic pitch compensation reflex for playing as had been observed in singing (Zarate, Wood, and Zatorre 2010). Second, we sought to determine whether cellists recruited the same network for ignoring as was reported in the literature for ignoring during singing. Ignoring incorrect feedback, unlike compensating for it, requires reliance on the forward model of the note to be maintained. Our hypothesis was that cellists would once again draw on the same brain structures to accomplish this task. However, if any condition were to show a clear difference between the two instruments we hypothesized that it would be this one since the pitch regulatory reflex is phylogenetically old whereas cello playing, and musical instrument playing more generally, is truly unique to humans.

1.3.4 Experiment 3: EEG study of perturbed pitch feedback

Much of the literature regarding the pitch regulatory reflex has been done using EEG to take advantage of its temporal precision. The rationale and methodology for the EEG experiment exactly mirrored those of the fMRI experiment. Participants were asked to listen to and subsequently sing individual target tones or play them on a Yamaha silent electric cello. On 66% trials, the pitch of the feedback was shifted suddenly up or down by 100 cents. Participants were instructed either to compensate for or to ignore the introduced perturbation. 64-channel EEG was recorded continuously throughout the experiment. Our primary motivation was to test the hypothesis that the electrophysiological correlates of processing incorrect pitch feedback, and selecting and planning a motor response (P1-N1-P2 ERP complex) would replicate previous research (Liu and Larson 2007) and would be similar in latency and magnitude for singing and cello playing. We also sought to test the hypothesis that singing would elicit a pitch-shift reflex even when participants were instructed to ignore, and to explore the hypothesis that this early component would differ between singing and cello playing.

Chapter 2 - Partially Overlapping Brain Networks for Singing and Cello Playing

2.1 Preface

Experiment 1 was designed to test the hypothesis that cello playing, a phylogenetically new cultural task, would make use of areas within the phylogenetically old singing network for the purposes of pitch regulation. This experiment was the first to make use of a specially designed MR-Compatible cello interface. This research is published in Frontiers in Auditory Cognitive Neuroscience (Segado et al. 2018)

2.2 **Abstract**

This research uses an MR-Compatible cello to compare functional brain activation during singing and cello playing within the same individuals to determine the extent to which arbitrary auditory-motor associations, like those required to play the cello, co-opt functional brain networks that evolved for singing. Musical instrument playing and singing both require highly specific associations between sounds and movements. Because these are both used to produce musical sounds, it is often assumed in the literature that their neural underpinnings are highly similar. However, singing is an evolutionarily old human trait, and the auditory-motor associations used for singing are also used for speech and non-speech vocalizations. This sets it apart from the arbitrary auditory-motor associations required to play musical instruments. The pitch distribution of the cello is uniquely similar to that of the human voice, but cello playing is completely independent of the vocal apparatus and can therefore be used to dissociate the auditory-vocal network from that of the auditory-motor network. While in the MR-Scanner, 11 expert cellists listened to and subsequently produced individual tones either by singing or cello playing. All participants were able to sing and play the target tones in tune (<50 cent deviation from target). We found that brain activity during cello playing directly overlaps with brain activity during singing in many areas within the auditory-vocal network. These areas include dorsal motor and premotor areas, SMA and Pre-SMA, STG, ACC, aINS, IPS(R), and Cerebellum but, notably, exclude the PAG and Basal Ganglia (Putamen). Second, that activity within the overlapping areas is positively correlated with, and therefore likely contributing to, both singing and playing in tune as shown by correlations with performance measures. Third, that activity in auditory areas is functionally connected with activity in dorsal motor and pre-motor areas, and that the connectivity between them is positively correlated with good performance on this task. This functional connectivity suggests that the brain areas are working together to contribute to task performance and not just coincidently active. Last, our findings showed that cello playing may directly co-opt vocal areas (including larynx area of motor cortex), especially if training begins before age 7.

Introduction

Playing musical instruments and singing both result in musical pitch patterns by integrating auditory perception with fine-motor control. Thus, an interesting question is whether the neural systems that control these two types of musical expression are similar. Auditory-motor integration for singing relies on neural systems for vocalization, where there is a relatively direct link between a motor action and the pitch produced. This evolutionarily old auditory-vocal system comprises auditory, motor and pre-motor regions in the dorsal stream, as well as the cerebellum, basal ganglia and brainstem structures (Figure 2-1) (Kleber and Zarate 2014). However, the auditory-vocal system is used for both speech and non-speech sounds as well as for singing. Furthermore, the vocal motor system follows a developmental sequence and does not require explicit training, at least for production of simple songs, which children produce by imitation, as with speech (Tourville, Reilly, and Guenther 2008). Nonetheless, musical instrument playing, which does require explicit training and often thousands of hours of practice, has been shown to rely on many of the same structures in neuroimaging studies (Zatorre, Chen, and Penhune 2007). However, no previous studies have directly compared the brain networks engaged by singing and instrument playing. This comparison would allow us to assess whether learned auditory-motor associations involved in playing an instrument build on existing brain networks that are in place for vocal production, or whether they engage different or additional systems.

Current models of auditory-motor integration for music and speech comprise a feed-forward and a feed-back component (S. Brown et al. 2004; Zatorre, Chen, and Penhune 2007; Tourville and Guenther 2011; Kleber and Zarate 2014). The feed-forward component encompasses brain areas that are responsible for motor planning and motor execution. These include primary motor, dorsal pre-motor, and supplementary motor cortices (M1, dPMC, SMA), brainstem nuclei including the periaqueductal gray (PAG), and the cerebellum. The feedback component encompasses brain areas that process sensory feedback and compare it to the expected output. Notably these include the primary and periprimary auditory cortices within the superior temporal gyrus (STG) including Heschl's gyrus, the planum temporale, and planum polare (HG, PT, PP), as well as the superior temporal sulcus (STS). These also include the anterior insula (aINS), anterior cingulate cortex (ACC), and intraparietal sulcus (IPS). These models were informed by research on speech and non-speech vocalizations, singing, and musical instrument playing. The earliest work done on the neural correlates of vocalizations was done in non-human primates. This work showed that stimulation of the PAG induces vocalizations, and that lesioning this region leads to muteness (Dujardin and Jürgens 2005; Uwe Jürgens 1976). Work in non-human primates has also shown that the SMA/pre-SMA and ACC are important for initiating voluntary vocalizations (Kirzinger and Jürgens 1982; Gooler and O'Neill 1987). This same work helped to identify a larynx specific region of primary motor cortex in non-human primates (for review see: Uwe Jürgens 1976). Several studies have since characterized a larynx-specific area of M1 in humans (S. Brown, Ngan, and Liotti 2008; Grabski et al. 2012) and replicated the finding that vocalization tasks recruit the PAG, ACC, and Pre-SMA/SMA in humans (Schulz et al. 2005) (for review see: Kleber and Zarate 2014).

Building on findings from early animal models of vocal control, research on singing has been instrumental in characterizing both behavioral features of vocal control in humans (Parlitz and Bangert 1999) and the associated brain regions. These brain regions comprise the auditory-vocal network.(Perry et al. 1999; S. Brown et al. 2004; Zarate and Zatorre 2005; Kleber and Zarate 2014). Some studies have specifically investigated the feed-forward components of the auditory-vocal network by masking sensory feedback during singing (Kleber et al. 2017), asking participants to ignore perturbed auditory feedback (Zarate and Zatorre 2008), or anesthetizing the vocal cords (Kleber et al. 2013). These studies have highlighted the role of the aINS, showing that its activity is modulated during singing when sensory feedback is masked or perturbed. Components of these studies have specifically focused on the effects of perturbed auditory feedback and found that the ACC and IPS are involved in compensating for these perturbations (Zarate and Zatorre 2008).

These studies also highlight the effects of musical expertise. Areas within the basal ganglia (putamen) were more active for experts than for non-experts (Zarate and Zatorre 2008), while the aINS was less active for experts than for non-experts (Kleber et al. 2013). However, the network of areas recruited is extremely stable throughout the singing literature and even shows some degree of overlap with the areas recruited for speaking in both auditory and motor regions (Hickok et al. 2003; Ozdemir, Norton, and Schlaug 2006). Differences between these two systems, like the relative right lateralization of auditory cortex for singing compared to speech, are thought to be related to the increased dependence on pitch processing during singing (Ozdemir, Norton, and Schlaug 2006). It therefore would seem likely that musical instrument playing, which is also highly dependent on pitch processing, would show a high degree of overlap with the brain areas recruited of singing.

Functional imaging research on musical instrument performance has also helped inform our understanding of the auditory-motor integration system. Studies on piano and keyboard playing (Haslinger et al. 2004; Baumann et al. 2005; Parsons et al. 2005; Bangert et al. 2006; R. M. Brown et al. 2012), simulated violin performance (Lotze et al. 2003, 1999), simulated guitar playing (Buccino et al. 2004; Vogt, Higuchi, and Roberts 2007; Vogt et al. 2004; Higuchi et al. 2012), and playing a trumpet (Gebel et al. 2013) show activation in many of the same core auditory and motor regions as are seen in the auditory-vocal network, but notably does not include the brainstem. However, as noted above, no direct comparison has been performed between the brain areas recruited for vocal and instrumental pitch production. To directly compare singing and instrument playing we need to use an instrument that has a continuous (as opposed to discrete) pitch mapping like the voice, but whose control is completely independent of the vocal apparatus. Instruments in the violin family fit both criteria. fMRI research on violin playing has relied on imagined performance or finger tapping as a proxy for real performance because a violin could not be played

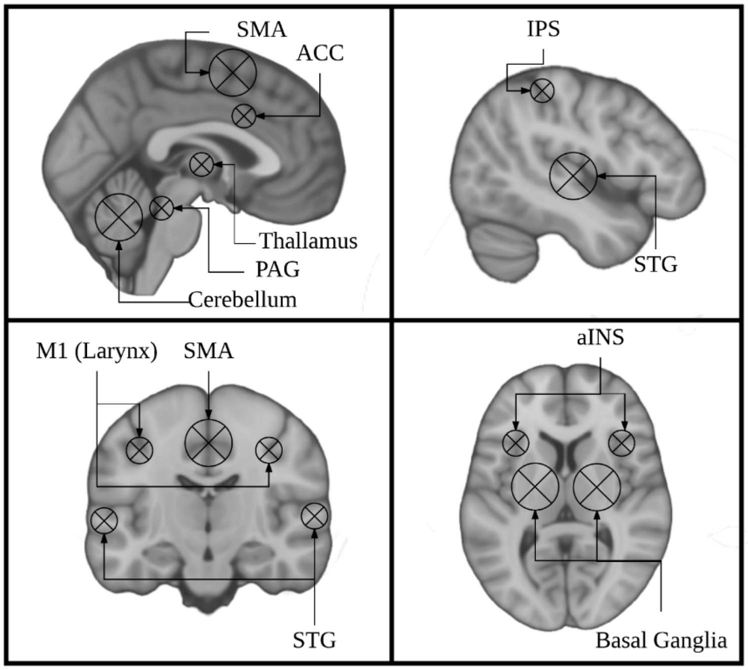

Figure 2-1 Key Regions | Key regions identified using singing as a model instrument provide a framework for interpreting findings from research on music and instrument playing.

in the scanner (Lotze et al. 2003). Consequently, this research lacked the auditory feedback necessary to directly compare singing to playing. One study made use of an fMRI compatible trumpet, but trumpet playing is not entirely independent of the vocal apparatus (Gebel et al. 2013). Other fMRI research on musical instrument playing has focused on the keyboard (J. L. Chen, Rae, and Watkins 2012; R. M. Brown, Zatorre, and Penhune 2015). However, the keyboard differs from the voice in two key ways. First, the keyboard is a discrete pitch instrument, in which the pitch of each note is mapped to a key in a one-to-one manner, and cannot be changed once a key has been pressed. In contrast, the human voice is a continuous pitch instrument which can produce any pitch within a range, not just discrete values, and can change this pitch on an ongoing basis (notably, in response to a feedback perturbation) (Burnett et al. 1998; Parlitz and Bangert 1999). This continuous nature of pitch is paralleled in fretless string instruments from the violin family. Second, in keyboard playing, pitch and timing are controlled by the same movement, a key press. In singing, pitch and timing are often decoupled; pitch is controlled by the tension of the vocal cords sound onsets and offsets are controlled primarily by the diaphragm. This distinction is paralleled by the asynchronous movements of the left and right hands required to play instruments from the violin family, which control pitch value and timing of sound onsets and offsets, respectively.

In the present study we directly test the hypothesis that singing and cello playing recruit an overlapping network of brain regions in auditory, motor, and auditory-motor integration regions. Specifically, we are interested in the auditory-vocal network. To accomplish our goal, we have developed an MR compatible cello device where sound feedback is delivered directly to the player during scanning (Hollinger and Wanderley 2013, 2015) (Figure 2-2). This cello uses optical sensors embedded in the fingerboard and bridge to capture finger position and string vibration, respectively. The optical sensor on the bridge of the cello provides real time, analog sound feedback. The MR cello lacks a resonant body, but the length of the fingerboard and strings, and, by extension, the locations of where fingers would be placed to produce specific notes are the same as those found on a standard full-size cello. We also constructed a miniaturized bow to fully leverage the continuous pitch nature of the cello. By comparing the neural correlates of singing to those of cello playing directly in the same individuals, we can determine the extent to which musical instrument playing makes use of functional brain networks that have evolved for singing.

We expect that cello playing and singing will recruit largely overlapping areas in premotor and supplementary motor areas, auditory cortices, and the cerebellum. We also expect that activation in primary motor cortex will be specific to the hand and larynx areas, respectively. Additionally, because playing in tune requires auditory-motor integration, we expect to see activation in areas that are thought to be responsible for perceptual monitoring and error correction (IPS, ACC, aINS, and thalamus). We also expect to see functional connectivity between these regions during both singing and cello playing, reflecting the network properties of these regions acting in concert to accomplish the task.

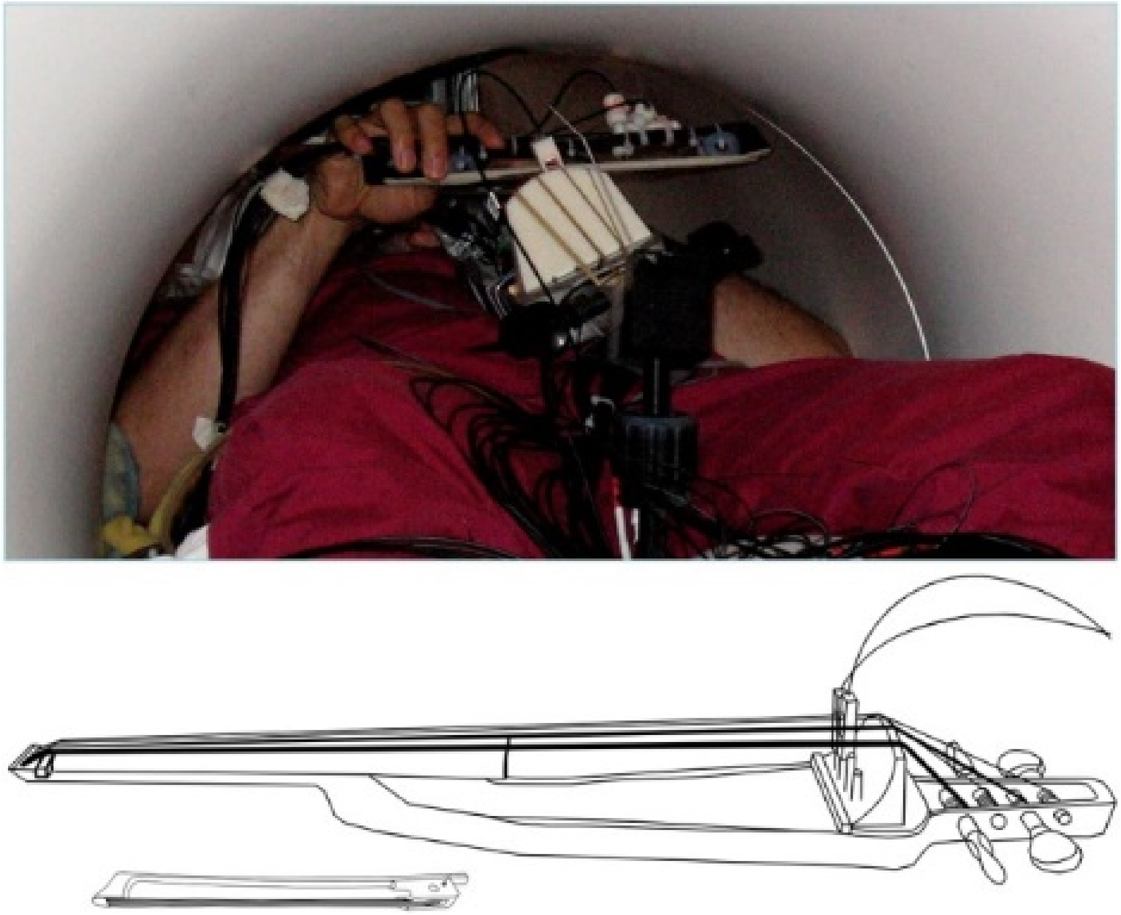

Figure 2-2 MR-Compatible Cello | (Top) View of the MR-Compatible cello and bow inside the scanner during experiment. (Bottom) Line drawing of the cello and bow developed for use inside the MR Scanner. Cello fingerboard has the same dimensions as a full sized cello to preserve the location and spacing of notes. Optic sensors at the bridge are used to detect string vibrations and are converted to analog audio output by a custom ARM board (not shown). The MR Compatible bow is 20cm long, which is just under a third of the length of a typical cello bow (approximately 70cm).

2.3 Materials and methods

2.3.1 Subjects

A total of 12 expert cellists (6 Female) were recruited from the Montreal community. All participants were right handed, had normal hearing, did not report any neurological disorders, and had no contraindications for the MRI environment (mean age = 21.4, mean years experience = 13.9, mean starting age = 7.4, mean practice hours per week = 24.8). Eleven participants were included in the final analysis, one participant was excluded from the analysis due to an equipment failure. This study was approved by and carried out in accordance with the recommendations of Montreal Neurological Institute Research Ethics Board and the McConnell Brain Imaging Centre. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

2.3.2 Experimental Paradigm

Stimuli and Task Conditions

Participants were asked to listen to and subsequently produce target tones both singing and on an MR-Compatible cello. There were three experimental conditions: Sing/Play, Listen, and Rest. For all three conditions, we presented auditory target tones that were 2 s long. Presented tones were either E3, F$^{\#}$3, G$^{\#}$3 for cello, and E3, F$^{\#}$3, G$^{\#}$3 or E4, F$^{\#}$4, G$^{\#}$4 for singing depending on participant's vocal range. Tones were recorded by either a female vocalist, male vocalist, or on a cello. Tone presentation was followed by a pre-recorded auditory instruction to either Listen, Sing/Play, or Rest. On some trials auditory feedback was either masked or pitch shifted (see below).

Participants underwent a familiarization session followed by an fMRI session. For both the familiarization session and the fMRI session, a microphone was suspended approximately 2 inches from their mouth, the MRI compatible cello device was placed along their torso using a specialized support, and headphones were provided (Sensimetrics S14 fMRI insert headphones, Dayton Audio DTA-1 amplifier). The microphone and cello were connected to a Mackie 802VLZ4, 8-channel mixer and to a midi controlled TCHelicon VoiceOne pitch shifter which was used to prevent the audio feedback from reaching the headphones on certain trials. Pink noise was played through the headphones to reduce bone conduction so that audio was being delivered exclusively through the headphones. All volume levels were adjusted on a per subject basis, but on average pink noise was presented at 78.3 dB SPL A and auditory targets were presented at approximately 15.6 dB above the noise floor. The experiment was run using custom scripts written in python. All vocalizations and cello sounds were digitally recorded using a Sound Devices 744T digital recorder (Figure 2-4).

A sparse sampling paradigm was used for the fMRI session (Belin et al. 1999), where a long delay in TR was used to allow tasks to be carried out in the relatively silent period between functional volume acquisitions, thus minimizing acoustical interference and also avoiding movement-related artifacts since the scanning takes place after the motor production for each trial (Figure 2-5).

Procedure

To allow participants to adjust to the fMRI-compatible cello and the constraints of playing it in the scanner, each person underwent a 45 min familiarization session no more than 1 week prior to their session in the MR-scanner. During the familiarization session each participant was asked to lie on a foam mat inside a structure that simulated the space constraints of the MRI environment. All participants were asked to perform a series of scales, which contained all of the target notes that would be presented during the experiment and a reduced duration (10 min version) of the experimental task.

On performance trials, participants were instructed to sing or play back the target tone for 4 s. For singing trials, participants were instructed to sing with closed lips in order to reduce breathing artifacts in the recorded signal and movement artifacts in the fMRI signal. For cello playing trials they were instructed to use as few bows as possible to reduce movement artifacts (approximately 1s per bow). Between trials participants were instructed to keep their hands on the cello and to move as little as possible. During familiarization, participants went through 2 reduced-length experimental runs (one cello, one singing), with all conditions included in each run). Three participants underwent a second familiarization session due to equipment problems during their first scheduled session.

Within 1 week of the familiarization session, participants were tested in the Siemens Trio 3T magnetic resonance (MR) scanner at the Brain Imaging Center of the Montreal Neurological Institute. Each participant was fitted with MR-compatible headphones. The MR-compatible microphone was attached to the mirror support system. The MR-compatible cello was laid across the torso using a special MR-compatible stand. Sound levels were adjusted on a per participant basis so that, during trials with masked auditory feedback, participants could not hear their voice or the sound of the cello above the pink noise.

During the fMRI session, participants performed two cello playing runs and two singing runs. Run order was counterbalanced across participants. During each of these runs, trial order was pseudo-randomized. Following the presentation of each target tone, participants were instructed to sing or play the cello, or to listen. On rest trials no auditory target was presented. On some of the performance trials, auditory feedback was pitch shifted (40 trials, up to one semitone) and on others it was fully masked (20 trials). Due to technical issues with the pitch shifter during data acquisition, these data were not included in the final analysis. The no audio condition and listen condition served as controls for the auditory and motor aspects of the performance trials.

MRI Acquisition

A high resolution T1-weighted anatomical scan (voxel = 1 mm^3) was collected between runs 2 and 3. During the 4 functional runs, one whole-head frame of 28 contiguous T2*-weighted images were acquired (Slice order = Interleaved, TE = 85 ms, TR = 10 s, Delay in TR = 7.7 s, 64 × 64 matrix, voxel size = 4 mm^3). All tasks were performed during the 7.7 s silent period between functional volume acquisitions. As such, the tasks were done in silence. Timing of the auditory stimulus presentation was varied randomly by up to 500 ms to increase the likelihood of obtaining the peak of the hemodynamic response for each task. Within each run, each condition was presented 10 times for a total of 20 acquisitions per condition for singing and 20 acquisitions per condition for cello playing. A high resolution whole brain T1-weighted anatomical scan (voxel = 1mm^3) was collected between runs 2 and 3.

2.3.3 Behavioral Analyses

Individual trials of singing and cello playing were analyzed with pitch information extracted from audio signals. Each trial's audio was first segmented from the MR-compatible microphone recording by hand using Audacity software. The trials were then processed using a custom analysis pipeline implemented in Python, with a GUI for visualizing and optimizing analysis parameters.

The ambient noise in the scanner room had a peak resonance of 160 Hz that interfered with the extraction of fundamental target pitches, so harmonics 3–10 of the cello and singing tones were used for pitch extraction. To reject room noise and to isolate harmonics of interest, the raw microphone signal was high-pass filtered with a cutoff at 367 Hz and low-pass filtered with a cutoff at 4,216 Hz. Pitch estimation was then performed using the YinFFT algorithm provided in the Python module Aubio (Brossier 2007), producing a time-series of pitch estimations (detected harmonic, in Hz) and confidence ratings (between 0 and 1). Estimates were adjusted to their representative fundamental pitches before selecting stable pitch regions for further analysis. Stable pitch regions were defined as: segments of at least 150 ms in which the rate of change of the pitch did not exceed 100 Hz/s (or approximately 0.07 Hz per 32-sample pitch estimation window at the sampling rate of 44,100 Hz). Of these regions, only those that maintained a confidence rating of at least 0.7 were included. Trials were rejected if no regions were found to meet the stability and confidence criteria (see Figure 2-3). In total, 92.4% of trials were retained. Rejected trials were excluded from the fMRI analysis.

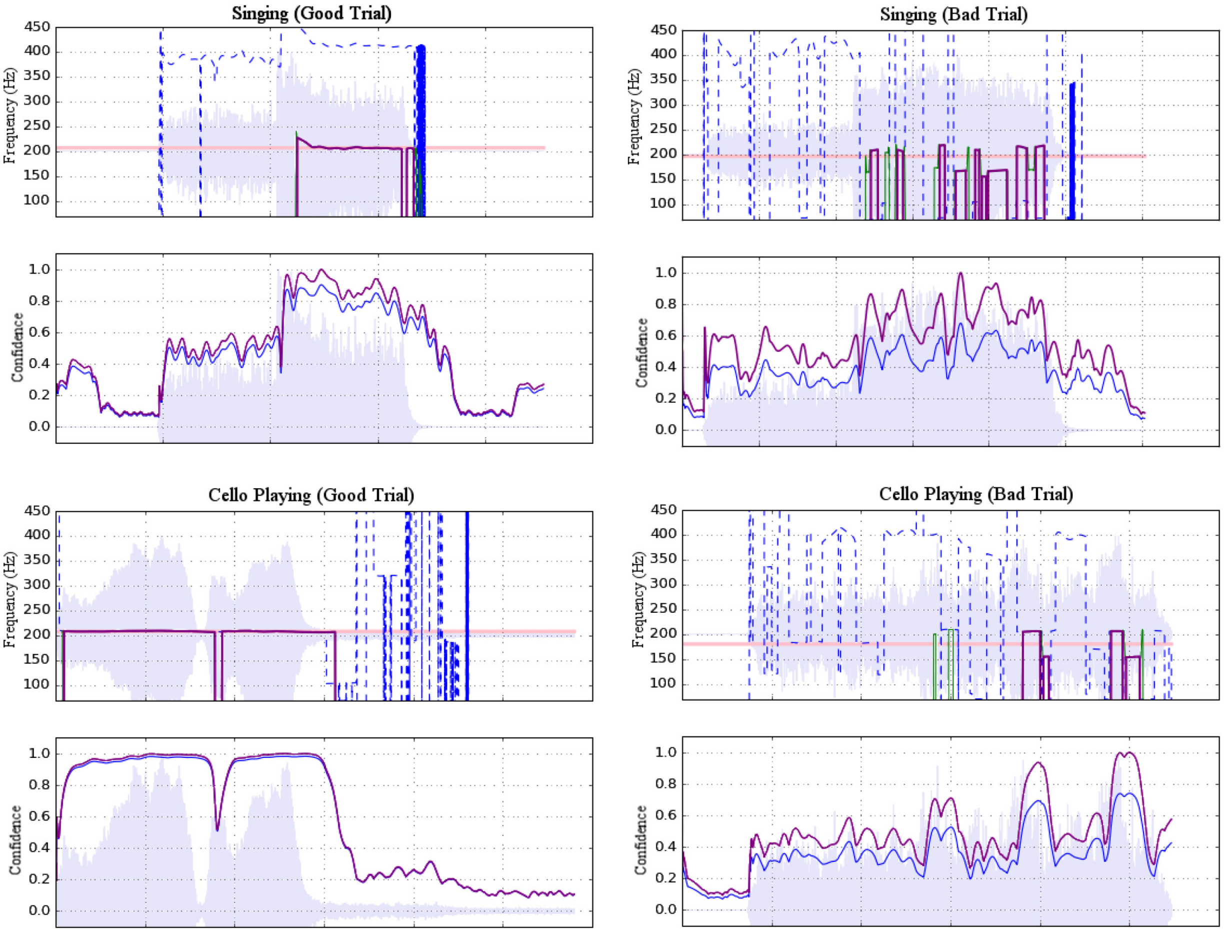

Figure 2-3 Behavioural Analysis Software | Examples of accepted **(left)** and rejected **(right)** trials for both singing **(top)** and cello playing **(bottom)**. Each example shows a plot of pitch estimations above a plot of confidence ratings. For the former, dashed blue lines represent raw pitch estimation whereas solid purple lines represent potentially stable pitch estimates adjusted to the fundamental. For the latter, raw confidence ratings are shown in blue and normalized confidence ratings are shown in purple. Shading indicates amplitude envelope.

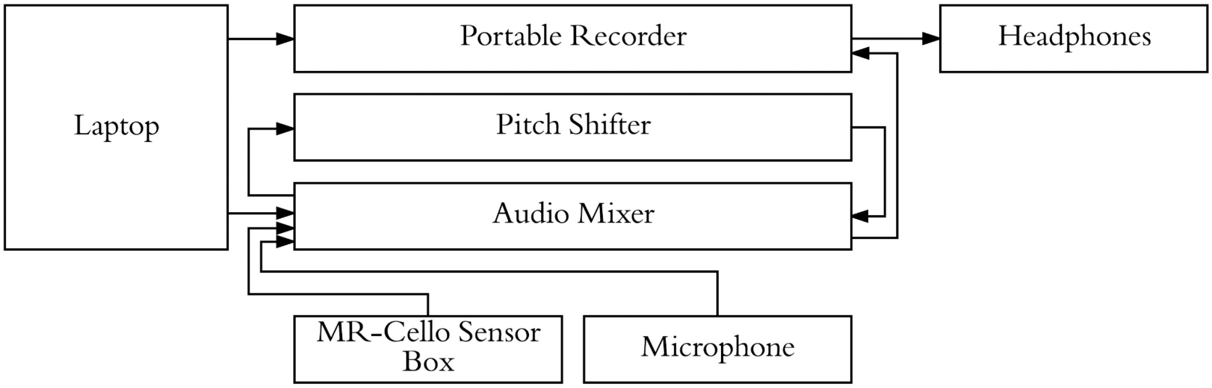

Figure 2-4 Experimental Setup | Experimental setup used to present stimuli, cello audio, and singing audio through headphones. Pitch shifter allowed for audio feedback from cello/singing to be blocked on specific trials while still presenting audio stimuli.

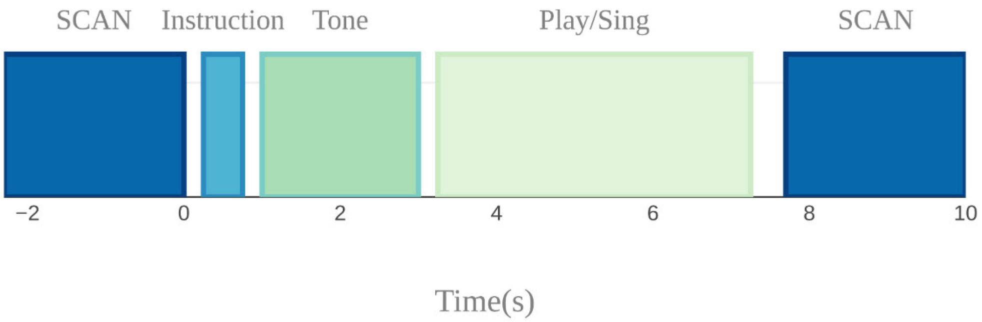

Figure 2-5 Sparse Sampling Design | Sparse sampling design used to avoid auditory and motion artifacts. Auditory target presentation, instruction presentation, and singing and cello playing were done in the silent 7.7 s period between scans.

Pitch accuracy was calculated on a per-trial basis as the deviation between the produced and target pitches, expressed in cents using Equation (Equation 2-1). An overall accuracy score for each participant was then determined by calculating their mean pitch accuracy across all trials. Finally, scores were analyzed using a two way (instrument x target) ANOVA implemented in R (R Development Core Team, 2008).

$$cents = 1200^{*}log_2\left(\frac{Freq_{produced}}{Freq_{target}}\right)$$

Equation 2-1 Conversion from absolute frequency difference to normalized cent difference

2.3.4 fMRI Analyses

fMRI data were analysed using the FSL5.0 FEAT toolbox (Jenkinson et al. 2012). Brain extraction was carried out using BET2. Functional volumes were aligned to the high resolution anatomical and then to MNI152 standard space using FLIRT linear registration with 12DOF. Motion parameters were estimated using MCFLIRT and fMRI timecourse was temporally filtered to remove drifts greater than 300ms. To boost the signal to noise ratio, images were spatially smoothed with an 8 mm FWHM kernel. FLAME-1 mixed effects modeling was used to fit the GLM to the fMRI signal.

Significance was determined using an FSL cluster probability threshold of $p < 0.05$ with a voxel-wise significance level of $z = 3.3$ ($p < 0.01$). The cluster probability threshold serves as a correction for multiple comparisons. Four contrasts were carried out: Singing vs. Rest, Cello playing vs. Rest, Cello Playing vs. Singing, and Singing vs. Cello Playing. Additionally, task performance (on a per-subject basis) was regressed against the BOLD signal for each of these contrasts. Statistical conjunctions were carried out using the conjunction script created by the Warwick University Department of Statistics, which also made use of the FSL tools (Nichols et al. 2005). This script carries out a voxel-wise thresholding of $p < 0.05$ in both conditions of interest, and then carries out a cluster correction of $p < 0.05$.

Functional Connectivity analyses were carried out using the FSL5.0 FEAT toolbox. A seed region in auditory cortex was identified by masking the conjunction of singing and cello playing from the functional data with an anatomically defined mask of Heschl's Gyrus (Harvard Structural Brain Atlas). Seed regions in primary motor cortex (M1) were identified by masking singing and cello playing with an anatomically defined mask of post-central gyrus (Harvard Structural Brain Atlas). An additional seed region was identified using a functionally defined mask of larynx area in M1 from Kleber et al. The activation timecourse in each of these regions was extracted and correlated with the whole-brain timecourse for each task of interest, which was estimated using the GLM. Correlated voxels were thresholded as described above. Regions that showed a correlated timecourse were then linearly regressed with task performance to determine whether those areas were contributing directly to good intonation.

2.4 **Results**

2.4.1 Task Performance

We first carried out a behavioral analysis to confirm that our participants, expert cello players, could sing 3 target tones and also play them on the MR-Compatible cello with at least quarter tone accuracy (50 cents). By performing a two way ANOVA (instrument by target tone), we found that participants could produce each of the three target tones within the specified accuracy both when singing (mean deviation from target = −8.9 cents, stdev = 58.9 cents) and when playing the cello (mean = −7.79 cents, stdev = 85.39 cents) (Figure 2-6). There was no significant effect of instrument ($p < 0.65$, $F = 0.21$), but there was a significant effect of tone ($p < 7.14 \times 10-4$, $F = 11.54$) and a significant tone by instrument interaction ($p < 4.21 \times 10-8$, $F = 30.65$). *Post-hoc* tests showed that, when playing the cello, participants tended to be flat on the highest tone (mean = −39 cents) and that, in singing, they tended to be flat on the lowest note (mean = −15 cents). The undershoot on the highest note for cello playing was likely because the note was the most difficult to reach within the confines of the scanner. The undershoot on the lowest note for singing, while significant, is within an eighth tone of the target pitch which is well below the quarter tone threshold for considering a note in tune. There was no within-subject correlation between performance on the cello and performance on the singing trials.

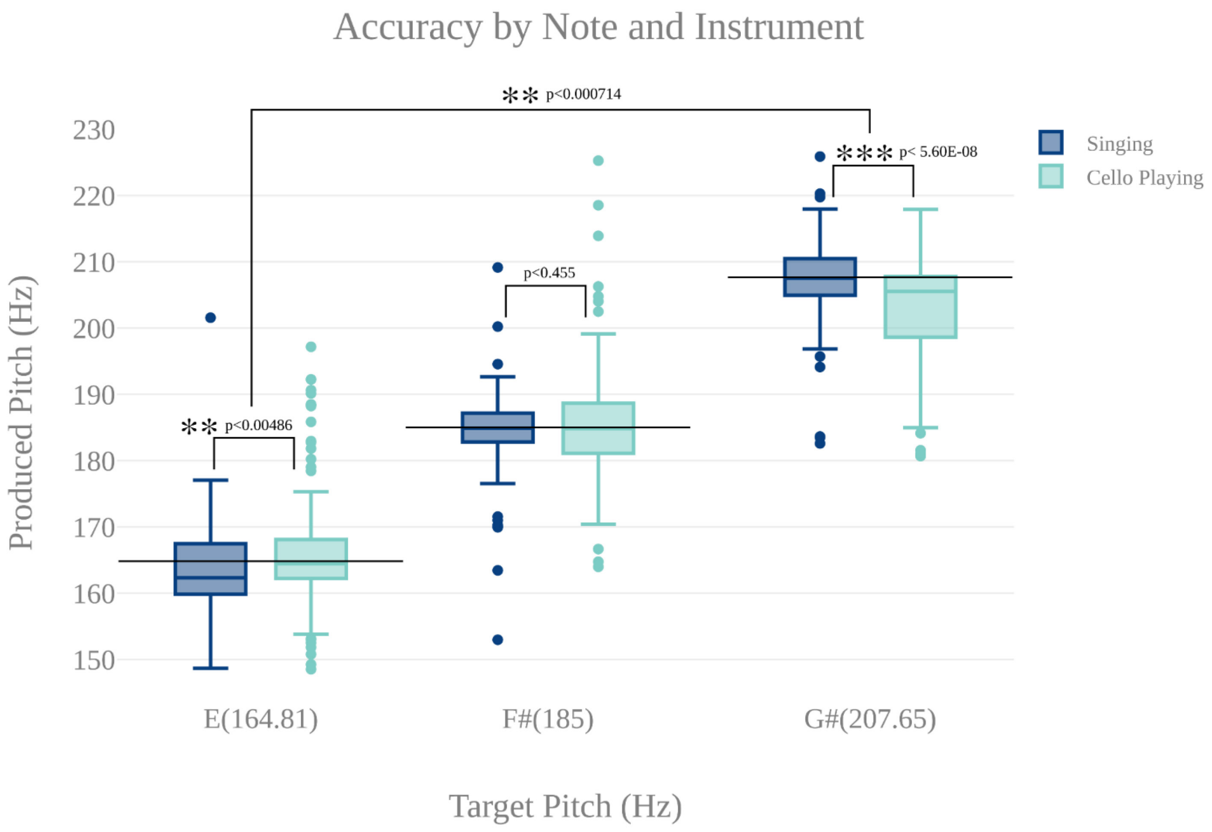

Figure 2-6 Accuracy by Note | Accuracy by note for singing (dark blue) and cello playing (light blue). There is a highly significant main effect of note, but no main effect of instrument. There is a significant note by instrument interaction showing that participants were slightly flat on the lowest note when singing and slightly flat on the highest when playing the cello. However, mean produced tones were within a quarter tone accuracy (50 cents).

2.4.2 fMRI Findings

Similarities Between Singing and Cello Playing

To test the hypothesis that singing and cello playing activate a shared set of brain areas, we first identified the respective networks by performing two contrasts: Singing vs. rest and Cello playing vs. Rest. The Singing vs. Rest contrast was used to test the hypothesis that the singing task would activate the auditory-vocal network identified in previous literature (Kleber and Zarate 2014). We found that, consistent with the areas reported in the literature, singing activated Pre and Post-central gyrus (R > L), SMA extending to ACC (bilateral), IPS (R), the length of the STG extending to supramarginal gyrus, STS, aINS (bilateral), cerebellum VI, VIIa, and CrusI-II (bilateral), thalamus extending into caudate (bilateral), the globus pallidus (bilateral), putamen (bilateral), and PAG, and the pons. In addition to areas within the auditory-vocal network, we also saw activity in middle frontal gyrus (right) (Figure 2-9). This contrast did not show larynx-specific activation in M1; however, we found that larynx activation was present at an uncorrected threshold of $p < 0.05$. To confirm this finding with a more focused though statistically stringent approach, we performed a fixed-effects analysis consistent with what has been done in previous studies of larynx function (S. Brown, Ngan, and Liotti 2008). This analysis showed significant activation in larynx area of both right and left M1 ($\text{zstat}_{\text{right}} = 5.3$, $\text{zstat}_{\text{left}} = 4.4$), in locations consistent with that reported by Brown et al in M1 for phonation ($d_{\text{right}} = 9.38$ mm, $d_{\text{left}} = 9.16$ mm) (S. Brown, Ngan, and Liotti 2008) Equation (Equation 2-2).

$$\mathrm{d}(\mathrm{p},\mathrm{q}) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}$$

Equation 2-2 Euclidean distance between two points

For cello playing vs. rest, many of the same auditory-vocal areas were active compared to singing (Figure 2-9). Again, activity clusters are seen throughout the pre- and post-central gyri (bilateral), SMA extending into ACC, and IPS. Activity also extends the length of the STG and superiorly into the supramarginal gyrus. Clusters are also seen in middle frontal gyrus (right), and in the anterior insula (bilateral). In the cerebellum, activation extends through VI, VIIIa, VIIIb, VIIb, CrusI-II, Vermis CrusII, and Vermis VI. Clusters of activity are also seen in thalamus (bilateral) extending into caudate (right), and in the globus pallidus (right).

To specifically test the hypothesis that singing and cello playing both engage some of the same components of the auditory-vocal network, we carried out a statistical conjunction of the cello vs. rest and singing vs. rest conditions (Nichols et al. 2005). The conjunction showed overlapping activation throughout the auditory-vocal network. However, no overlap was seen in Putamen, or Brainstem. In the cerebellum, overlapping activity was seen in VI, CrusI-II, VIIIa, and VIIIb (Figure 2-9).

Task accuracy was then linearly regressed against both singing vs. rest and cello playing vs. rest to determine which of the active areas were correlated with, and therefore likely to be contributing to, task accuracy. In both cases, areas that were active for the task were positively correlated with task performance (no negative correlations were observed). Specifically, for singing, all regions of the auditory-vocal network were more active as participants performed better. For cello playing, pre- and post-central gyri, STG extending into the supramarginal gyrus, aINS, cerebellum, and thalamus were more active in participants that performed better (Figure 2-9).

We further hypothesized that singing and cello playing would show overlap in areas whose activity contributed to higher accuracy. To test this hypothesis, we carried out a statistical conjunction of the singing vs. rest and cello vs. rest regressions by task performance. This conjunction showed overlap in SMA extending into ACC, IPS, middle frontal gyrus (right), STG, supramarginal gyrus, the aINS, and thalamus extending to caudate. In the cerebellum, there was overlapping activation in VIIb, VIIIa, VIIIb, and VI (Figure 2-9).

The previous analyses found that singing and cello playing activate a shared set of brain regions, and that activity in many of these regions is positively correlated with task accuracy. Building on these findings, we decided to test the hypothesis that areas within this set are functionally connected in both singing and cello playing (Figure 2-10). To accomplish this goal we performed a functional connectivity analysis using the activity in Heschl's gyrus (bilateral, from Harvard Brain Structural Atlas) as a seed region and correlating activity within this seed with activity in the rest of the brain on a voxelwise basis. For singing, we saw correlated activity in auditory cortices of the STG (bilateral) both within and around the seed area, supramarginal gyrus (bilateral), pre-central gyrus (right), inferior frontal gyrus (bilateral), and in VIIIa of the cerebellum (left). For cello playing, we saw correlated activation within the seed region and also in posterior STG extending into the supramarginal gyrus, pre- and post-central gyrus, and VIIIa, VIIIb, VIIb, VI (left), and Vermis VI of the cerebellum. The conjunction of singing and cello playing functional connectivity analyses showed shared connectivity in Heschl's gyrus, pre-central gyrus, and VIIIa of the cerebellum (left) (Figure 2-10).

Functional connectivity in both singing and cello playing using the same seed region was then linearly regressed with task performance to determine the correlation between connectivity and pitch accuracy. In singing, higher functional connectivity was positively correlated with performance in the supramarginal gyrus, and in cerebellar VIIa (left) and VIIb (right). In cello playing, higher functional connectivity was positively correlated with performance in pre- and post-central gyrus, the supramarginal gyrus, and cerebellar VIIIa and VI (bilateral) (Figure 2-10).

Differences Between Singing and Cello Playing

In order to characterize the differences in brain activity between singing and cello playing, two contrasts were first carried out: singing vs. cello playing and cello playing vs. singing (Figure 2-9). For singing vs. cello playing, no significant clusters of activation were observed after correcting for multiple comparisons. However, the PAG and Putamen were active in the Singing-Rest contrast, and not in Cello Playing-Rest, or the conjunction of Cello-Rest and Singing-Rest. Additionally, at an uncorrected threshold of $z = 1.8$ differences can be seen in larynx area of motor cortex. In order to directly test the hypothesis that the larynx area of motor cortex was more active for singing than for cello playing, which was one of our predictions, we performed a region of interest analysis of post-central gyrus using the mask of laryngial motor area described above. This analysis showed more activation for singing than for cello playing in larynx area. We further tested this hypothesis by performing a 2 factor ANOVA (Instrument by Region of Interest) with spherical ROI in hand motor regions (bilateral) and larynx area of motor cortex (bilateral). This analysis showed significant effects of both Instrument ($p < 8.28 \times 10\text{-}5$, $F = 19.68$) and ROI ($p < 2.42 \times 10\text{-}2$, $F = 5.53$), as well as a significant Instrument x ROI interaction ($p < 1.03 \times 10\text{-}6$, $F = 34.47$) driven by a highly significant difference between cello playing and singing in hand motor regions ($p < 1.71 \times 10\text{-}5$, $F = 33.6$). However, no significant difference was observed in larynx area ($p < 0.13$, $F = 2.45$) (Figure 2-7). To understand why the contrast of singing vs. cello playing showed sub-threshold differences in larynx area but the ANOVA did not, we decided to consider potential covariates. Based on previous research showing that early trained musicians show advantages in pitch tasks, we hypothesized that early trained musicians may use larynx area of motor cortex to a greater extent than late trained musicians. To test this, we included starting age as a covariate. While starting age itself did not significantly correlate with larynx activation in either singing ($r = -0.31$, $p < 0.34$) or cello playing ($r = -0.41$, $p < 0.26$), the contrast of singing vs. cello playing in larynx area was significant when accounting for the effect of starting age ($p < 0.0091$, $F = 8.3$)

(Figure 2-7). Additionally, when data were divided into early and late starting groups (as in Penhune 2011), cellists that began musical training before age 7 (*n* = 5) showed significantly more activation in larynx area during cello playing than did those that started age 7 and up (*n* = 6) ($p < 0.01$, $F = 10.26$) (Figure 2-7).

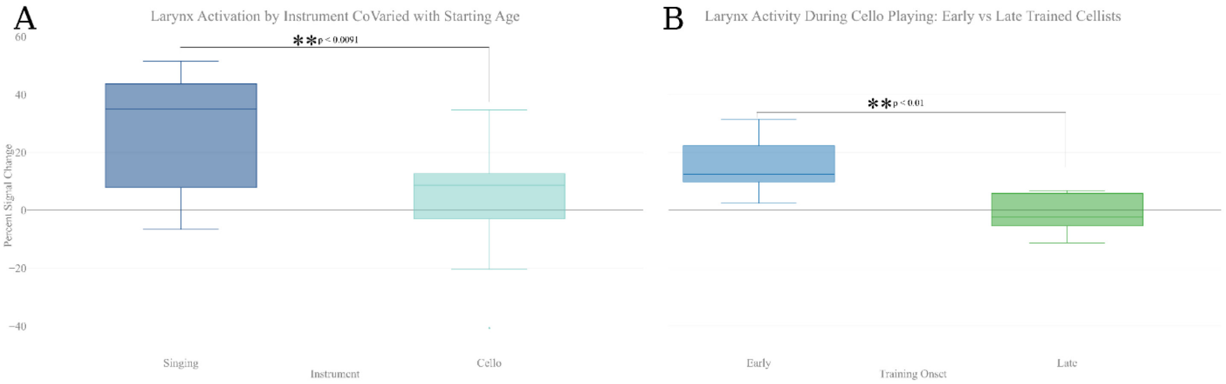

Figure 2-7 Larynx Activation by Instrument (A) Contrast of parameter estimates (units arbitrary) in larynx area of motor cortex for singing (blue) and cello playing (turquoise). When starting age is included as a covariate, ANOVA shows a significantly more activity in larynx area during singing than cello playing. (B) Mean contrast of parameter estimate values in larynx area of motor cortex during cello playing for Early trained (green) and Late trained (gray) cellists. Cellist that started training before age 7 show significantly more activity in larynx during cello playing than do those that started age 7 and later.

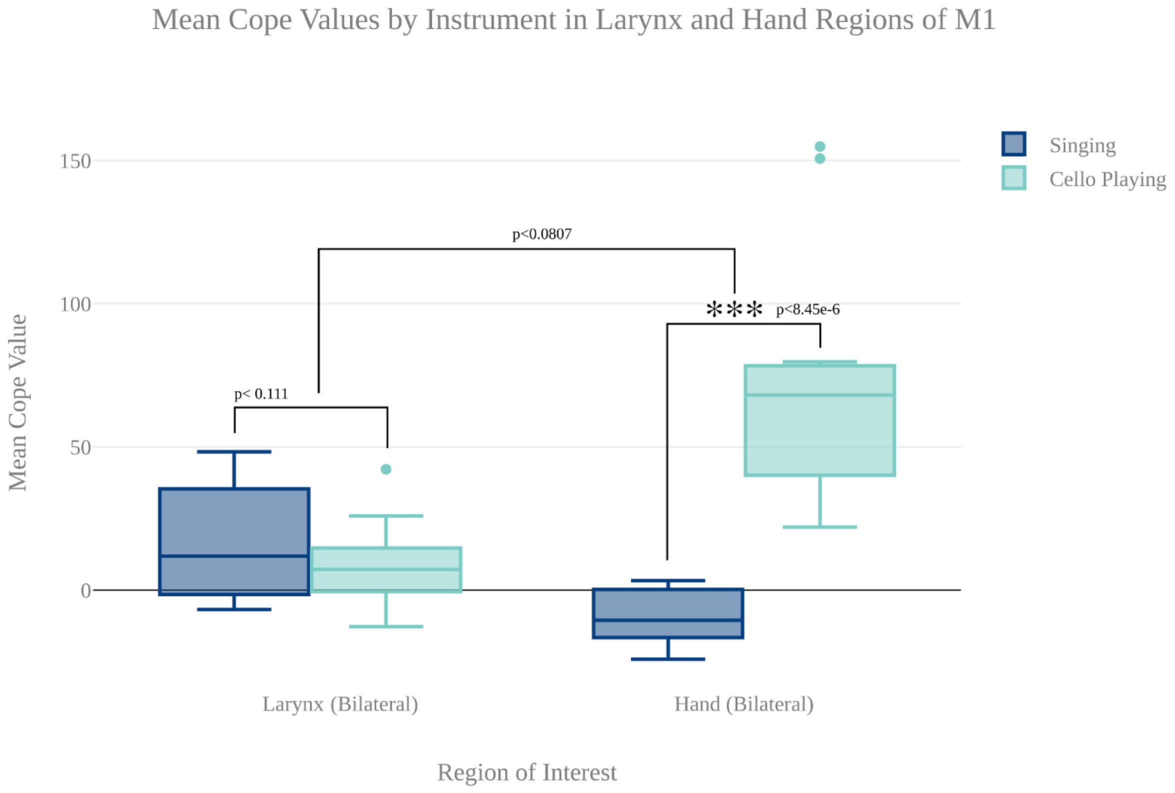

Figure 2-8 Motor Effector by Instrument Interaction | Contrast of parameter estimates (units arbitrary) in larynx and hand area of motor cortex for singing (blue) and cello playing (turquoise). The ANOVA showed no significant effect of instrument or brain area, but did show a significant instrument by region interaction.

For cello playing vs. singing, there was more activation in STG extending to the SMG, pre- and post-central gyri, and in the posterior insula. In the cerebellum, cello playing showed more activation than singing in I–IV, in VI and in VIIIb (bilateral). In pre- and post-central gyri as well as in the cerebellum, the activation peaks were centered on areas associated with hands/arms (Yousry et al. 1997). No significant differences in functional connectivity were observed when directly contrasting singing and cello playing.

2.5 **Discussion**

2.5.1 Findings

Four conclusions may be reached from this research study. First, that brain activity during cello playing directly overlaps with brain activity during singing in many areas within the auditory-vocal network. These areas include dorsal motor and premotor areas, SMA and Pre-SMA, STG, ACC, aINS, IPS(R), and Cerebellum but, notably, exclude the PAG and Basal Ganglia (Putamen). Second, that activity within the overlapping areas is positively correlated with, and therefore likely contributing to, both singing and playing in tune as shown by correlations with performance measures. Third, that activity in auditory areas is functionally connected with activity in dorsal motor and pre-motor areas, and that the connectivity between them is positively correlated with good performance on this task. This functional connectivity suggests that the brain areas are working together to contribute to task performance and not just coincidently active. Last, our findings showed that cello playing may directly co-opt vocal areas (including larynx area of motor cortex), especially if training begins before age 7.

2.5.2 Questions Answered

Neural Correlates of Learned, Arbitrary Associations

This study provides evidence that relatively new auditory-motor integration tasks like stringed instrument playing make use of the auditory-vocal network, which is thought to be an evolutionarily old system (Figure 2-9). The interpretation that cello playing makes use of neural mechanisms that evolved for singing is consistent with the theory of Neuronal Recycling proposed by Dehaene and Cohen (2007). This theory proposes that cultural tasks (like arithmetic) are too new to be the product of evolution and that, as a result, they have to make use of cognitive mechanisms that are in place for more basic tasks (like direction processing). We propose that our findings are an example of the same concept but in the auditory-motor domain. The auditory-vocal network used for singing develops without explicit training, much like spatial processing in the visual domain. After explicit instruction, cello playing brain activity patterns overlap with singing throughout the auditory-vocal network. Potentially the best point of evidence supporting this interpretation is our finding that cellists that began training before age 7 playing activated the larynx area of motor cortex during cello playing despite cello playing not relying on the larynx. We cannot rule out the possibility that cellists were humming subvocally during the cello playing task, though we did rule out the possibility that they were actually humming using the continuous microphone recording. However, if subvocal humming was responsible for the larynx activation observed during cello playing, it seems likely that other vocalization specific areas like the PAG would also be recruited, which is not the case, and seems unlikely that we would see a starting age effect. In addition, no significant activation was observed in basal ganglia or brainstem areas that, in singing, are active even during imagined singing (Kleber et al. 2007).

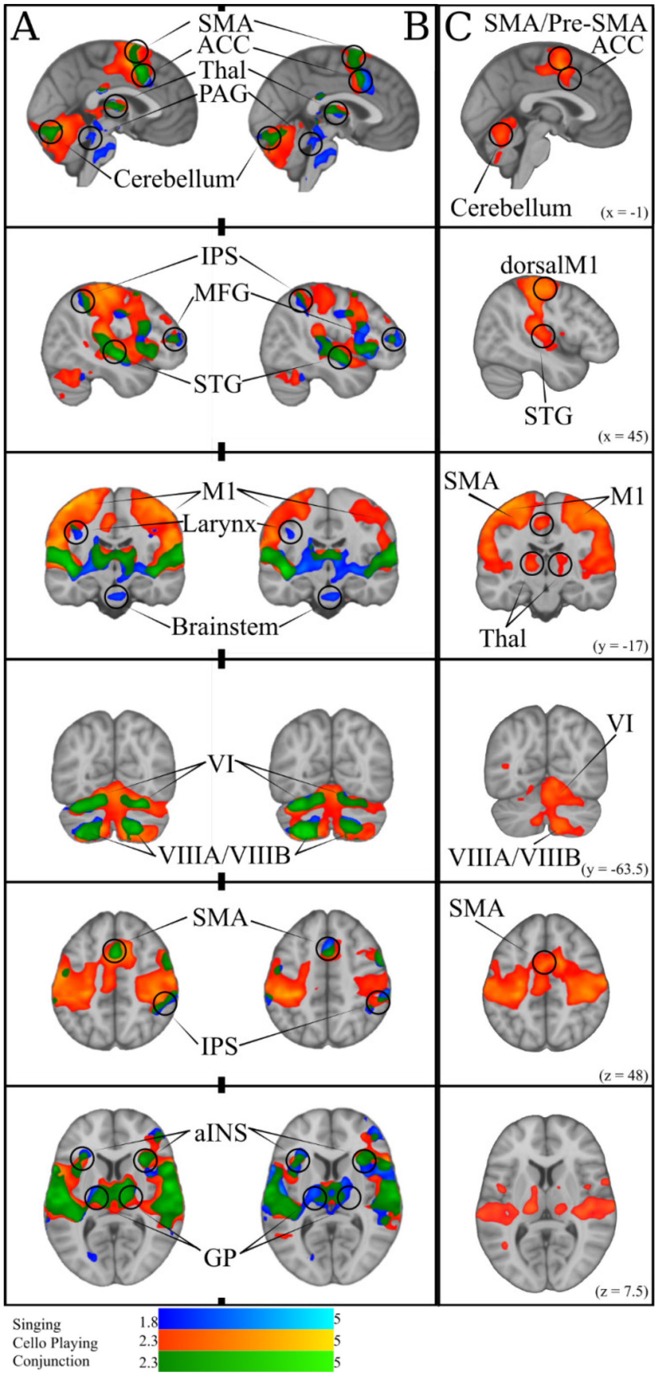

Figure 2-9 Brain Network for Singing and Cello Playing | Simple Singing (blue), cello playing (orange), and conjunction (green). **(B)** Simple singing (blue), cello playing (orange) and conjunction (green) regressed by task accuracy with better performance correlated with more activation. The conjunction of singing and cello playing shows overlapping activation in SMA, ACC, Thalamus, PAG, IPS, MFG, STG, MI (larynx area for singing), aINS, GP, and VI, and VIIIA/VIIIB of the cerebellum. All of these regions were positively correlated with better intonation. **(C)** Cello playing > Singing throughout dorsal motor and premotor regions, SMA/Pre-SMA., posterior STG, Thalamus, and Cerebellum.

The Role of the ACC and aINS

Some brain areas, like the ACC and aINS, are reported to be active both in studies of musical instrument performance and singing, but are given different interpretations according to the task. In studies of piano performance, the activity of the ACC is thought to reflect coordination of hand movements (Parsons et al. 2005) while in studies of singing the ACC is thought to be specifically involved in initiating vocalizations (Kleber et al. 2007) similar to previous work done in animal models (Uwe Jürgens 2002). Similarly, the aINS is thought to integrate somatic information from the body to support bi-manual coordination in piano playing (Parsons et al. 2005) and to coordinate the vocal musculature during singing (Kleber et al. 2007) together with other interoceptive inputs (Kleber et al. 2013). In our study, we found that activity in the aINS and ACC was directly overlapping during cello playing and singing. Consequently, we propose that the role of the aINS and ACC is to coordinate the movements of whichever motor effectors are required in order to produce pitched sounds, and that their activity is not specific to any one motor system.

The Role of the Brainstem

One of most prominent differences observed between singing and cello playing was that singing activated the brainstem, including in the PAG, while cello playing did not. The PAG is one of the key regions identified through singing research as instrumental to producing vocalizations (Kleber and Zarate 2014), a finding which we replicated in the present study. However, the lack of activation in cello playing suggests that not all regions in the vocal-motor network are co-opted in cello playing. While it may be the case that regions like the ACC and aINS, or even the larynx area of motor cortex, can be re-purposed to coordinate activity of motor systems other than those required to produce vocalizations, lower level physiological systems like the brainstem are likely exclusive to vocal control and respiration. Early animal work on the PAG found that it was the first point at which stimulation produced “normal sounding” vocalizations (Uwe Jürgens and Pratt 1979b), and later work found that different types of electrical activity in the PAG directly correlated with adduction and abduction of the laryngial and respiratory musculature during vocalizations in non-human primates (Larson 1991). While systems that coordinate breathing may come into play for more complex instrumental performance aspects like phrasing which are tightly coupled to respiration (Watkins and Scott 2012), these systems may not be directly involved in the hand/arm movements required to produce single notes during the investigated task. Without recording muscle activity of the larynx during cello playing, we cannot rule out the possibility that cello playing causes larynx activity. However, we can say conclusively that playing the cello did not incidentally produce vocalization during this task and, consequently, that the descending signals from the brainstem to the musculature were specific to each instrument.

The Role of the Putamen

Another difference observed between singing and cello playing was that singing activated the putamen while cello playing did not, though both cello playing and singing activated the GP. The finding that cello playing and singing both recruit the GP is consistent with previous work in both keyboard playing (e.g., Parsons et al. 2005) and singing (e.g., Zarate and Zatorre 2008). However, research in singing has also shown that recruitment of the putamen is specifically linked to expertise, with expert singers recruiting the putamen to a greater extent than novice singers when compensating for, or ignoring, introduced pitch perturbations (Zarate and Zatorre 2008). In this study, neither experts nor novices recruited the putamen when simply singing single notes without an introduced manipulation. The interpretation given to this finding in Zarate and Zatorre (2008) is that the putamen is likely involved in correcting for perceived errors in auditory feedback, and that singing single tones was simple enough that no real error correction was needed. They also note that lesions of the putamen have been linked to dysarthria (Uwe Jürgens 2002). Putamen activity has also been linked to over-learned automatic responses in motor sequence learning across a number of studies (Lehéricy et al. 2005; Penhune and Steele 2012) . In our research, no feedback manipulation was introduced either during cello playing or singing. As such, it could be the case that participants were correcting for incorrect intonation during singing and not cello playing. However, both singing and cello playing show a higher degree of pitch variability at the start of trials, compared to the end. This suggests that corrections to produced pitch were being carried out in both cases. Following from this, it could be the case that the putamen was not recruited for cello playing during our experiment because the function of the putamen is unique to the vocal domain. However, this would be in conflict with findings regarding the putamen's role in other types of sensorimotor adaptation (e.g., Seidler, Noll, and Chintalapati 2006) and auditory-motor integration tasks like tapping to the beat (Kung et al. 2012). More likely, it is the case that, similar to the aINS and ACC, the putamen servers a more domain general role in auditory-motor integration, and was not shown to be active for cello playing due to a lack of statistical power. We predict that the putamen will be recruited for both cello playing and singing in future studies involving tasks that more directly probe for auditory-motor integration.

The Association Between Auditory and Motor Regions

This study also provides direct evidence supporting the idea that playing single notes on the cello not only recruits many of the same brain areas as singing, but that it makes use of the same network of brain regions. First, we found that good intonation is positively correlated with functional brain activity within the areas that are recruited both for singing and cello playing (Figure 2-9). This shows that the same activity in both instruments plays a role in accomplishing the same behavioral goal. Second, we found that auditory (bilateral HG) and motor (dorsal pre-motor, SMA) regions within the areas common to singing and cello playing were functionally connected during both tasks, and that the degree of functional connectivity is positively correlated with good intonation (Figure 2-10). In other words, the same brain areas are working together to accomplish both singing and playing the cello in tune during the presented task. This finding addresses the potential criticism that the brain areas observed in the GLM analyses are coincidently active but not necessarily interacting. The functional connectivity findings are also consistent with previous work showing that singing recruits a functionally connected network of brain areas (Zarate, Wood, and Zatorre 2010)

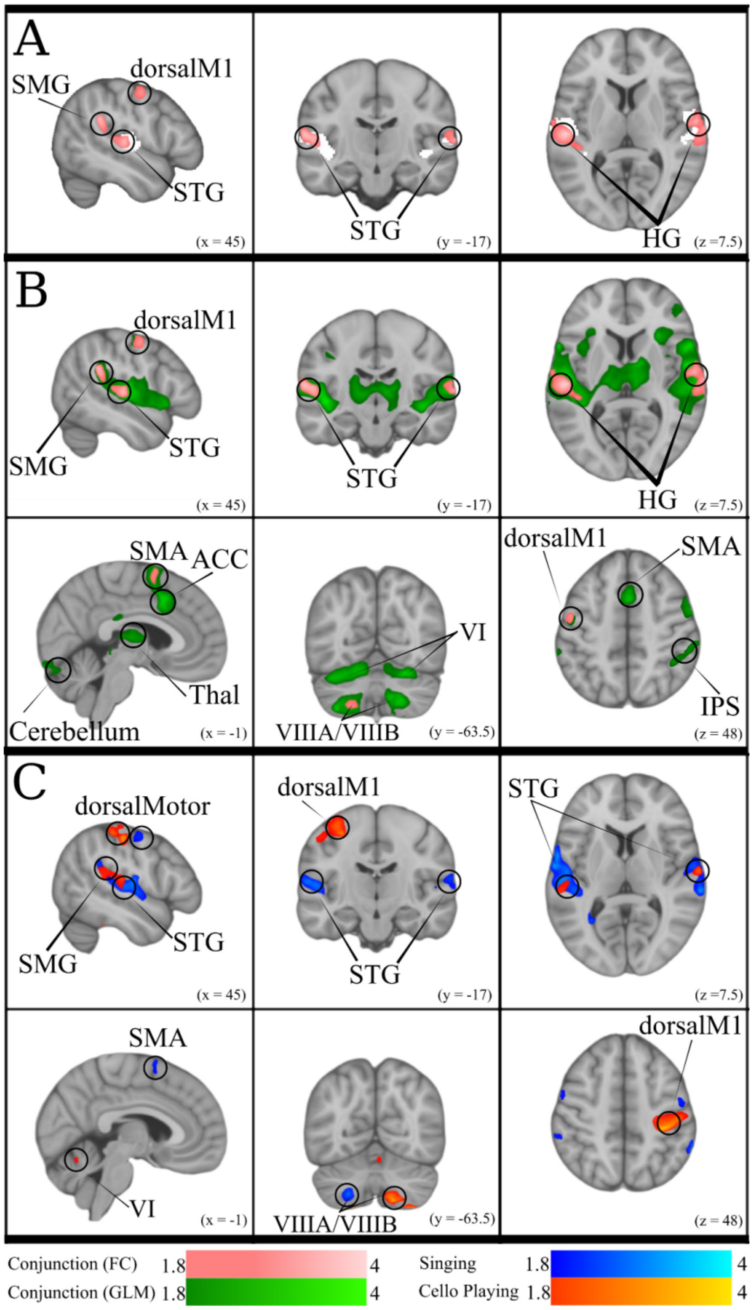

Figure 2-10 Functional Connected Network for Singing and Cello Playing (A) Conjunction of functional connectivity during Singing and Cello Playing (pink) overlayed on bilateral Heschle's Gyrus seed region (white). (B) Conjunction of Functional Connectivity during Singing and Cello Playing (pink) overlayed on Conjunction of simple singing and cello playing (green). **(C)** Functional connectivity for Cello Playing > Singing (orange) and Singing > Cello Playing (blue).

2.5.3 Questions Not Answered

STG Recruitment

Even when directly comparing singing and cello playing within the same individuals, some questions are still unanswered. For instance, we found increased activation in cello playing relative to singing in STG. There are two main ways that increased activation may be interpreted when activity falls within regions that are active for both tasks. The first is that increased activation is a sign of enhanced processing. The second is that it is a sign of decreased cognitive efficiency. We propose that the years of explicit training required to play the cello results in enhanced processing during cello tasks relative to singing. This interpretation is consistent with previous work in singing, which has shown that expert singers recruit STG to a greater extent than do novices (Zarate and Zatorre 2008) and previous work in trumpet players showing preferential recruitment of STG during trumpet playing (Gebel et al. 2013). Following from this, we would predict that if expert singers were compared to expert cellists there would be less of a difference in STG activity levels. A third possible interpretation is that the difference is the result of a confounding factor like intensity or another physical feature of the sound.

Single Notes vs. Melodies

This experiment used a listen/play paradigm with single tones as opposed to melodies, which may limit how well these findings generalize to musical performance in a more naturalistic context. In this regard it may be important to consider our findings in the context of continuous feedback hit-track paradigms. For instance, work done on goal directed movements has characterized an open loop, goal directed and closed loop, feedback oriented system for motor control (for review: Gaveau et al. 2014) and brain activity measured using single note reproduction tasks like the one used in this study may be biased toward the open loop component. However, we do not believe that playing a single tone would bear no similarity at all to playing a pattern of tones. Our findings did show activation in line with that of both singing melodies (Kleber et al. 2007) and playing melodies on musical instruments(Lotze et al. 2003). Most importantly, we recently carried out an fMRI study on learning to produce a four-note sequence on the cello which shows very similar auditory-motor activation in dorsal motor and pre-motor areas to those observed here (Wollman et al. 2018). We therefore conclude that the neural systems are similar for production of a single tone as they are for production of a short sequence.

Discrete vs. Continuous Pitch Instruments

One of the premises we out forth in our introduction is that cello playing is uniquely similar to the human voice. However, one could argue that the linear arrangement of keys and, consequently, pitches along the length of a keyboard is more related to the monotonic arrangement of pitch along the human vocal cord than the many-to-one mapping of pitches on the cello. We would argue that the differences between singing and keyboard playing are much more compelling, given that motor control over the larynx entails muscular contraction of the vocal cords to different degrees, coordinated with breathing, whereas to play a keyboard requires coordinated action of muscles, joints, limbs, and possibly body posture. Furthermore, with respect to the point about a monotonic mapping, it is possible to play any of the 88 notes on a piano with any of the ten fingers of the two hands. Therefore, there is no one-to-one mapping between motor action and sounded pitch; rather, there is always more than one fingering combination to produce the same pitch. We acknowledge that when all strings and all hand positions are used to play the cello it also creates a many-to-one mapping of action to pitch. However, in our study we specifically limited the task to the use of the index finger on one string to maximize the similarity between our cello and singing tasks. By imposing these limitations (string, finger movement) we control for both the many-to-one mapping of location to pitch, and of action to pitch present in everyday cello playing.

Our research cannot directly address the question of how these findings would generalize to discrete pitch instruments like the keyboard or guitar. These instruments do not allow for the online pitch adjustments that are integral to singing or playing continuous pitch instruments in tune. Therefore, the most important differences are likely to emerge in paradigms that exploit this aspect of on-line correction. However, based on the fact that singing and cello playing show such a high degree of overlap in recruited brain areas despite being such physically different tasks, we would speculate that discrete pitch instruments would show a high degree of overlap as well when no online correction is required. This prediction would also be consistent with the large body of research showing that many of the same brain areas are recruited during both piano playing

(Parsons et al. 2005) and guitar playing (Vogt et al. 2004; Vogt, Higuchi, and Roberts 2007) (for review: Zatorre, Chen, and Penhune 2007).

2.5.4 Future Directions

Using auditory feedback to meaningfully alter movements is one of the core features of auditory-motor integration. In the present study we did not specifically test how auditory feedback affected motor output. As such it is possible, though unlikely, that our participants were relying exclusively on the feed-forward component of the auditory-motor integration network and that auditory feedback was not being used to inform their movements. One of the classic ways of studying the neural correlates of auditory-motor integration is to use pitch perturbation paradigms, where participants are specifically instructed to compensate for introduced perturbations in auditory feedback (Burnett et al. 1998; Zarate, Wood, and Zatorre 2010). In so doing, researchers can directly observe which brain regions are involved in integrating auditory feedback with motor planning and execution. Using such paradigms in future experiments will allow us to observe how this auditory-motor integration occurs in cello playing and once again compare these findings with singing.

Another axis along which singing and cello playing might differ, even if the same brain areas are recruited, is the timing of the different processing steps. For instance, it could be the case that the auditory-motor integration network processes and responds to pitch perturbations more quickly during vocal tasks than cello playing due to the evolutionary significance of the voice, and/or due to connectivity differences between auditory and motor systems involved. Directly comparing the latency of event related potentials during both singing and cello playing would allow us to address this question. In doing so, we would gain a more complete understanding of how new skills make use of existing mechanisms in the brain for accomplishing similar tasks.

Chapter 3 - Pitch regulation in cello playing relies on singing-network brain structures

3.1 **Preface**

In Experiment 1 we showed that cello playing recruits areas within the singing network for a simple single note performance task, thereby supporting our hypothesis that the two tasks make use of a shared network for auditory-motor integration. In Experiment 2 we asked participants to compensate for or to ignore an introduced pitch feedback perturbation to specifically test the hypothesis that cello playing also makes use of structures within the singing network for linking perception to action. Experiment 2 has been prepared as a manuscript for publication, and has been formatted for inclusion in the thesis.

3.2 Abstract

Singing and cello playing have previously been shown to rely on overlapping brain structures within the Singing Network for playing individual target notes. These brain regions include those for auditory feedback monitoring (STG), motor planning (SMA, ACC), and auditory-motor integration (IPS). Activity in areas related to motor execution (M1, PAG), conversely, was found to be largely distinct, suggesting that differences between singing/playing are primarily related to the use of the relevant motor effectors and brain structures used for higher level sound to movement transformations are shared. However, it is unclear the extent to which singing/playing individual tones requires auditory-motor integration. Research has shown that singers and musicians alike can continue singing/playing in tune even in the absence of auditory feedback (Kleber et al. 2013), especially for short periods of time (J. Chen et al. 2013). Pitch feedback perturbation paradigms have been used throughout the auditory-motor integration literature to address this possible confound by using the changes in auditory feedback to drive a compensatory motor response. Studies specifically looking at pitch regulation in singing have shown that both singers and non-singers can compensate for incorrect pitch feedback, and that the ACC, aINS, and IPS are recruited for accomplishing this task. Critically, these studies also found that only expert singers could ignore incorrect feedback. The pSTG and SMG were found to be key to this experience dependent behavioural effect. Here we replicated our previous finding that cellists can play/sing individual target tones equally in tune when no perturbation is introduced and extend this finding by showing that cellists can also compensate for and ignore introduced pitch feedback perturbations equally well on the cello and when singing. We also replicated key parts of the overlapping brain network for singing/playing in auditory (STG), motor (SMA, ACC), and auditory-motor integration regions (dPMC, SMG, IPS). When participants compensated, we found overlapping activation in IPS and SMG compared to rest. When ignoring, participants showed overlapping activation only in STG and dPMC. Differences between singing/playing across all three conditions were most prominent in M1, centered around the relevant motor effectors (hand, larynx). These findings support the hypothesis that pitch regulation during cello playing relies on structures within the singing network and suggests that differences arise only at the level of forward motor control.

3.3 **Introduction**

Playing a musical instrument and singing are complex auditory-motor tasks that require the production of accurate pitch patterns over time; as such they depend both on fine motor control and on auditory feedback processing. These two domains have been shown to rely on many of the same brain circuits despite widely varying motor demands, but questions remain regarding how auditory-motor integration takes place when sounds are produced via the vocal tract, or by manipulating an external object such as an instrument.

Research on vocalizations, including singing, has characterized a network of feedforward and feedback brain structures that act in concert to plan, execute, and adjust generated musical pitch patterns (Kleber and Zarate 2014). The feedforward network includes primary motor, dorsal pre-motor, supplementary motor cortices (M1, dPMC, SMA), the basal ganglia (BG) and cerebellum for motor planning and executing, and the feedback network encompasses primary and periprimary auditory cortices within the superior temporal gyrus (STG) including Heschl's gyrus, the planum temporale, and planum polare (HG, PT, PP) for monitoring auditory feedback. At the junction of the two are the anterior insula (aINS), anterior cingulate cortex (ACC), and intraparietal sulcus (IPS) for comparing feedback to the expected output.

Critically, these components all interact with one another to enable sensory-motor integration; that is, the ability to monitor the sensory feedback that is the outcome of a set of actions in order to adjust the motor response to achieve a given sensory goal (Kleber and Zarate 2014). Many of these same structures are recruited during musical instrument performance (R. M. Brown, Zatorre, and Penhune 2015; Zatorre, Chen, and Penhune 2007), and in an fMRI study were found to be directly overlapping during singing and cello playing (Segado et al. 2018), raising the possibility that musical instrument playing may exploit the same sensory-motor system as vocalization.

In that previous study we directly compared, within the same individuals, which brain structures were recruited for both singing and cello playing during a simple single note performance task that followed a listen/play trial structure. We found that singing and cello playing recruit several common structures. Specifically, we found these areas include dorsal motor, premotor, and supplementary motor areas, auditory areas within STG as well as the ACC, aINS, and IPS. These findings allow us to conclude that cello playing makes use of the auditory-vocal network. However, because in that study no feedback manipulation was introduced, we were not able to ascertain whether these structures were directly contributing to auditory-motor integration. In the absence of a feedback perturbation one could argue that the findings only reflect the feed-forward component of the auditory-motor integration model since it is impossible to show that the auditory feedback is being used to influence the subsequent movements. This is especially important given that somatosensory, kinesthetic, and vibrotactile cues allow people to sing and play musical instruments (relatively) in tune even when auditory feedback is attenuated (Kleber et al. 2017; J. Chen et al. 2008; Zarate and Zatorre 2008, 2005).

Research done on poor pitch singers and amusics has shown that such people can still rely on sensorimotor associations to correctly mimic songs (Pfordresher and Mantell 2014) or sung sentences (S. Hutchins and Peretz 2012) even if they cannot produce them independently or perceive that they are incorrect. Studies on mechanical perturbations of the orofacial articulators have found that participants can compensate for perturbations based on somatosensory feedback alone (Tremblay, Shiller, and Ostry 2003); and anesthetizing the vocal chords during singing, thereby attenuating sensorimotor but not auditory feedback, has been shown to negatively impact pitch accuracy in non-singers, but not in trained singers (Kleber et al. 2013). Kinesthetic feedback (Askenfelt and Jansson 1992) and vibrotactile feedback (Goebl and Palmer 2008) have also been shown to be important for accurate musical instrument performance. Nevertheless, it also has been clearly demonstrated that monitoring of auditory feedback is essential for singing and playing in tune. For instance, research on singing has shown that pitch is less accurate and more variable in the absence of auditory feedback except in the case of expert singers who are just as accurate without feedback (Kleber et al. 2017) In cello playing, performance is accurate at first, but gets less accurate the longer performance continues in the absence of auditory feedback (J. Chen et al. 2013). As such we will focus on the role of auditory feedback in the present investigation.

Auditory feedback perturbation paradigms are used throughout the auditory-motor integration literature as a way to ensure that auditory feedback is being used to modify future movements. Research done on speech has shown that participants will compensate for introduced auditory feedback perturbations of F0 as well as in formants (Tourville, Reilly, and Guenther

2008). In singing, research has shown that participants compensate for introduced pitch perturbations both when holding a constant pitch (Burnett et al. 1998) and when executing dynamic pitch changes (Burnett and Larson 2002). In fact, there is a wealth of evidence to suggest that compensating is the default response to perturbed auditory feedback. Burnett (1998) characterized an early and a late phase of vocal compensation, the former being fully involuntary.

Further evidence comes from work done in singing where non-singers were able to compensate for, but not ignore, introduced pitch perturbations of 200 cents (Zarate and Zatorre 2008), and that even expert singers were unable to fully ignore perturbations of 25 cents (Zarate, Wood, and Zatorre 2010). In addition to recruiting brain regions that govern motor planning and execution (M1, pre-SMA/SMA, dPMC), compensating for pitch perturbations was found encompass brain areas that process sensory feedback (STG) and compare it to the expected output (IPS, ACC, aINS). Subcortical structures were also recruited for this task, notably the putamen was recruited to a greater extent when compensating than just singing back a target note. A network involving IPS, SMG, STG/STS, Insula and vPMC was found to be specifically relevant for successfully ignoring pitch shifts of 200 cents (Zarate and Zatorre 2008), while the rCZa and pSTS were found to be related to the involuntary pitch shift reflex (Zarate, Wood, and Zatorre 2010).

Musical instrument research has also looked at the neural correlates of incorrect pitch feedback. A study on pitch and timing feedback manipulations during keyboard playing found that participants recruit the SMA, ACC, and the Cerebellum when pitch feedback was perturbed compared to simply playing a simple melody (Pfordresher et al. 2014). While this study did not allow for the online pitch corrections that occur during continuous pitch instrument performance, this finding still demonstrates that these brain structures are linked to the sound to movement transformations necessary for using incorrect pitch feedback to plan future movements.

In addition, several studies have found that when participants with no musical training learn to play simple sequences on the keyboard, they quickly form a link between sounds and movements, which is reflected in coordinated activation of motor and sensory brain regions even when only sensory input is provided (J. L. Chen, Zatorre, and Penhune 2006; Stephan, Lega, and Penhune 2018; Lahav et al. 2005; Herholz and Zatorre 2012). While keyboard studies cannot directly address the role of these structures in pitch regulation since the keyboard is a discrete-pitch instrument, the same finding has recently been replicated on the cello, which like the voice has a continuous pitch distribution. When complete novices learned to play a set of simple cello sequences they recruited the auditory to dorsal-cortical pathway after only 1 week of cello learning (Wollman et al. 2018). Moreover, the link between auditory and SMA/pre-SMA after one week of training was found to be predictive of better learning outcomes after 4 weeks of training. The latter finding in particular speaks to the functional importance of sensory-motor links in the production of accurate pitch sequences.

In the present study we directly test the hypothesis that singing and cello playing recruit overlapping brain regions in the same individuals during auditory-motor integration, which we elicit using a compensate/ignore perturbed pitch paradigm. Our aim is not only to use an independent sample to replicate our previous finding that cello playing recruits structures within the auditory-vocal network during simple single-note reproduction, but also to extend these findings by including the critical new conditions in which a pitch perturbation is introduced, and listeners are asked either to compensate or ignore the perturbation, as has been used in singing studies (Zarate and Zatorre 2008; Zarate, Wood, and Zatorre 2010). To accomplish our goal, we used an MR compatible cello device where sound feedback is delivered directly to the player during scanning (Hollinger and Wanderley 2013, 2015). By comparing the neural correlates of singing to those of cello playing directly in the same individuals, we can determine the extent to which musical instrument playing makes use of functional brain networks that have evolved for singing.

We expect that cello playing and singing will recruit largely overlapping areas in premotor and supplementary motor areas, auditory cortices, and the cerebellum when compensating for and ignoring perturbations since both of these conditions involve the production of target tones. However, we expect that activation in primary motor cortex for cello and singing will be specific to the hand and larynx areas, respectively as was the case in Experiment 2 (Segado et al. 2018). When compensating for the introduced perturbations, we expect that playing/singing will both preferentially recruit the IPS, ACC, and aINS as well as pre-SMA/SMA relative to the simple performance condition since these regions were most strongly implicated in auditory-vocal integration by singing research and compensating accurately specifically requires an accurate auditory to motor transformation. When ignoring, we expect to see increased recruitment of STG/STS, aINS relative to simple performance since successfully accomplishing this task requires increased reliance on the forward model and may require increased auditory feedback monitoring to suppress a tendency towards compensation. Regarding the differences between singing and cello playing, we are most interested in the recruitment of the basal ganglia and the brainstem, as well as larynx area of M1, all of which were shown to be recruited by singing more than by cello playing in our previous study.

3.4 Materials and methods

3.4.1 Subjects

A total of 15 expert cellists (10 Female) were recruited from the Montreal community, reported normal hearing, did not report any neurological disorders, and had no contraindications for the MRI environment (mean years experience = 12.93 +- 4.46, mean starting age = 9, mean practice hours per week = 12.07). Thirteen participants were included in the final analysis; two were excluded from the analysis due to technical problems. This study was approved by and carried out in accordance with the recommendations of Montreal Neurological Institute Research Ethics Board and the McConnell Brain Imaging Centre. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

3.4.2 Experimental Paradigm

Stimuli and Task Conditions

Participants were asked to listen to and subsequently reproduce target tones both by singing and by playing them on the MR-Compatible cello. There were three experimental conditions: Sing/Play, Listen, and Rest. For the first two of these conditions, we presented auditory target tones that were 1 s long. Presented tones were (using Western music notation) E3 and F$^{\#}$3 for cello, and either E3 and F$^{\#}$3, or E4 and F$^{\#}$4 for singing depending on participant's vocal range. Target tones for the cello playing condition were recorded on a cello, while those for the vocal condition were recorded by either a female vocalist, male vocalist. Participants heard and imitated the pitch sung by someone of their own gender. During the Sing/Play conditions (but not listen or rest) between 1 and 1.5 seconds after the trial onset auditory feedback was shifted in pitch up or down by 100 cents (one semitone) on certain trials in an unpredictable manner (jittered up to 500ms), and participants were instructed in separate blocks of trials to either compensate for or to ignore the introduced pitch shift. Visual instructions to compensate or ignore were displayed on a black screen during the presentation of the auditory cue.

Participants underwent a behavioral familiarization session prior to scanning, followed some days later by an fMRI session. For both the familiarization session and the fMRI session, a microphone was suspended approximately 5 cm from the mouth, the MRI compatible cello device was placed along the torso using a specialized support, and bilateral sound delivery was provided via insert earphones (Sensimetrics, Dayton Audio DTA-1 amplifier). The microphone and cello were connected to a Zoom F8 portable field recorder, and to a midi controlled TCHelicon VoiceOne pitch shifter. The shifter was used to shift the pitch feedback on some trials, or mask the feedback on other trials. Pink noise was played continuously through the headphones to reduce bone conduction so that audio was being delivered exclusively through the earphones to the greatest extent possible. All sound levels were adjusted on a per subject basis, but on average pink noise was presented at 78.3 dB SPL A and auditory targets were presented at approximately 15.6 dB above the noise floor. The experiment was run using custom scripts written in python. All vocalizations and cello sounds were digitally recorded to the Zoom F8 digital recorder.

A sparse sampling paradigm was used for the fMRI session (Belin et al. 1999), where a long delay in TR was used to allow tasks to be carried out in the relatively silent period between functional volume acquisitions, thus minimizing acoustical interference and also avoiding movement-related artifacts since the scanning takes place after the motor production for each trial (Figure 3-2).

Procedure

To allow participants to adjust to the fMRI-compatible cello and the constraints of playing it in the scanner, each person underwent a 10 min familiarization session no more than 1 week prior to their session in the MR-scanner. During the familiarization session each participant was asked to lie inside a structure that simulated the space constraints of the MRI environment. All participants were asked to play both target notes, and to try warming up as they would on a regular instrument.

On performance trials, participants were instructed to sing or play back the target tone for 2.5s. For singing trials, participants were instructed to sing as quietly as possible with their jaw relaxed to reduce breathing artifacts in the recorded signal and movement artifacts in the fMRI signal. For cello playing trials they were instructed to use as few bows as possible to reduce movement artifacts (approximately two bows per trial). Between trials participants were instructed to keep their hands on the cello and to move as little as possible.

Within 1 week of the familiarization session, participants were tested in the Siemens Trio 3T magnetic resonance (MR) scanner at the Brain Imaging Center of the Montreal Neurological Institute. Each participant was fitted with MR-compatible headphones. The MR-compatible microphone was attached to the mirror support system and, as was previously mentioned, the MR-compatible cello was laid across the torso using a special MR-compatible stand. During the fMRI session, participants performed ten runs, each of which was composed of four blocks that alternated between the three conditions (Compensate, Ignore, Attenuated) for a total of approximately 60 trials per condition per instruments. Within each block, participants underwent one passive listening trial and six performance trials per instrument (2 no-shift, 2-shift up, 2-shift down), and one rest trial at the end for a total of fifteen trials. Trial order was pseudo randomized within each block, block order was pseudo randomized within run, and run order was pseudorandomized across participants. Timing of the target tone was jittered randomly by up to 500ms to increase the likelihood of obtaining the peak of the hemodynamic response for each task. Listening and attenuated audio trials were not included in the analyses presented in this manuscript as they did not relate to the central hypotheses and will be analysed separately at a later point in time.

MRI Acquisition

During the 10 functional runs, one whole-head frame of 28 contiguous T2*-weighted images were acquired (Slice order = Interleaved, TE = 85ms, TR = 6.7s, Delay in TR = 4.4s, 64 ×

64 matrix, voxel size = $4mm^3$). All tasks were performed during the 4.4s silent period between functional volume acquisitions. On each trial, the target tone was delivered for one second together with the visual instruction; the rest of the trial was taken up by reproduction of the cello or sung tone, or by no action on rest trials (See Figure 3-2). A high-resolution whole-brain T1-weighted anatomical scan (voxel size = $1mm^3$) was collected between runs 5 and 6.

3.4.3 Behavioral Analyses

Individual trials of singing and cello playing were analyzed with pitch information extracted from the recorded audio signals. Each trial's audio was first segmented from the continuous 6-track audio using Audacity software. The trials were then processed using a custom analysis pipeline implemented in Python, with a GUI for visualizing and optimizing analysis parameters (Figure 2-3 Behavioural Analysis Software).

The ambient noise in the scanner room had a peak resonance of 160 Hz that interfered with the extraction of fundamental target pitches, so harmonics 3–10 of the cello and singing tones were used for pitch extraction. To reject room noise and to isolate harmonics of interest, the raw microphone signal was high-pass filtered with a cutoff at 367 Hz and low-pass filtered with a cutoff at 4,216 Hz. Pitch estimation was then performed using the YinFFT algorithm provided in the Python module Aubio (Brossier 2007), producing a time-series of pitch estimations (detected harmonic, in Hz) and confidence ratings (between 0 and 1). Estimates were adjusted to their representative fundamental pitches before selecting stable pitch regions for further analysis. Stable pitch regions were defined as: segments of at least 150 ms in which the rate of change of the pitch did not exceed 100 Hz/s (or approximately 0.07 Hz per 32-sample pitch estimation window at the sampling rate of 44,100 Hz). Of these regions, only those that maintained a confidence rating of at least 0.7 by the algorithm were included. Trials were rejected if no regions were found to meet the stability and confidence criteria, if participants started playing after the pitch shift was introduced, or if they stopped playing before the end of the trial. In total, 86.5% of trials were retained. Rejected trials were excluded from the fMRI analysis.

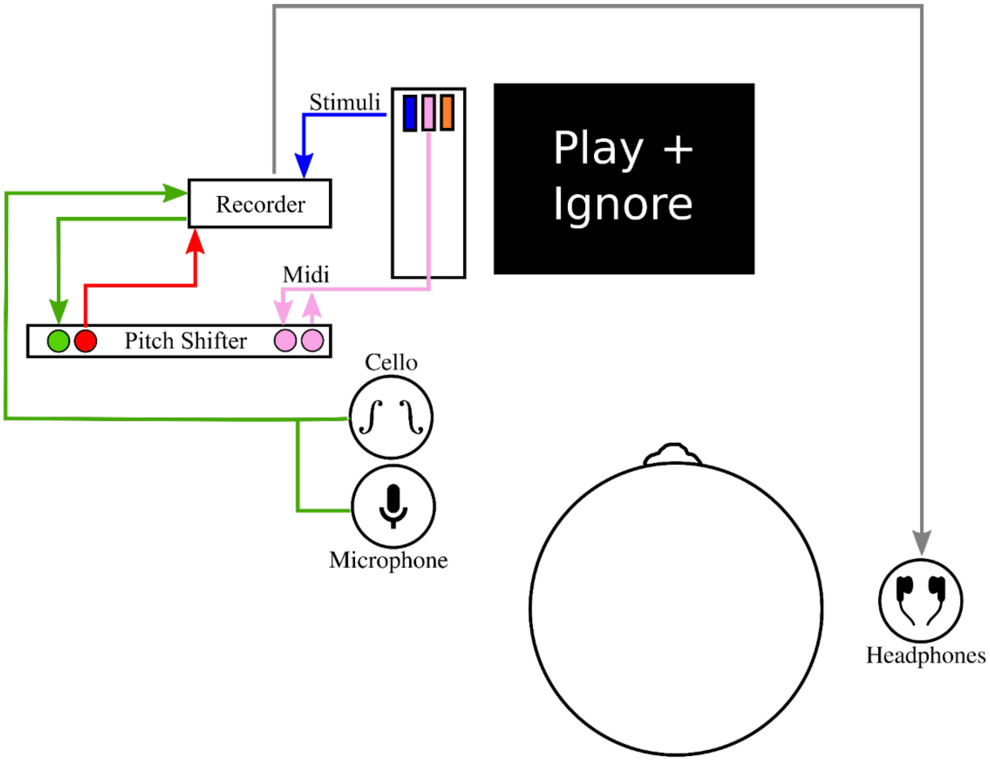

Figure 3-1 Experimental Setup | Experimental setup used to present stimuli, cello audio, and singing audio through headphones. Pitch shifter allowed for audio feedback from cello/singing to be shifted or attenuated on specific trials while still presenting audio stimuli and masking noise.

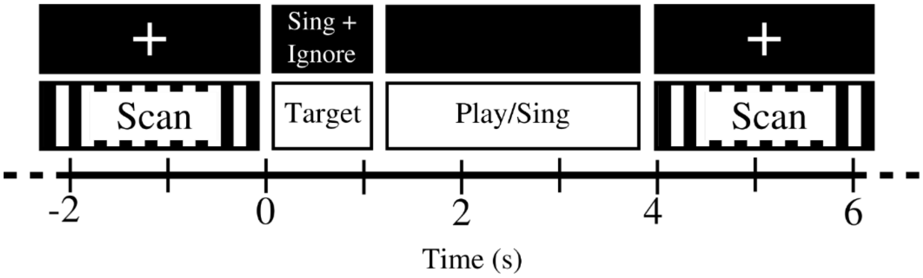

Figure 3-2 Sparse sampling design Sparse sampling design used to avoid auditory and motion artifacts. Auditory target presentation, instruction presentation, and singing and cello playing were done during the silent 4s period between scans.

Pitch extraction was performed for both the unshifted raw output track (produced) and for the shifted headphone track (perceived) to allow for a direct comparison between the produced and perceived audio. Performance scores were calculated on a per participant basis as follows:

The difference between the produced tone from the target tone (expressed in cents, or 100^{ths} of a semitone) was analyzed within three 150ms windows, Pre-Shift, Early Post-Shift, and Late Post-Shift for each of the Simple, Compensate, and Ignore conditions (Equation 3-1).

An overall accuracy score for each participant was then determined by calculating their mean pitch accuracy across all trials. Finally, scores were analyzed using a two way (instruction by timepoint) ANOVA implemented in GraphPad.

$$cents = 1200^{*} log_2\left(\frac{Freq_{produced}}{Freq_{target}}\right)$$

Equation 3-1 Conversion from absolute frequency difference to normalized cent difference

Simple tone reproduction

In these trials, participants were instructed to either ignore or compensate, but no pitch shift was introduced. In other words, the difference between perceived and produced audio should have remained at 0 throughout the trial. Trial accuracy was determined by averaging the cent deviation from the target tone at the pre- and late-post timepoints. This was done to account for any natural drift in pitch throughout the trial.

Compensate and Ignore

In these trials participants were instructed to either compensate for or to ignore a 100 cent (quartertone) pitch shift introduced between 1 and 1.5 seconds after the trial onset. Shift onset was jittered to prevent participants anticipating the shift. On compensate trials, participants were instructed to compensate for the introduced perturbation in order to go back to hearing themselves play the original target pitch. In other words, in response to a 100 cent downwards shift, participants would need to increase their produced pitch by 100 cents, and vice-versa for an upward shift. On ignore conditions they were asked to ignore the perturbation and continue playing or singing the pitch they played initially.

Each participant was given a score for the simple, ignore, and compensate conditions based on how well they were able to accomplish each task. All scores we calculated based on the cent difference between the produced output and the target tone at the early and late timepoints. For the compensate condition, perfect performance was considered to be 100 cents in the direction opposite to the introduced pitch perturbation, whereas for the ignore condition perfect performance was considered to be 0 cents deviation from the target tone. All three scores were averaged to give a global participant score, which was used in the fMRI analysis to look at BOLD activity as a function of task performance.

3.4.4 fMRI Analyses

fMRI data were analysed using the FSL6.0 FEAT toolbox (Jenkinson et al. 2012). Brain extraction was carried out using BET2. Functional volumes were aligned to the high resolution anatomical and then to MNI152 1mm standard space using FLIRT non linear registration with 12 degrees of freedom and a warp resolution of 10mm. To improve the registration quality, b0 unwarping was carried out using a percent signal loss threshold of 50%. Motion parameters were estimated using MCFLIRT, and fMRI timecourse was temporally filtered to remove drifts greater than 160 ms. To further decrease the number of motion-related active voxels outside the brain, a brain mask based on the MNI152 1mm brain was applied prior to thresholding. To boost the signal to noise ratio, images were spatially smoothed with an 8 mm FWHM kernel. FLAME-1 mixed effects modeling was used to fit the GLM to the fMRI signal.

Statistical significance was determined using an FSL cluster probability threshold of $p < 0.05$ with a voxel-wise significance level of $z = 2.3$ ($p < 0.05$). The cluster probability threshold serves as a correction for multiple comparisons. A total of 21 contrasts were carried out. Contrasts were carried out for each of the instruction conditions (simple vs rest, compensate vs simple, and ignore vs simple) within each of the instruments (singing, cello playing). Additional contrasts were carried out to directly compare the two instruments within each of the instruction conditions (e.g. singing compensate vs cello compensate), and between instruction conditions ([singing compensate vs ignore] vs [cello compensate vs ignore]). Additionally, task performance (on a per-subject basis) was regressed against the BOLD signal for each of these contrasts to determine whether or not activity in these areas is positively correlated with good performance on the task, as would be expected if the region in question is specifically linked to the cognitive/motor demands of the task of interest. Statistical conjunctions were carried out to identify commonalities in singing vs cello playing, using the conjunction script created by the Warwick University Department of Statistics, which also made use of the FSL tools (Nichols et al. 2005). This script carries out a voxel-wise thresholding of $p < 0.05$ in both conditions of interest, and then carries out a cluster correction of $p < 0.05$.

Functional Connectivity analyses (underway) were carried out using the FSL6.0 FEAT toolbox. Six anatomically defined seed regions of interest were chosen due to their involvement in simple singing and cello playing: DPMC, M1 (bilateral hands, larynx), IPS, SMA, and HG. A seed region in auditory cortex was identified by masking the conjunction of singing and cello playing from the functional data with an anatomically defined mask of Heschl's Gyrus (Harvard Structural Brain Atlas). Seed regions in primary motor cortex (M1) were identified by masking singing and cello playing with an anatomically defined mask of post-central gyrus (Harvard Structural Brain Atlas). An additional seed region was identified using an ROI of larynx area based on coordinates for the larynx phonation area in Brown 2008 (S. Brown, Ngan, and Liotti 2008). The activation timecourse in each of these regions was extracted and correlated with the whole-brain timecourse for each task of interest, which was estimated using the GLM. Correlated voxels were thresholded as described above. Regions that showed a correlated timecourse were then linearly regressed with task performance to determine whether those areas were contributing directly to good intonation.

3.5 **Results**

3.5.1 Behavioural Findings

We first carried out a behavioral analysis to confirm that our participants, who were all cello players rather than professional singers, could sing target tones accurately, and also play them on the cello with an average tone accuracy of at least a quartertone (50 cents). The behavioural analysis further allowed us to determine the extent to which participants could compensate for introduced pitch perturbations and also the extent to which they could ignore them. Because there were no significant performance differences between singing and cello playing, the two instruments were analysed independently in order to decrease the degrees of freedom.

Accuracy during Simple condition

We first replicated our previously published finding that cellists could both sing ($\text{mean}_{\text{sing}}$ = 0.86 ± 0.27 cents) and play target tones on the cello ($\text{mean}_{\text{cello}}$ = -3.53 ± 0.33 cents) with a high degree of accuracy when no pitch shift was introduced. In this simple condition, participants stayed within 15 cents deviation from the target tone at all three timepoints both when playing the cello and singing. Additionally, there was no significant difference between the early and late timepoints (p_{sing} = 0.76, p_{cello} = 0.88), which shows that their produced pitch was stable. We also found no significant difference between the pitch accuracy in the simple condition compared to the pre-shift timepoint in the compensate and ignore conditions for both singing (p_{comp} = 1.00, p_{ign} = 0.82) and cello playing (p_{comp} = 0.95, p_{ign} = 0.98), showing that participants were able to accurately reproduce the target tone before the pitch perturbation was introduced.

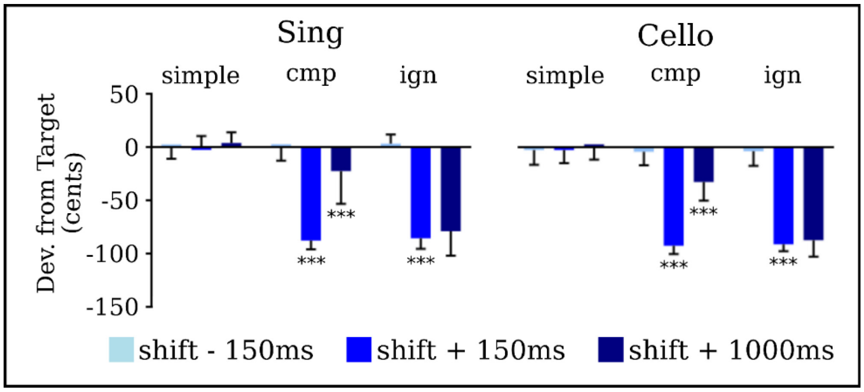

Figure 3-3 Deviation from target tone during simple, compensate, and ignore conditions | Deviation from target tones at 150ms prior to the shift (light blue), 150ms post shift (blue), and 1s post shift (dark blue) as heard through the participant's headphones. For the Simple condition where no shift was introduced the 'shift' was considered to be at the midpoint of the trial. For Compensate and Ignore conditions, perturbation was 100 cents.

Accuracy during Compensate and Ignore conditions

Our next hypothesis was about participants' ability to compensate for a sudden onset 100 cent pitch perturbation. Because we had no specific hypotheses related to the directionality of these perturbations, upwards and downwards perturbations were treated as equivalent, and values from the downwards perturbation were therefore multiplied by -1. Cent deviation from the pre-shift baseline (shift – 150 ms) was considered at 2 different timepoints. Time point 1 (post-early) was selected to capture any immediate response to the perturbation (shift + 150 ms), and timepoint 2 (post-late) was selected to capture the completed compensatory response (shift + 1000ms). The result was analysed using a two way repeated measures ANOVA including two factors: instruction (compensate, ignore) and timepoint (pre, post-early, post-late). Cello playing and singing were considered separately.

In singing, we found a significant effect of instruction (F = 125.4(2), $p < 0.0001$) and timepoint (F= 138.2(2), $p < 0.0001$), with a highly significant instruction by timepoint interaction (F = 51.79(4), $p < 0.0001$). Similarly, in cello playing, we found a significant effect of instruction (F = 103.1(2), $p < 0.0001$) and timepoint (F =443.7(2), $p < 0.0001$), with a highly significant instruction by timepoint interaction (F =149.9(4), $p < 0.0001$). Pairwise contrasts were carried out using Tukey's multiple comparisons tests. For both compensate and ignore instructions in both cello playing and singing, participants showed a highly significant difference between the pre-shift baseline and the 150ms following the introduced perturbation ($p_{(\text{post-early})} < 0.0001$). These post-early responses did not differ from one another in singing ($p = 0.93$) nor in cello playing ($p = 0.88$). In other words, in the 150ms following the introduced pitch shift, participants were affected by the pitch change as much in the compensate condition as they did in the ignore condition. However, when instructed to compensate, participants showed a highly significant ($p < 0.0001$) compensatory change from the post-early to post-late timepoints ($\text{mean}_{\text{cello}}$ = -59.92 cents, $\text{mean}_{\text{sing}}$ = -65.32 cents) which did not occur when they were instructed to ignore. When instructed to ignore, participants showed no significant change ($p_{\text{cello}} = 0.70$, $p_{\text{sing}} = 0.52$) from post-early to post-late timepoints ($\text{mean}_{\text{cello}}$ = -3.77 cents, $\text{mean}_{\text{sing}}$ =-6.70 cents).

3.5.2 fMRI findings

Simple tone reproduction

Overall, the fMRI results replicated many of our previous findings, as well as many of the findings from the vocal pitch literature. Cello playing and singing both recruited dorsal motor and pre-motor areas bilaterally, as well as the SMA and STG (also bilaterally). Singing compared to Rest specifically activated larynx area of motor cortex, whereas Cello Playing compared to Rest recruited more dorsal regions related to hand and arm movements. These findings support the hypothesis that activity in these regions is contributing to good task performance as opposed to them being coincidently activated by nonspecific task demands.

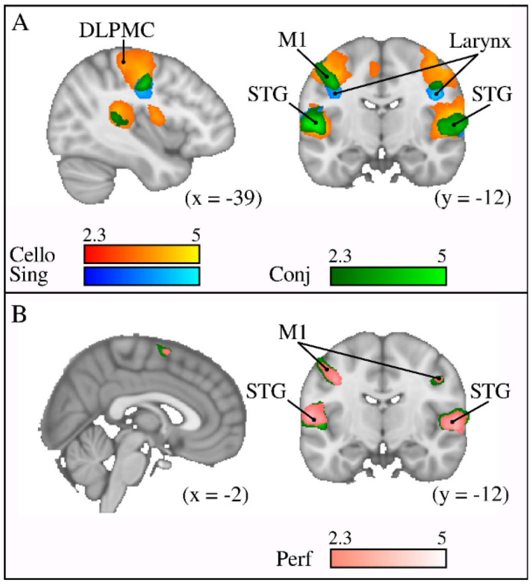

Figure 3-4 Simple performance condition A) Simple cello playing (orange), singing (blue) and conjunction (green). B) Conjunction (green) and areas that were positively correlated with good task performance (pink). Singing and cello playing were found to have overlapping activity in pre- post-central gyrus, STG and SMA. This activity was found to be positively correlated with good performance on the task.

Compensating: similarities

Only trials where participants successfully compensated for the perturbation were included in this analysis. When singing and when playing the cello, as compared to the simple no shift condition, participants recruited a number of dorsal-stream structures: DPMC, pre-central gyrus, IPS, SMG extending to posterior STG, and SMA extending to pre-SMA (all bilateral). These regions all correspond to those of the auditory-vocal integration network reported previously. A statistical conjunction of singing and cello playing activity showed that many of the areas recruited for compensating were directly overlapping (Figure 3-5A). Moreover, these regions of overlap were positively correlated with good performance on the task (Figure 3-5B).

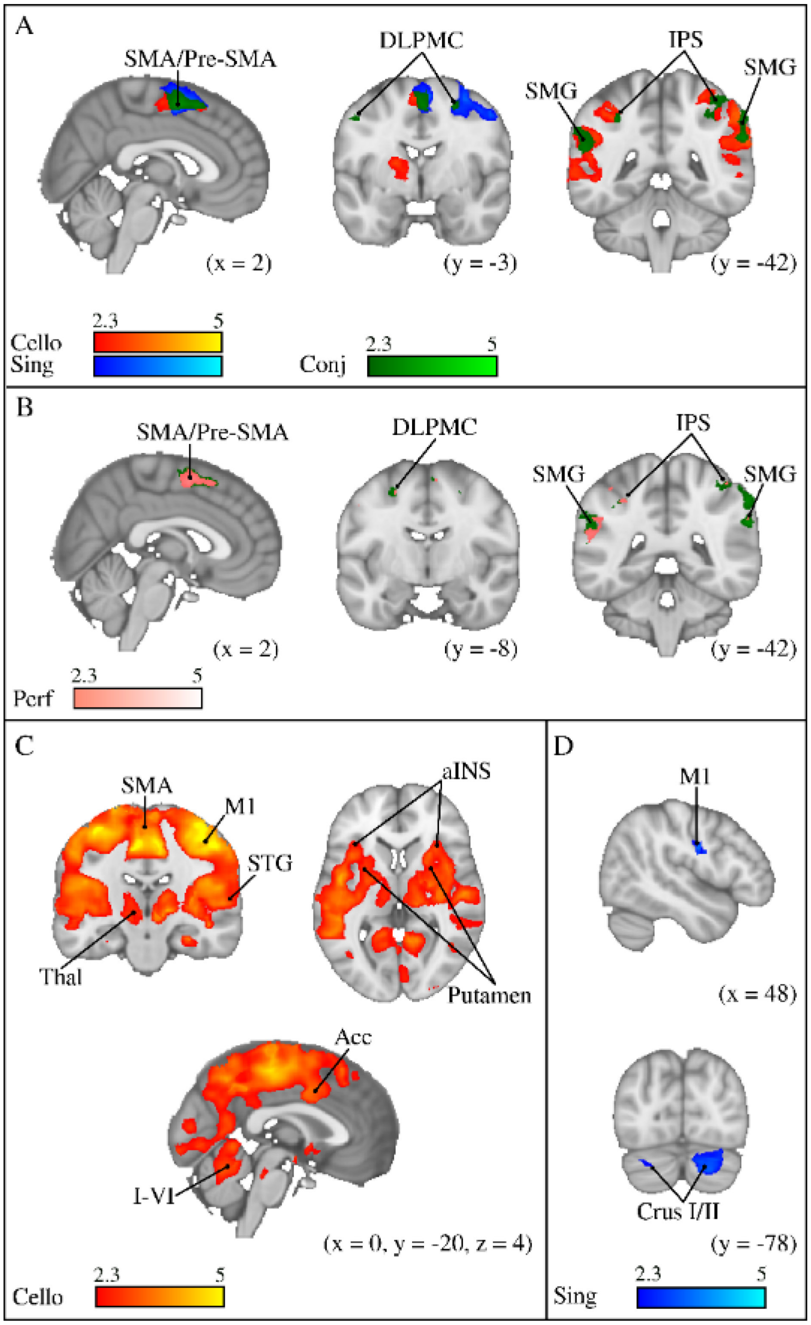

Figure 3-5 Compensating for pitch perturbations | A) Cello playing (orange), singing (blue), and Conjunction (green) of Compensate vs Simple contrast B) Areas within this conjunction (green) where activity is positively correlated with good task performance (pink) C) Cello playing > singing and D) singing > cello playing. Conjunction and regression show that SMA/pre-SMA, DPMC, SMG, and IPS are all contributing to good task performance in singing and in cello playing. Cello playing recruits more activity throughout the auditory motor integration network including in the BG and aINS. Singing shows more activity in vocal areas of motor cortex and cerebellum.

Compensating: Differences

We also observed some differences between singing and cello playing in the contrasts of compensate to simple performance. Cello playing showed more, and more extensive activation in bilateral dorsal motor and pre-motor regions extending posteriorly to the superior parietal lobule. This was also the case in the SMA and parietal operculum. More activity was also seen in the pre-cuneus. In the cerebellum cello playing preferentially recruited bilateral Crus V (Figure 3-5C). Singing preferentially recruited vocal areas of motor cortex as compared to cello playing. In the cerebellum singing preferentially recruited Crus I extending to Crus II (Figure 3-5D).

Ignoring: similarities

In the ignore condition, singing and cello playing both recruited dorsal motor cortex and STG when contrasted with Rest (Figure 3-6A). The statistical conjunction of the two conditions (cello ignore vs rest and singing ignore vs rest) showed overlapping activation in planum temporale and in dorsal pre-central gyrus. Within this conjunction, the PT showed a positive correlation with good task performance (Figure 3-6B). Overall, many fewer similarities were observed in the ignore condition than for the compensate condition.

Ignoring: differences

As in the Compensate condition, cello playing elicited more and more widespread activation relative to singing throughout bilateral dorsal motor and pre-motor regions extending posteriorly to the superior parietal lobule. This was also the case in the SMA, the parietal operculum, and the pre-cuneus. In the cerebellum cello playing preferentially recruited bilateral Crus V and bilateral VIIIa and VIIIb (Figure 3-6C). Singing showed more activation in ventral M1 (larynx area) and in Crus I/II of the cerebellum, as was the case in the Compensate condition, and also showed increased activation of the hippocampus (bilaterally) (Figure 3-6D).

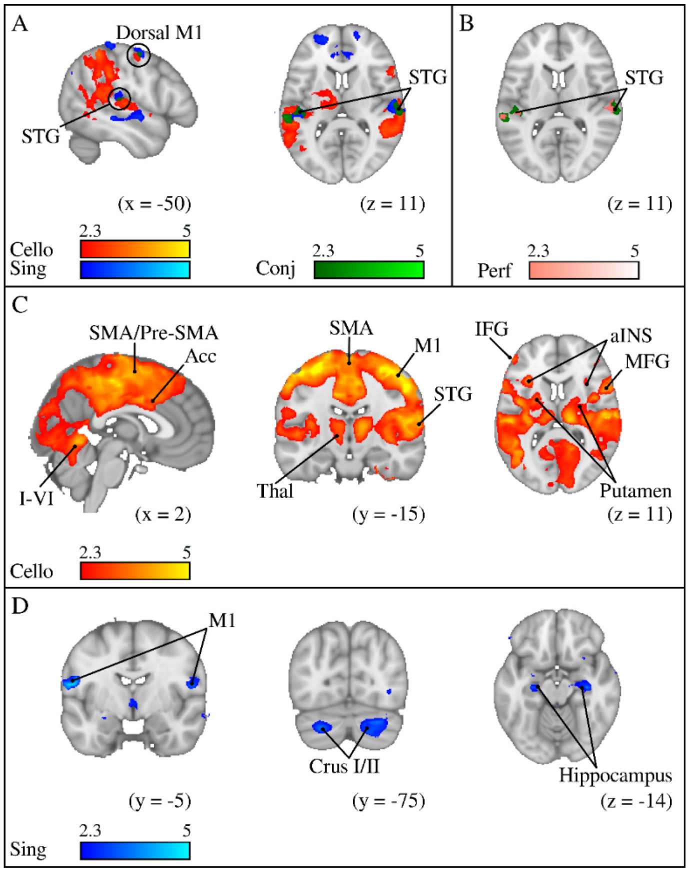

Figure 3-6 Ignoring incorrect pitch feedback A) Cello playing (orange), singing (blue), and Conjunction (green) of Ignore vs Simple contrast B) Areas within this conjunction (green) where activity is positively correlated with good task performance (pink) C) Cello playing > singing and D) singing > cello playing. Conjunction and regression show that only posterior STG is contributing to good task performance in both singing and in cello playing. Cello playing recruits more activity throughout the auditory motor integration network including in the BG and aINS, MFG, and IFG. Singing shows more activity in vocal areas of motor cortex and cerebellum, and in the hippocampus.

3.6 **Discussion**

In our previous paper (Segado et al. 2018) we found that the brain areas for singing and cello playing directly overlap in key areas related to auditory perception (HG, STG), auditory-motor integration (IPS, ACC, aINS), and movement planning (dPMC, SMA/pre-SMA) and execution (M1). However, we could not determine whether those regions were specifically contributing to auditory-motor integration, as opposed to contributing separately to sensory and motor processes, because no feedback manipulation was introduced. Here we not only replicate some of the key findings from our previous work, but also extended those findings by showing that brain regions recruited for singing and cello playing directly overlap in motor planning areas (SMA/pre-SMA, dPMC) and auditory-motor integration areas (IPS, SMG) when compensating for perturbed pitch feedback, and in auditory areas (STG) when ignoring it. This new set of findings reinforces our central assertion that structures within the auditory-vocal network are being used to regulate pitch in continuous pitch instrument performance as well as in singing. While we also observed some differences between the two instruments within each experimental condition (compensate, ignore), they were mostly expected differences in primary motor cortex related to the relevant motor effectors.

There are three ways of interpreting the broader significance of these findings. One is that cello playing scaffolds off of existing vocal infrastructure for pitch regulation, but that there is no direct relationship between the use of vocal infrastructure and vocalizations per se. In other words, cello playing is in no way directly related to singing but uses domain-general infrastructure that developed for singing. Two, that cello playing makes use of vocal regions in a more direct manner because there is an inherent link between cello playing and vocalization. For instance, it is possible that subvocalization during musical instrument performance is the default action. Three, that some components of the vocal network aren't specific to vocalizations after all, and that both singing and cello playing are making use of domain-general infrastructure that evolved for an even more global computational demand such as sensory-motor transformations. The findings of this paper continue to support the Neuronal Recycling interpretation that newer cultural tasks like playing musical instruments are simply reusing the systems that developed for phylogenetically older tasks (in this case singing) much like mathematical operations reuse areas that process direction (Dehaene and Cohen 2007).

3.6.1 What we found

Compensating and Ignoring in cello playing and singing: Behavioral findings

Our first finding was that our participants, who were all expert cellists but not professional singers, could accurately match target tones and maintain a stable pitch throughout the trial equally well when singing and when playing the cello (Figure 3-3). This finding, while expected, is important as it replicates our previous work (Figure 2-6) and the use of the MR compatible cello as a model instrument. The finding is also important in allowing us to interpret the neuroimaging findings, since performance on both instruments was comparable, meaning that the two tasks can be compared without confounds associated with behavioral differences.

In addition to finding that participants could match tones, we found that they were able to follow the instruction to ignore or to compensate accurately for unexpected sudden onset pitch shifts of 100 cents (Figure 3-3). For the compensate condition, participants were able to return to within the 50 cent 'correct' threshold of the target tone both in singing and cello playing. The 50 cent correct threshold was chosen based on previous work done in singing (Pfordresher and Brown 2017). For the ignore condition, the difference between what participants heard and what they produced stayed at 100 cents (+/- 10 cents) for the duration that the pitch shift was applied. We expected to see a compensatory response at the early timepoint especially in singing, but our analysis did not show evidence of this response. This could be because the pitch shift was quite large compared to the shifts that have been used to elicit a reflexive response in the literature (100 compared to 25 cents), and previous work has shown that large shifts are easier to ignore than small ones (Zarate, Wood, and Zatorre 2010). It could also be because the analysis window was quite large (150ms) and may have averaged across otherwise interesting initial responses to the pitch perturbation. This possibility can be addressed in future work using a more fine grained behavioural analysis approach.

Overlapping activation: simple target tones

The individual contrasts of singing and cello playing comparted to rest, together with the conjunction analysis (Figure 3-4) allowed us to independently replicate our previous findings regarding the similarities between singing and cello playing during simple single-tone reproduction (no perturbation). We again found overlapping recruitment of brain areas throughout dorsal motor regions and the STG in essentially identical regions. In addition, we found that this overlap includes the dorsal part of the larynx area of M1. Replication is important given the prominent critiques of neuroimaging that have emerged in recent years (Turner et al. 2018) especially those that object to relatively small sample sizes, as is necessarily the case here because we are studying highly specialized populations (cellists) where it would be hard to obtain large samples sizes. Given the similarity of results across two separate samples, at least at the cortical level, we are confident in asserting that cello playing depends on very similar cortical infrastructure as for auditory-vocal tasks like singing.

Overlapping activation: compensating for pitch perturbations

The contrast of compensating compared to simply playing the cello or singing without perturbation, together with the corresponding conjunction (Figure 3-5), shows that auditory-motor integration regions are not only coincidently active during cello playing and singing, but also that they are actively contributing to pitch regulation. Introducing an unpredictable pitch shift and having participants compensate for it is a common way to ensure that auditory feedback is being monitored and used in a meaningful way to guide subsequent movements. This manipulation is essential if the goal is to study the brain regions associated with this transformation from errant pitch feedback to compensatory action.

Our behavioural results demonstrate that this was, in fact, the case and hence validate the procedure used for both instruments. That being said, the nature of playing continuous pitch instruments like the cello is such that playing in tune requires continuous error feedback monitoring, because the instrument can go out of tune easily due to subtle changes in finger position, bow pressure, and ambient temperature; this factor is most likely why even without feedback manipulations our results show overlapping activation in what are traditionally believed to be auditory-motor integration areas like SMG. This effect was even more evident in our previous study, which showed recruitment of the IPS in the absence of a feedback manipulation (Segado et al. 2018). This is likely because participants were holding the note for longer in that study and, as such, had more opportunity to modify the ongoing note.

The role of the IPS & IPL

Our finding that the IPS and IPL (specifically the SMG) are recruited in both singing and cello playing when compensating for introduced pitch perturbations supports the wealth of previous evidence showing that the IPS is the is recruited for sensory-motor integration, including pitch regulation. Research done in singing and using a keyboard has been especially important in implicating the IPS in pitch regulation. In singing, Zarate et. al (2008) showed that the IPS and IPL (SMG) were recruited both when ignoring and compensating for introduced pitch perturbations of 200 cents. More generally, this has also been shown in non-singing vocalizations.

Toyomura at. al (2007) showed recruitment of IPS and SMG when pitch feedback was shifted while participants vocalized /a/.

Even outside the domain of vocalization the IPS is recruited during a broad range of keyboard/piano playing tasks, and studies using the keyboard have shown that this recruitment is likely related to pitch regulation as opposed to just movement coordination. Perturbing pitch feedback during keyboard performance specifically elicited activity in IPL (Pfordresher 2014), and the repetition suppression effect for pitch repetition was shown to be most prominent in IPS (Brown 2012). In fact, the IPS is recruited for pitch transformations even in the absence of a set performance task. Studies on melody transformations have shown that the IPS is likely contributing to pitch frequency transformations (transposition) in a similar manner to visual spatial transformations (Foster 2010).

Research on human and non-human primates has found that within IPS there are organized subdivisions that are responsible for tasks varying from visually guided saccades to grasp orientation adjustments (Grefkes and Fink 2005), and that there is a topography of these representations going from posterior portions of the IPS (more involved in visuomotor tasks) to anterior portions of the IPS, close to the postcentral gyrus (more involved in grasping and other sensory-motor transformations). In the present study we show that, within the IPS, cello playing and singing recruit directly overlapping regions primarily in anterior portions of the IPS, which has been found to relate to sensorimotor transformations and motor execution(Culham and Valyear 2006). This overlapping activity extended ventrally to the SMG which, as mentioned above is consistently active in the musical instrument literature (Bangert et al. 2006; Gaab et al. 2003; Toyomura et al. 2007) at least in part due to its role in skilled object related actions like tool use (Binkofski, Klann, and Caspers 2016; Culham and Valyear 2006). The mid to anterior localization of activity within the IPL is consistent with work showing that these regions are most related to motor planning and execution, simple motor behaviours, tactile reception, and a number of speech tasks like phonological short term memory (Tremblay, Shiller, and Ostry 2003).

Overlapping activation when ignoring pitch perturbations

This manipulation allowed us to test the hypothesis that singing and cello playing would also use similar brain regions for ignoring incorrect pitch feedback and maintaining the correct target tone. For this we looked at brain activity when ignoring compared to simple cello playing and singing, together with the conjunction (Figure 3-6). We found that the two instruments overlapped in dorsal pre-motor cortex and in posterior STG, and that only activity in posterior STG was positively correlated with good performance on the task. Our interpretation of this is that the increased activity in auditory cortex resulted from the absence of they typical inhibition of auditory cortex typically seen for monitoring of self-generated sounds (Bendixen, SanMiguel, and Schröger 2012; Sanmiguel, Todd, and Schröger 2013; Mathias, Gehring, and Palmer 2019; Christoffels, Formisano, and Schiller 2007). If participants were no longer linking the perceived sounds to their produced output, and instead increasing their reliance on a forward model and on non-auditory sensory feedback, then it would follow that inhibition of auditory cortex would be weaker and activity in STG would be stronger as a result.

Based on the extensive overlap seen in the simple and compensate conditions, we had predicted somewhat more overlap for the ignore condition, for instance in the aINS or ACC, but the observed regions are consistent with literature on error monitoring and pitch control. We would therefore conclude that the mechanisms involved in representing the target pitch stably in the face of competing inputs, and perhaps inhibiting corrective responses, are shared for both singing and cello, just as the mechanisms for compensation also appear to be shared.

The finding that both instruments recruit dorsal stream regions to a greater extent when ignoring a pitch shift than when simply playing a tone suggests that some motor planning takes place in response to the perturbed feedback even if no movement follows from it. It is well established in singing that, in response to incorrect pitch feedback, people exhibit compensatory reflexes (Burnett et al. 1998; Zarate and Zatorre 2008; Zarate, Wood, and Zatorre 2010). However, the same has not been established in musical instrument performance. This could be preliminary evidence suggesting that a tendency towards compensation exists. However, to demonstrate that such a phenomenon exists more conclusively would require more fine-grained behavioral evidence, as discussed above, and also a neural measure more sensitive to temporal modulations on a fine scale, such as EEG.

The only region we found to be contributing to better task performance when ignoring compared to just playing in tune was the pSTG. This finding suggests that enhanced auditory encoding is likely a key factor in the extent to which participants could ignore the errant auditory feedback. This finding is consistent with the proposal that auditory error maps or auditory state maps are stored within the pSTG (Tourville and Guenther 2011). These auditory error/state maps are hypothesized to compare incoming auditory feedback with the intended auditory target and send an error signal to motor control structures so they can update the positions of the relevant effectors as necessary. While the DIVA model is based mostly on speech research, the same pSTG recruitment has also been shown during musical instrument performance (Baumann et al. 2005; S. Brown et al. 2004) including for the processing of perturbed pitch feedback (Pfordresher and Mantell 2014). PSTG is also named as a key locus for auditory-motor mapping in other models of auditory-motor integration as well (Rauschecker 2011; Baumann et al. 2005).

3.6.2 What we found but can't explain

Subcortical Structures

One of the central findings from our previous paper was that singing recruited the brainstem while cello playing did not. We were not able to replicate that finding likely because the reduced delay in TR (4s compared to 7s used previously) prevented us from catching the peak of the haemodynamic response. Based on our previous work, and the fact that singing explicitly necessitates breathing in order to produce the tone during exhalation while cello playing does not, our hypothesis was that singing would preferentially recruit structures that directly innervate the musculature that controls breathing and vocalization, namely, the Pons and the PAG. In the basal ganglia, we had predicted that the Putamen would be recruited for both singing and cello playing due to its role in sensorimotor adaptation across modalities and effectors. In the current study, we did not find that this was the case. We did find that cello playing showed more extensive activation throughout the brain, including in the basal ganglia, but because these differences were so widespread it is difficult to make any claims about their significance specifically in the basal ganglia.

More, and more extensive cello activation

We once again found that cello playing elicited more, and more extensive activation throughout all areas of interest than did singing in the simple, compensate, and ignore conditions despite increased efforts on our part, compared to our previous study, to ensure that the loudness was comparable. This finding, as we argued previously, supports the hypothesis that the years of explicit cello training that our participants have undergone results in enhanced processing for cello playing relative to singing, which would be in line with related findings, for example that trumpet players show more activation in STG for their own instrument as compared with playing and instrument on which they had less experience (in this case the keyboard), which the authors interpreted as a sign of enhanced processing (Gebel et al. 2013). However, it is still difficult to accept the interpretation that participants have more experience on the cello than they do with their own voice. A complementary interpretation relates to the auditory suppression effect, which is when neuronal activity in STG is actively inhibited for self-generated vocalizations (Christoffels, Formisano, and Schiller 2007). This inhibition occurs even in non-human primates (Müller-Preuss and Ploog 1981; Eliades and Wang 2008). Conversely, STG activity has been found to be enhanced for self-initiated sound producing movements, like playing the keyboard (Reznik et al. 2014). So the observed difference between singing and cello playing, at least in STG, could be accounted for by a combination of auditory suppression during self-initiated vocalization and enhanced processing during for cello playing. Increased activation in brain areas related to motor planning and execution during cello playing may be related to the larger, more extensive movements, involving multiple muscle groups required to play the cello compared to singing.

3.6.3 Future Directions

Subvocalization

As before we cannot directly rule out the possibility that cellists were subvocalizing when playing the cello, or imagining themselves playing the cello while singing, and that these covert activities are partially responsible for the observed overlap in recruited brain regions. However, we can say conclusively that they were not overtly vocalizing during the cello condition or producing sounds on the cello while singing, as we have the recordings to verify it; so, at the very least, the descending motor commands were specific to the relevant effectors, as also reflected in the M1 regions recruited. Future work could directly address the open question of subvocalization in a few ways. The most straightforward would be to measure EMG of the hands and larynx during a cello task to see if participants are subvocalizing as a strategy to keep the target tone in memory (Pruitt, Halpern, and Pfordresher 2019) and/or during performance of the task itself. A more direct approach would be to attenuate the descending motor commands using transcranial magnetic stimulation over larynx area of motor cortex. In both cases, the effect on cello performance accuracy could act as a measure of the extent to which laryngeal activity is necessary for musical instrument performance.

Compensatory Reflex

Our finding that ignoring a pitch shift recruits dPMC to a greater extent than just playing a tone in the absence of a pitch shift supports the idea that some compensatory movement is planned in response to the perturbed pitch feedback. in cello playing as well as in singing. Future behavioural analyses will take a more fine-grained approach to explore if such a compensatory reflex exists. However, EEG is much better suited to exploring the timecourse of responses than is fMRI. Our future work will follow up on this question by looking at the latency and magnitude of the ERP complexes that occur in response to sudden onset pitch perturbations.

Limitations

A less exciting but still possible interpretation of the overlapping activity discussed in this paper is that the resolution of the chosen fMRI sequence, and the design of this study, were not fine grained enough to tease apart processes that are occurring very near one another but aren't actually overlapping. There are several ways to address this issue. One would be to use a repetition suppression approach, as used for example by Brown et al. in a piano production study (R. M. Brown et al. 2012), which allows activity elicited by different conditions within the same region of interest to be dissociated. A more technical solution would be to use higher resolution imaging (eg. choosing smaller voxels or using a 7T scanner), and a multivariate analysis approach, the degree to which the pattern of activity within a patch of voxels contains information about a given task could be tested, even if the average activity within that patch is similar.

While the similarities between singing and cello playing can still be interpreted despite the different baseline levels of activation, differences between the two instruments are somewhat more difficult to interpret, because they could be due either to fundamental differences in the cognitive demands associated with each, or they could simply be a more trivial consequence of differences in motor control required, or the sound quality produced. In the present study the overall trial duration was reduced compared to what we used previously so that each participant could perform each condition. As a result, the sparse sampling design was not optimized to catch the peak of the hemodynamic response and signals that already had lower signal strength to begin with, like in the basal ganglia, did not reach significance in this study.

Chapter 4 - Similar ERPs index pitch regulation in cello playing and singing

4.1 Preface

In Experiment 1 and Experiment 2 we used fMRI to show that cello playing and singing elicit activation within very similar brain structures for singing/playing individual target tones, compensating for introduced pitch perturbations, and ignoring them. These brain structures include those related to movement planning/execution, feedback monitoring, and auditory-motor integration. However, fMRI is not known for its high temporal resolution, and pitch regulation is a temporally precise, multi-step process that occurs over the course of milliseconds, not seconds. In Experiment 3 we build on our previous findings by using EEG, a temporally precise brain imaging modality, to test the hypothesis that the timecourse of the brain responses for singing/playing is also similar. Experiment 3 has been prepared as a manuscript for publication and has been formatted for inclusion in the thesis.

4.2 Abstract

Sensory feedback allows us to rapidly adjust ongoing movements to increase the likelihood of attaining a sensory goal. Vocal pitch regulation is a prominent example of this phenomenon in the auditory-motor domain and has been studied extensively in both speech and singing. Research done in these domains using pitch feedback perturbation paradigms has shown that people can use auditory feedback to rapidly adjust their vocal output in order to decrease the discrepancy between the sound they intend to produce and the sound they are actually producing. We recently demonstrated that this behaviour generalizes to performance of a continuous pitch instrument (the cello; and using fMRI, that the brain regions for singing/playing are directly overlapping in areas related to motor planning/initiation (SMA, ACC), feedback monitoring (STG), and sensory to motor transformations (SMG, IPS, dPMC). It follows that, if cello playing is using the same brain network to accomplish the same task, then the activation timecourse of singing/playing should be similar as well. Perturbations of the fundamental frequency (F0) during speech vocalizations have been shown to elicit a P1-N1-P2 ERP complex in EEG (Liu and Larson 2007; Behroozmand, Liu, and Larson 2010), which is thought to reflect the detection of incorrect pitch feedback (P100) and the selection of a compensatory response (N100-P200). In the present experiment we implemented a pitch perturbation paradigm identical to that used in our previous fMRI experiments (Experiments 1, 2) to test the hypothesis that singing/playing would elicit the P1-N1-P2 complex, and that the responses would be similar between them. What we found was that singing/playing did elicit a 3-compnonet ERP complex that was similar between the two instruments, but that the components that were elicited may be different in latency from what we expected based on the speech vocalization literature. The observed similarity between the two instruments supports our central hypothesis, that cello playing makes direct use of the singing network to accomplish pitch regulation. In the current study, we propose that the observed components reflect error detection followed by a context updating process, with participants updating their sensory context in response to the pitch feedback perturbation as has been shown in piano playing (Mathias, Gehring, and Palmer 2019) and in other sensory contexts (Polich 2007).

4.3 Introduction

Singing and playing musical instruments in tune requires a skilled execution of rapidly timed movements and fine, ongoing adjustments in response to the sounds those movements produce. A number of brain networks contribute to this process including those for motor planning, motor execution, feedback monitoring, and error correction. These networks have been studied extensively in both singing (Zarate and Zatorre 2005, 2008; Zarate, Wood, and Zatorre 2010; Kleber and Zarate 2014) and musical instrument playing (Parsons et al. 2005; Gebel et al. 2013; Lotze et al. 2003; Segado et al. 2018) and these studies have shown that both modalities rely on many of the same brain structures, as specifically shown for playing the cello in chapters 2 and 3 of this thesis.

However, the various components associated with pitch production and adjustment happen within milliseconds, which is well beyond the timing resolution of the BOLD signal measured with fMRI. As such, much of what we know about the temporal dynamics of auditory-motor integration comes from work done with EEG, which has a much higher temporal resolution than fMRI. Studies of speech and vocalization have found a characteristic series of ERP responses associated with feedforward control and error adjustment (Liu and Larson 2007; Behroozmand, Liu, and Larson 2010; Scheerer and Jones 2018). To our knowledge no previous study has investigated the EEG correlates of pitch regulation in singing or instrument playing. We propose that studying the EEG correlates of pitch regulation in singing and comparing them with those of playing a continuous pitch instrument that is independent of the vocal apparatus (in this case the cello) will extend current models of auditory-motor integration by exploring whether the time-varying EEG responses are similar or different than those observed for vocalization. To address the question of the dynamics of the pitch production and correction process, in the present study we used a pitch-perturbation paradigm previously developed for an fMRI study of singing (Chapter 3) and measured the corresponding brain response with EEG.

Research on singing expertise has shown that when auditory feedback is altered, untrained singers have more difficulty suppressing a pitch-shift reflex than do expert singers (Zarate, Wood, and Zatorre 2010), which is in line with the interpretation that the default action is to compensate for incorrect pitch feedback. This regulatory pitch reflex is phylogenetically old and is seen not just in humans but also in several animal models as well. For instance, bats exhibit a highly developed vocal pitch shift reflex to help them compensate for the Doppler shift that they experience during flight (Metzner 1996; Moss and Sinha 2003; Schuller and Radtke-Schuller 1990). Bats are not known, however, for playing the cello. Musical instrument playing is a relatively new, cultural task unique to humans that is unlikely to have any direct evolutionary significance. Although expert cellists were shown in Chapter 3 to compensate for introduced pitch perturbations equally well when singing and when playing notes on the cello, studies have also shown that cellists cannot maintain a stable pitch when playing the cello in the absence of auditory feedback (J. Chen et al. 2008, 2013) in the same way that expert singers have been shown to do when singing (Kleber et al. 2017). It is therefore reasonable to expect that, even if cello playing recruits brain areas within the singing network to accomplish a pitch regulation task, there may also be some differences between the two tasks at the level of online error monitoring and correction that are detectable using EEG.

Previous EEG studies of pitch feedback perturbations during vocalizations have shown that three ERP components, the P100, N100, and P200, are specifically related to perception-action coupling (Liu and Larson 2007; Behroozmand et al. 2015; Scheerer et al. 2013; Scheerer and Jones 2018). The P100 has been shown to be insensitive to the magnitude of feedback perturbations, and as such is thought to reflect a generic difference or error detection process (Jones, Scheerer, and Tumber 2013; Scheerer and Jones 2018). Conversely, the magnitudes of the N100 and P200 have been shown to be closely linked to the magnitude of introduced feedback perturbations, and to increase during auditory-motor tasks compared to perception-only or motor-only conditions. Consequently, they are thought to be more closely linked to the selection and execution of a compensatory motor response. The N100 is specifically thought to reflect the difference between the sensory prediction of a planned motor action and the resulting sensory feedback since this component has been shown to be attenuated (lower amplitude) during self-initiated vocalizations compared to passive listening and enhanced (higher amplitude) during vocalization under frequency-altered feedback conditions (Heinks-Maldonado et al. 2005; Jones, Scheerer, and Tumber 2013; Scheerer and Jones 2014; Scheerer et al. 2013; Scheerer and Jones 2018). The P200 has similarly been found to have a positive, linear correlation to pitch feedback perturbation magnitude, and to be similarly enhanced during auditory-motor conditions relative to auditory only conditions (Behroozmand and Larson 2011; Scheerer et al. 2013; Jones, Scheerer, and Tumber 2013).

Studies of pitch regulation in instrumental playing have looked at the EEG correlates of keyboard performance. These studies examined brain responses to self-generated pitch errors, rather than responses to perturbed feedback. Here they examined the Error Related Negativity (ERN) and the Error Positivity (Pe), rather than the P1-N1-P2 complex. The ERN is a negative deflection that has been shown to occur around 100ms after an incorrect response has been initiated, while the Pe is a positive deflection occurring at around 200-500ms post-error and is thought to reflect the conscious perception of an incorrect response (Waszak and Herwig 2007). A 2009 study of keyboard performance in the presence and absence of auditory feedback showed that auditory feedback did not have a significant influence on the ERN, and that the ERN was detectable up to 70ms prior to an error being committed, suggesting that this component relates to a feedforward model of pre-planned actions (Ruiz, Jabusch, and Altenmüller 2009). Conversely the Pe was significantly larger during auditory-motor conditions than motor only, suggesting that it is of greater relevance for the feedback component of online error monitoring and, potentially, for the selection of an appropriate compensatory response. However, the authors themselves acknowledge that this effect may not generalize well to continuous pitch string instruments like the cello, given that continuous monitoring of auditory feedback is required in order to play in tune on the cello. They contrast this with piano production, which is relatively stable in the absence of auditory feedback.

The idea that vocal production is more independent of auditory feedback is less true in the context of singing than it is in speech. Research on singing has shown that only expert singers are able to maintain pitch accuracy in the absence of auditory feedback, presumably due to their ability to rely on a more well-developed forward model of the associated motor commands. In a study looking at singing with and without auditory feedback, non-singers required feedback in order to maintain a stable pitch (Kleber et al. 2017). It is also important to consider that the previously discussed keyboard study was looking at sequence production as opposed to single tone production, given that error expectancy has been shown to play a role in EEG activity (Ruiz, Jabusch, and Altenmüller 2009). Research done on unexpected perturbations of auditory feedback (specifically using the oddball paradigm) has shown that this type of violation can elicit a Mismatch Negativity (MMN) response, which is a negative deflection around 200ms following the expectancy violation, as well as a P300 between 200-500ms post expectancy violation (Waszak and Herwig 2007; Ruiz et al. 2014; Nittono 2006). While the MMN is a negative deflection that occurs at a similar latency to the N200, it is thought to be a distinct component due to the differing distribution (fronto-central as opposed to parietal). Similarly, the P300 response is very similar to the Pe, but is thought to occur in response to external manipulations of feedback, like pitch feedback perturbations, as opposed to self-initiated errors (Ruiz, Jabusch, and Altenmüller 2009).

In the current experiment, we hypothesized that pitch feedback perturbations during singing and cello playing will elicit the P1-N1-P2 response characterized for other vocalizations (Liu and Larson 2007; Behroozmand, Liu, and Larson 2010; Scheerer et al. 2013), in accordance with our general concept derived from the findings in chapters 2 and 3, that cello playing and singing rely on similar neural substrates. Behaviourally, we expected participants to be able to sing and play back individual target tones as was the case in Experiment 1 and Experiment 2. We also expected them to be able to compensate for and to ignore sudden onset pitch feedback perturbations (100 cents) as in Experiment 2. Regarding the EEG correlates of compensating, we predicted that singing and cello playing would have a similar activation timecourse, much as they had overlapping brain activity observed in Experiments 1 and 2. We had initially hypothesized that the pitch feedback perturbation would elicit a pitch shift reflex in singing but not in cello playing, and that this difference would be indexed by a difference in the N1 and P2 responses. However, because we did not see any behavioural evidence of such a reflex in either singing or cello playing in Experiment 2 (3.5.1), no differences were expected. We predicted that both ignoring and compensating for perturbed pitch feedback would elicit a P100 component that would not differ between instruments or conditions. This prediction follows from the premise that the P1 is not sensitive to perturbation magnitude and is rather thought to be a generic error detection process. Further, we predicted that the amplitude of the N1 would be less for the ignore condition than the compensate condition, following from the premise that the N100 is related to a mismatch detection between predicted sensory consequences of an action and incoming sensory feedback. Finally, we predicted that the P2 would be larger in the compensate condition relative to the ignore condition, as the link between perception and action is preserved and previous research has shown that the

P2 is larger for auditory-motor conditions than for passive detection of a feedback error (Behroozmand, Liu, and Larson 2010).

4.4 **Materials and methods**

4.4.1 Subjects

Fifteen expert cellists (10 Female) were recruited from the Montreal community. Three were excluded from the analysis due to technical problems. The remaining twelve were highly trained (mean years experience = 12.93 +- 4.46, mean starting age = 9, mean practice hours per week = 12.07), reported normal hearing, no neurological disorders, and had no contraindications for MRI scanning. This study was approved by and carried out in accordance with the recommendations of Montreal Neurological Institute Research Ethics Board and the McConnell Brain Imaging Centre. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

4.4.2 Experimental Paradigm

Stimuli and Task Conditions

Participants were asked to listen to and subsequently reproduce target tones both by singing and by playing them on a battery-powered Yamaha silent electric cello. Unlike in Experiments 1 and 2 where participants were laying down, participants in this experiment carried out the task in a natural seated position. There were three experimental conditions: Sing/Play, Listen, and Rest. For the first two of these conditions, we presented target tones that were 1 s long. Presented tones were (using Western music notation) E3 and F$^{\#}$3 for cello, and either E3 and F$^{\#}$3, or E4 and F$^{\#}$4 for singing depending on participant's vocal range. Target tones for the cello playing condition were recorded on a cello, while those for the vocal condition were recorded by either a female or male vocalist. Participants heard and imitated the pitch sung by someone of their own gender. Pitch perturbations of 100 cents (one semitone) up or down were introduced between 1 and 1.5 seconds (jittered in steps of 500 ms) during the Sing and Play conditions. In separate blocks participants were instructed to either compensate for (Compensate) or to ignore (Ignore) the introduced pitch shift. Each block had 6 sing trials and 6 play trials. The pitch perturbation was introduced on 2/3 of the trials (8 per instrument per block, 80 trials total), the remaining third of the trials (4 per instrument per block, 40 trials total) had no pitch shift (Simple condition). In a separate condition (Attenuated), pink noise attenuated auditory feedback to the greatest extent possible, and participants sung/played an instructed tone with the masked feedback (1 block per run, 120 trials total). Visual instructions to compensate or ignore were displayed on a black screen during the presentation of the auditory cue.

Participants underwent a brief training session prior to the start of the experiment. First, they were familiarized with the Listen and Play trial structure. Next, they were familiarized with the Ignore and Compensate conditions. In the Ignore condition, sudden onset pitch shifts were introduced, and participants were instructed to ignore the incorrect feedback and continue singing or playing the 'correct' note they heard as the cue. In the Compensate condition, sudden onset pitch shifts were introduced, and participants were instructed to compensate for the shift to return to hearing the original target note.

For both the familiarization session and the EEG session, a microphone was placed approximately 5 cm from the mouth behind a wind guard and the electric cello device was placed in a comfortable playing position that did not interfere with the EEG or microphone. Bilateral sound delivery was provided via insert earphones (Sensimetrics, Dayton Audio DTA-1 amplifier). The microphone and cello were connected to a Zoom F8 portable field recorder, and to a midi controlled TCHelicon VoiceOne pitch shifter. The shifter was used to shift the pitch feedback on some trials or mask the feedback on other trials. Pink noise was played continuously through the headphones to reduce bone conduction so that audio was being delivered exclusively through the earphones. All sound levels were adjusted on a per subject basis, but on average pink noise was presented at 78.3 dB SPLA and auditory targets were presented at approximately 15.6 dB above the noise floor. The experiment was run using custom scripts written in Python. All vocalizations and cello sounds were digitally recorded to the Zoom F8 digital recorder (Figure 4-1). To minimize movement artifacts, participants were instructed to sing with their mouth open but relaxed and to swallow only between trials. For cello playing, participants rested their elbows on the armrests of the chair and used only their lower arm and wrist to draw the bow.

On Sing and Play trials, participants were instructed to produce the target tone for 2.5s. For singing trials, participants were instructed to sing as quietly as possible with their jaw relaxed to reduce breathing artifacts in the recorded signal and movement artifacts in the EEG signal. For cello playing trials they were instructed to use as few bow movements as possible to produce the note in order to reduce movement artifacts (approximately one bow per trial). Between trials participants were instructed to relax their hands and to move as little as possible.

Listen and Rest trials, and Attenuated trials were not included in the analyses presented in this manuscript as they did not relate to the central hypotheses and will be analysed separately at a later point in time.

EEG Setup

EEG data were continuously recorded throughout the session at 10KHz using the BrainProducts 64-Channel Acticap system. Participants were fitted with the 64-channel EEG cap that best suited their head size and electrode position was recorded at the end of each session using a Polhemus digitizer. Electrodes were adjusted until the impedance of each was below 5Ω. Cables were clipped to the back of participants' shirts at the base of their neck and wrapped in sports wrap to minimize movement artifacts related to the cello playing task.

TTL pulses were used to timestamp the EEG data throughout the trial to allow for subsequent alignment of the trials to specific points of interest. For all trials, a TTL pulse was sent at the onset and offset of the auditory target presentation, the onset of tone reproduction, the timepoint at which the pitch shift was sent (on trials where a pitch shift was introduced), and at timepoint where the pitch shifter executed the pitch shift (90ms delay from command send to command execution).

4.4.3 Behavioral Analyses

Individual trials of singing and cello playing were analyzed with pitch information extracted from the recorded audio signals. The audio for each trial was first segmented from the continuous 6-track audio using Audacity software. The trials were then processed using a custom analysis pipeline implemented in Python, with a GUI for visualizing and optimizing analysis parameters (Figure 2-3).

Pitch extraction was also done using the custom GUI developed previously for use with the fMRI compatible cello interface to allow for a direct comparison between the two datasets. In this software harmonics 3–10 of the cello and singing tones were used for pitch extraction as it was found to give a more stable pitch estimate than f0. To reject any room noise and to isolate harmonics of interest, the raw microphone signal was high-pass filtered with a cutoff at 367 Hz and low-pass filtered with a cutoff at 4,216 Hz. Pitch estimation was then performed using the YinFFT algorithm provided in the Python module Aubio (Brossier 2007), producing a time-series of pitch estimations (detected harmonic, in Hz) and confidence ratings (between 0 and 1). Estimates were adjusted to their representative fundamental pitches before selecting stable pitch regions for further analysis. Stable pitch regions were defined as: segments of at least 150 ms in which the rate of change of the pitch did not exceed 100 Hz/s (or approximately 0.07 Hz per 32-sample pitch estimation window at the sampling rate of 44,100 Hz). Of these regions, only those that maintained a confidence rating of at least 0.7 by the algorithm (derived from the signal d prime) were included. Trials were rejected if no regions were found to meet the stability and confidence criteria, if participants started playing after the pitch shift was introduced, or if they stopped playing before the end of the trial. In total, approximately 83% of trials were retained. Rejected trials were excluded from the EEG analysis.

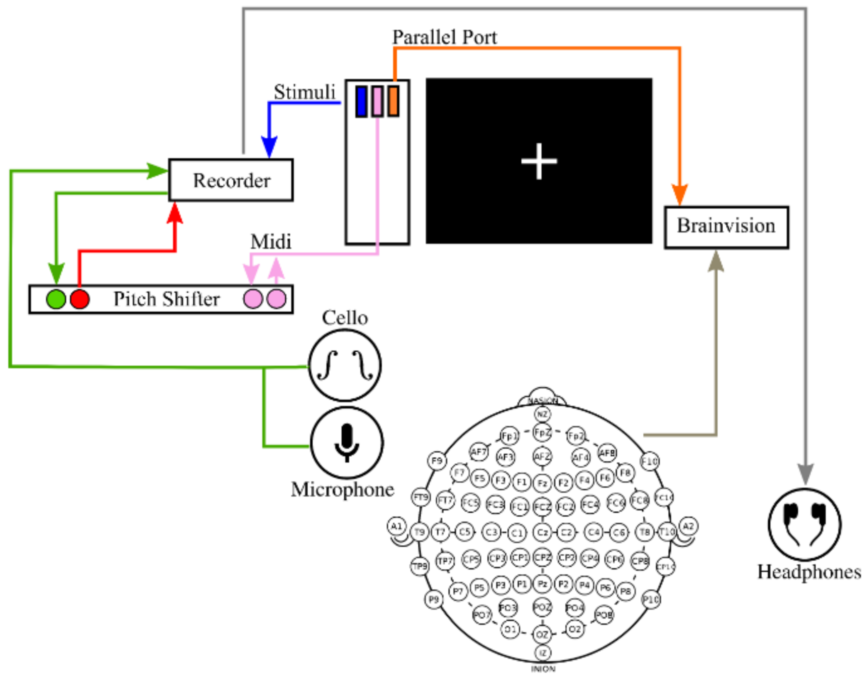

Figure 4-1 EEG Experimental Setup | Experimental setup used to present stimuli, cello audio, and singing audio through headphones. Pitch shifter allowed for audio feedback from cello/singing to be shifted or attenuated on specific trials while still presenting audio stimuli and masking noise.

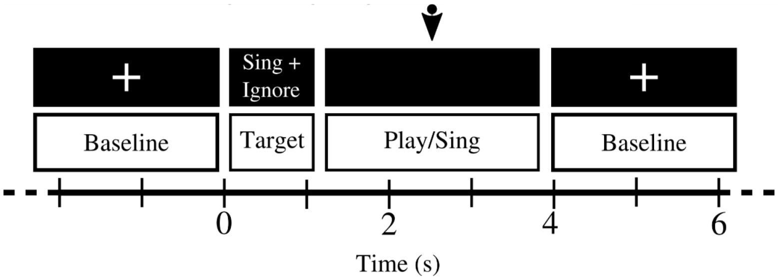

Figure 4-2 Listen/Play Trial Structure | 2.3s of silence (used as a baseline) preceded the presentation of the auditory target for each trial. The target onset was randomly jittered by up to 500ms to avoid an anticipatory response.

Pitch extraction was performed for both the unshifted raw output track (produced) and for the shifted headphone track (perceived) to allow for a direct comparison between the produced and perceived audio. Performance scores were calculated on a per participant basis as follows:

The difference between the produced tone from the target tone (expressed in cents, or 100^{ths} of a semitone) was analyzed within three 150ms windows before and after shift onset, Pre-shift(-150 to 0), Early post-shift (0-150), and Late post-shift (1000-1150) for each of the Simple, Compensate, and Ignore conditions (Equation 4-1).

An overall accuracy score for each participant was then determined by calculating their mean pitch accuracy across all trials. Finally, scores were analyzed using a two way (instruction by timepoint) ANOVA implemented in GraphPad.

$$cents = 1200^{*} log_2 \left(\frac{Freq_{produced}}{Freq_{target}} \right)$$

Equation 4-1 Conversion from absolute frequency difference to normalized cent difference

Simple tone reproduction

On 1/3 of the trials within the Ignore and Compensate blocks (80 trials total) no pitch shift was introduced (Simple condition). In other words, the difference between perceived and produced audio remained at 0 cents throughout the trial. No special instruction was given to inform participants that no perturbation would be introduced – Simple trials were randomly interspersed with the trials that did contain a pitch-feedback perturbation. Trial accuracy was determined by averaging the cent deviation from the target tone at the Pre- and Late post-shift timepoints. This was done to account for any natural drift in pitch throughout the trial.

Ignore and Compensate

In these trials (160 trials total: 80 ignore, 80 compensate) participants were instructed to either compensate for or to ignore a 100 cent (semitone) pitch shift introduced between 1 and 1.5 seconds after the trial onset. Shift onset was jittered to prevent participants anticipating the shift. On Compensate trials, participants were instructed to compensate for the introduced perturbation in order to go back to hearing themselves play the original target pitch. In other words, in response to a 100 cent downwards shift, participants would need to increase their produced pitch by 100 cents, and vice-versa for an upward shift. On Ignore trials they were asked to ignore the perturbation and continue playing or singing the initial pitch.

Each participant was given a score for the Simple, Ignore, and Compensate conditions based on how well they were able to accomplish each task. All scores we calculated based on the average cent difference between the produced output and the target tone at the Early- and Late post-shift timepoints across all retained trials. For the compensate condition, perfect performance was considered to be 100 cents in the direction opposite to the introduced pitch perturbation, whereas for the ignore condition perfect performance was considered to be 0 cents deviation from the target tone. All three scores were averaged to give a global participant score.

4.4.4 Preliminary EEG Analyses

EEG data were analyzed using Brainstorm analysis software. Data were notch filtered at 60Hz (including second and 3rd harmonics), and bandpass filtered from 0.5Hz to 40Hz. Eye movement, saccade, reference artifacts were characterized using ICA and removed from the data. The clean signal was segmented around the pitch shift (shift - 300ms to shift + 600ms). DC offset was removed based on the 300ms pre-trigger baseline. 4 average traces were generated for each participant: Sing Ignore (40 trials), Sing Compensate (40 trials), Cello Ignore (40 trials), and Cello Compensate (40 trials). These traces were then averaged across participants, and ERP components were characterized using the Matlab find-peaks function within 50ms windows around 100, 200, and 300ms. The amplitude at each of the peak locations was then extracted from the individual traces, and analysed using a 2-way (component x instruction) ANOVA.

4.5 **Results**

4.5.1 Behavioural Findings

We first carried out a behavioral analysis to confirm that our participants, who were all cello players, could sing target tones and also play them on the cello with an average tone accuracy of at least a quartertone (50 cents). The behavioural analysis further allowed us to determine the extent to which participants could compensate for introduced pitch perturbations and also the extent to which they could ignore them. Because in our previous research we established that there were no significant performance differences between singing and cello playing, the two instruments were analysed independently in order to decrease the degrees of freedom.

Accuracy during simple tone reproduction

We first replicated our previously published finding that cellists could both sing and play target tones on the cello with a high degree of accuracy when no pitch shift was introduced (Figure 4-3). In the Simple condition, participants stayed within 15 cents deviation from the target tone at all three timepoints both when playing the cello ($\text{mean}_{\text{cello}} = 0.046 \pm 0.286$ cents) and singing ($\text{mean}_{\text{sing}} = 9.783 \pm 0.383$ cents). Additionally, there was no significant difference between the Early and Late post-shift timepoints ($p_{\text{sing}} = 0.95$, $p_{\text{cello}} = 0.99$), which shows that their produced pitch was stable. We also found no significant difference between the pitch accuracy in the Simple condition compared to the pre-shift timepoint in the Compensate and Ignore conditions for both singing ($p_{\text{comp}} = 0.97$, $p_{\text{ign}} = 0.98$) and cello playing ($p_{\text{comp}} = 0.99$, $p_{\text{ign}} = 0.97$), showing that participants were able to accurately reproduce the target tone before the pitch perturbation was introduced.

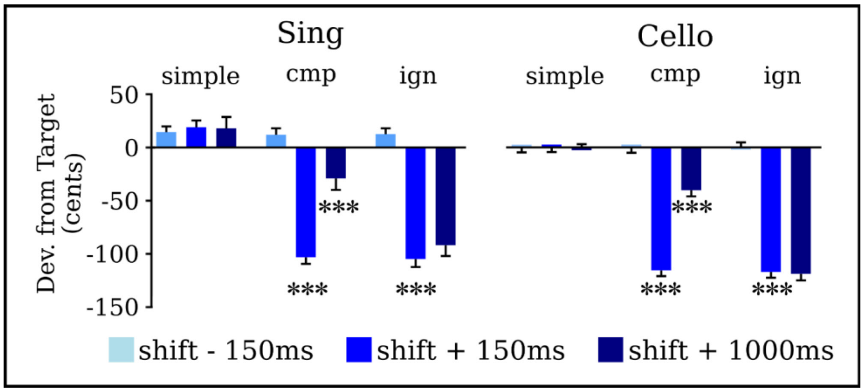

Figure 4-3 Deviation from target tone during Simple, Compensate, and Ignore conditions | Bins are each 150 ms long and represent average pitch at the Pre-Shift (light blue), Early Post-Shift (blue), and Late Post-Shift (dark blue) timepoints as heard through the participant's headphones. For the Simple condition where no shift was introduced the 'shift' was considered to be at the midpoint of the trial. For Compensate and Ignore trials, perturbation was 100 cents.

Accuracy during compensating and ignoring

Our next hypothesis was about participants' ability to compensate for a sudden onset 100 cent pitch perturbation. Because we had no specific hypotheses related to the directionality of these perturbations, upwards and downwards perturbations were treated as equivalent; values from the downwards perturbation were therefore multiplied by -1, and then averaged. Cent deviation from the pre-shift baseline (-150 ms to 0) was considered at 2 different timepoints. Time point 1 (post-early) was selected to capture any immediate response to the perturbation (0 to 150 ms), and timepoint 2 (post-late) was selected to capture the completed compensatory response (1000 to 1150 ms). The data was analysed using a two way Instruction (Compensate, Ignore) by Timepoint (Pre, Post-Early, Post-Late) ANOVA (Figure 4-3). Cello playing and singing were considered separately. In singing, we found a significant effect of Instruction ($F(2,106) = 79.46$, $p < 0.0001$) and Timepoint ($F(2,106) = 73.51$, $p < 0.0001$), with a highly significant instruction by timepoint interaction ($F(4,106) = 26.02$, $p < 0.0001$). Similarly, in cello playing, we found a significant effect of Instruction ($F(1.734,62.42) = 356.4$, $p < 0.0001$) and Timepoint ($F(2,36) = 113.7$, $p < 0.0001$), with a highly significant Instruction by Timepoint interaction ($F(4,72) = 128.2$, $p < 0.0001$). Pairwise contrasts were carried out using Tukey's t-tests. For both Compensate and Ignore instructions in both cello playing and singing, participants showed a highly significant difference between the pre-shift baseline and the 150ms following the introduced perturbation ($p(\text{post-early}) < 0.0001$). These post-early responses did not differ from one another in singing ($p = 0.99$) nor in cello playing ($p = 0.98$). In other words, in the 150 ms following the introduced pitch shift, participants responded to the pitch change as much in the compensate condition as they did in the ignore condition. However, when instructed to compensate, participants showed a highly significant ($p < 0.0001$) compensatory change from the post-early- to post-late timepoints ($\text{mean}_{\text{cello}} = -56.53$ cents, $\text{mean}_{\text{sing}} = -55.59$ cents) which did not occur when they were instructed to ignore. When instructed to ignore, participants showed no significant change ($p_{\text{cello}} = 0.70$, $p_{\text{sing}} = 0.52$) from post-early to post-late timepoints ($\text{mean}_{\text{cello}} = 1.55$ cents, $\text{mean}_{\text{sing}} =-9.76$ cents).

4.5.2 Preliminary ERP results

Our first analysis compared the ERP response to the sudden onset 100 cent pitch perturbation under compensate and ignore instructions for both singing and cello playing averaged across all 64 electrodes for all 12 participants. Averaged responses showed 3 main components, which were consistent across all instrument and instruction conditions. The first is a large positive response at a latency of ~100ms following the onset of the pitch perturbation. The second is a large negative deflection at ~200ms post perturbation and the third is a positive deflection at ~300ms post perturbation. The peak amplitude at each of these three timepoints was extracted from the individual traces to test for significant differences across conditions. The 2-way (instrument x component) ANOVA showed no significant differences between any of the 4 conditions; however, the ignore condition for the cello has a much higher variance in the magnitude of the second and third components relative to singing. For a subsequent analysis (Figure 4-5), we looked at the subset of 13-electrodes (Cz, C3, C4, T3, T4, Fz, F3, F4, F7, F8, Pz, P3, and P4) used in Liu and Larson (2007). The three-component response is still observed in this subset of electrodes, for each of the four conditions, though the third component occurs later and is smaller in the cello ignore condition than in the other three.

We further investigated the response by plotting the scalp distribution for each of the four conditions at each of the three timepoints (Figure 4-5). The distribution for the first two components seems highly consistent for all four conditions. However, the third component appears to be less pronounced and occur later for the both of the cello conditions (Figure 4-5A, 4-5C) than for the singing conditions (Figure 4-5B, 4-5D). This is especially evident for the Cello Ignore condition where the third component peaks ~200ms later than in the other three conditions.

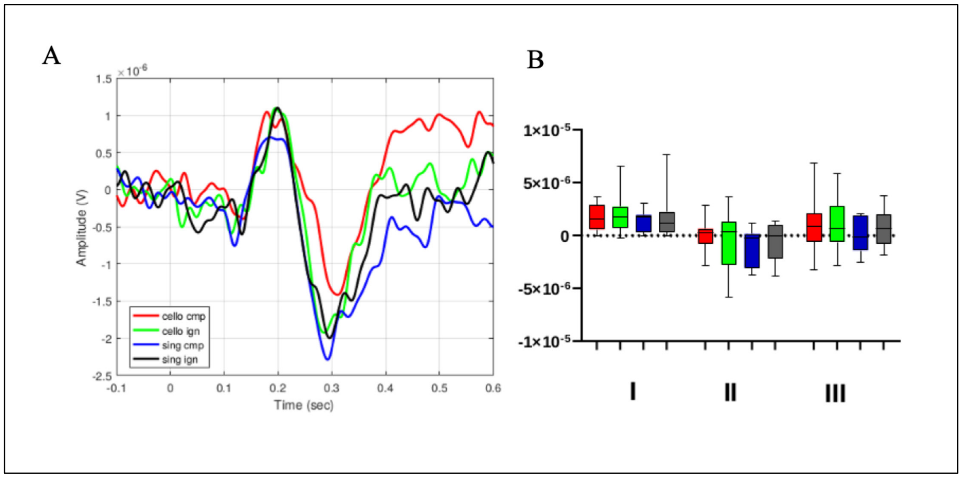

Figure 4-4 ERP Responses to pitch-perturbed feedback | A) ERP trace for 300ms and 600ms following a sudden onset 100 cent pitch shift when compensating on cello (red) or singing (blue) and when ignoring on cello (green) and singing (grey). Three ERP responses were identified (I,II,III) when ignoring or compensating for sudden onset pitch perturbations. 0 indicates the timepoint that the pitch perturbation was sent from the computer and the dotted line indicates the timepoint at which the pitch feedback perturbation was implemented (fixed 90ms delay) B) Mean amplitude of the three ERP components for the four conditions (cello compensate/ignore; sing compensate/ignore). Output of 2-way (Instrument by ERP component) analysis shows no significant effect of instrument or instruction condition, and no interaction. Bars indicate standard error

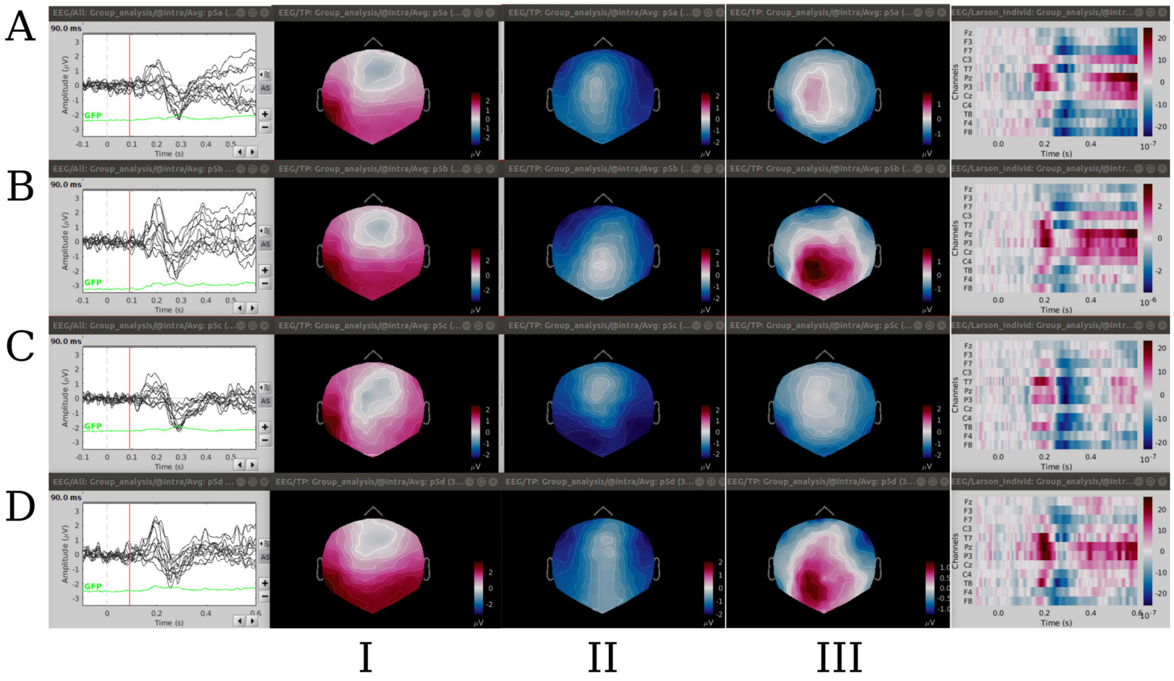

Figure 4-5 ERP Response by condition at three timepoints (I II II) | The first column shows the individual electrode traces for a subset of 13 electrodes averaged across 12 participants, columns 2-3 show the scalp distribution at each of three timepoints (I, II, III) and column 4 shows an image representation of the response at each of the 13 electrodes averaged across 12 participants for A) Cello Compensate B) Sing Compensate C) Cello Ignore D) Sing Ignore. Both instruments and timepoints show a three-component response to introduced 100 cent pitch feedback perturbation (red line, first column). Component I is a positive temporoparietal response that peaks around 100 ms post perturbation onset. Component II is frontotemporal response that peaks 200 ms post perturbation onset. Component III is a positive centroparietal response that begins around 300ms post-perturbation onset, later in (C) the Cello Ignore condition.

4.6 Discussion

Our goal was to test the hypothesis that the activation timecourse of pitch regulation in cello playing is similar to that of singing due to the recruitment of overlapping brain networks that we demonstrated in Experiment 2. We found that cello playing and singing both elicited a similar three-component ERP complex, and that the peak amplitudes of the components did not differ between instruments. This supports our main hypothesis that cello playing is making direct use of structures within the singing network for pitch regulation. However, the timecourse of the second and third peaks is delayed relative to those typically expected for the P1-N1-P2 elicited in previous studies of speech vocalizations (Behroozmand and Larson 2011; Liu and Larson 2007). While a difference in latency does not, on its own, allow us to determine whether or not the observed components reflect the same underlying process, we propose that the structure of our task, which differed from that of Liu and Larson by including a target tone, may have lead participants to compare incoming incorrect auditory feedback to a previously established sensorimotor memory (Waszak and Herwig 2007; Nittono 2006; Ruiz, Jabusch, and Altenmüller 2009), and use it to update the context of the sound to movement mapping (Polich 2007) leading to responses whose timecourse is more consistent with an N200- P300 despite not having been elicited by an oddball paradigm. However, the observed ERP components are still highly relevant to auditory-motor integration and provide insight into the brain mechanisms underlying pitch regulation in continuous pitch instruments.

4.6.1 What participants did

Behaviourally we found that our participants, who were all expert cellists, could both ignore and compensate for 100 cent sudden onset pitch feedback perturbations when singing and playing the cello. This replicates our previous finding from Experiment 2. The finding that cellists can compensate for pitch-perturbed feedback equally well on both instruments is important since it demonstrates that the behaviour we are measuring is comparable to that reported in the literature for speech vocalizations, and also that the behaviour we observed in Experiment 2 generalized to the use of an electric cello in a natural, seated position. If this were not the case, any EEG findings would be difficult to interpret in the context of the existing literature (Behroozmand and Larson 2011; Liu and Larson 2007; Elman 1981; Scheerer and Jones 2018), as would any comparisons between instructions or instruments.

The finding that cellists can ignore equally as well on cello and when singing is also important for interpreting our results. We had initially hypothesized that, when singing but not on the cello, participants would reflexively compensate for the pitch feedback perturbation and that this would be reflected in the EEG response. However, we did not see a difference between the Early-Post and Late-Post timepoints in either instrument during the Ignore condition. As such, we would not expect to see a difference in the neural response either. The behavioural finding that participants can successfully ignore the pitch feedback perturbation allows us to better interpret our findings from the Compensate condition as well since it shows that the compensatory response is purely voluntary.

The reason our manipulation did not elicit the reflexive compensatory response that has been demonstrated during both speech and non-speech vocalizations (Zarate, Wood, and Zatorre 2010; Scheerer and Jones 2018) is likely related to the magnitude of the pitch shift we selected (100 cents). The pitch regulatory reflex in singing was previously characterized for very small shifts of 25 cents (Zarate, Wood, and Zatorre 2010), and research done on F0 perturbations during speech vocalization has shown that people reflexively compensate for feedback perturbations of as little as 9 cents (Scheerer and Jones 2018). Moreover, previous research comparing expert singers to nonmusicians has shown that larger magnitude shifts are easier to ignore for expert singers than smaller shifts (Zarate, Wood, and Zatorre 2010; Zarate and Zatorre 2008). Although our cellists were not trained singers, they had received extensive musical training and it is likely that this experience also allowed them to be better able to ignore the 100 cent shift than would be the case with nonmusicians. This ability to avoid reflexive compensation may be related to the fact that participants are less likely to attribute errors that deviate significantly from a planned action to themselves and can therefore tune out the incorrect feedback (Pfordresher 2019). This is important for interpreting the neural correlates since studies have shown differences in ERPs elicited by self-generated errors compared to external manipulations (Waszak and Herwig 2007).

4.6.2 What ERP components were elicited

Preliminary EEG analyses showed that both singing and cello playing elicited a 3-component ERP complex during both compensating and ignoring incorrect pitch feedback. The first component was in line with our hypothesis and, given the nature of the task, likely reflects the P1 component that is typically observed in response to pitch feedback perturbation. The P1 response has been shown to accompany the detection of incorrect pitch feedback (Jones, Scheerer, and Tumber 2013; Scheerer and Jones 2018), but is insensitive to the magnitude of feedback perturbations. As such, it is thought to reflect a purely perceptual process (Jones, Scheerer, and Tumber 2013; Scheerer and Jones 2018). The observed similarity between singing and cello playing, and between compensating and ignoring is consistent with this interpretation. The second component we observed, which is similar in latency and polarity to an N200 response, may reflect the detection of a discrepancy between a sensory memory and sensory input (Waszak and Herwig 2007). This may be in contrast to the component we expected to see, the N100, which is thought to reflect a detected discrepancy between a sensory prediction and sensory feedback. Initially we had hypothesized that the pitch feedback perturbation would result in an N100 response based not only on the speech vocalization literature, but also on the premise that, during performance, musicians generate sensory predictions well in advance of their actions (Pfordresher and Palmer 2006, 2002; Mathias, Gehring, and Palmer 2017; Mathias et al. 2014). Our implicit assumption was that even after our participants had started singing/playing the correct target tone they would continue to generate a sensory prediction against which to compare incoming auditory feedback. However, in our experiment, participants knew that auditory feedback was unreliable and that they would likely have to compensate for or ignore a pitch perturbation part way through the trial. As such, they may have adopted a strategy of relying on a mnemonic representation of the tone as opposed to a sensory prediction. Such mnemonic sensory representations have previously been described for delayed visuomotor decision tasks (Constantinidis, Franowicz, and Goldman-Rakic 2001). If our participants were not relying on a sensory prediction to accomplish this task, then the new pitch feedback generated by the perturbation would have been compared to the sensory mnemonic, thereby eliciting the later negative response we observed.

The second component was followed by a positive deflection at ~300ms post perturbation. This is again slightly later than the P200 response that we had initially hypothesized would follow the N100 but is consistent with what we would expect to have follow an MMN. While the response we observed cannot be considered an MMN as it was not elicited by an oddball paradigm, it may still be instructive to consider the literature surrounding the MMN as the oddball paradigm is not entirely unlike the introduced pitch perturbation used in this experiment in that both reflect a sudden unexpected change in auditory feedback from a continuous state. Based on the Context Updating theory, the P300 (specifically the P300b) follows an MMN in cases where sensory feedback indicates a necessary update to the sensory prediction (Polich 2007). The P200 that we expected to see, conversely, follows from the N100 response and is thought to reflect the selection of a compensatory response to incorrect feedback (Liu and Larson 2007; Behroozmand, Liu, and Larson 2010) in cases where no context update is required (Polich 2007). In the case of our experiment, participants could correctly assume that, once the pitch perturbation was applied, their produced output would be 100 cents off from what they would expect based on their previous sensory prediction. As such, our finding that the third elicited component is a P300 (possibly P300b) is in line with the interpretation that participants switch their sensory context in response to the pitch feedback perturbation. A complementary interpretation for why our experiment may have elicited a P300 response is that the P300 is more likely to be elicited by external manipulations of sensory feedback (Nittono 2006; Waszak and Herwig 2007) whereas the P200 is more likely to be elicited by incorrect feedback resulting from a self-initiated action, as in the case of playing a wrong note during a piece of music (Ruiz et al. 2014; Ruiz, Jabusch, and Altenmüller 2009).

4.6.3 What this means

Taken together this preliminary evidence suggests that auditory-motor integration during singing and cello playing, presents similarities and differences from previous work done on vocalization. We propose that these differences are related to the differing sensory goals of matching a specific pitch in the case of music compared to producing a relatively pitch insensitive speech vocalization. Confirmation of this hypothesis could be achieved by direct comparison of the task used here together with a typical vocalization/perturbation task in the same participants. Our prediction would be that singing/playing would still elicit components that occur later, which require participants to rely on a sensory mnemonic, whereas we would expect to see an N1-P2 response for speaking. However, for the main purpose of this thesis, it is still an interesting finding that the responses are extremely similar for singing and cello playing. This is in line with our previous findings showing that cello playing recruits areas within the singing network in order to accomplish pitch regulation and extends these findings by showing that the time course of these responses is similar as well.

4.6.4 What we plan to do next

Further analysis is required in order to fully characterized the ERP response to the perturbed pitch feedback. Subsequent analyses will also look at the ERPs in response to trial onset and contrast them with those elicited by the introduced pitch feedback perturbation. Without comparing the ERPs from performance onset to those of pitch perturbation onset, it is difficult to say which components are deriving from the perturbation and which are related to the onset of performance itself. ERP analyses will also use the trials where no shift was introduced as a baseline against which to compare the responses to the pitch perturbations in both compensate and ignore conditions.

Source localization will be carried out for each of the characterized ERP components using sLoreta implemented in Brainstorm. Source localized components can then be compared with the fMRI findings from our previous study (Chapter 3). This comparison will lead to a more comprehensive understanding of the activation timecourse within the characterized brain network by showing which brain regions are specifically recruited during the auditory, motor, and auditory-motor integration phases of the pitch regulatory process

Finally, in addition to looking at the ERPs related to the pitch perturbation, we also intend to explore oscillatory synchronization in cello playing and singing, particularly within delta and theta bands as was used previously in Behroozmand (2015) to index the detection of vocal production errors (delta band) and sensory context updating (theta band) to determine future vocal actions. Our goal would be to continue testing the hypothesis that cello playing draws on the same neural mechanisms as singing, but also to assess more specifically the underlying mechanisms.

Chapter 5 - General Discussion

5.1 Overview

In this thesis we present three experiments that test a central hypothesis: that cello playing, a phylogenetically new cultural task, makes use of the phylogenetically old singing network of the brain in order to regulate pitch. In Experiment 1 we demonstrate that despite relying on completely distinct motor effectors and having very different evolutionary relevance, the brain areas recruited for cello playing directly overlap with areas in the singing network. In Experiment 2 we replicate and expand on this finding by showing that this overlap in recruited brain regions is specifically tied to auditory-motor integration that is necessary for pitch regulation when feedback perturbations are introduced. Finally, in Experiment 3 we show preliminary EEG evidence that the temporal dynamics of auditory-motor integration for both singing and cello playing are also similar. Taken together these findings expand our current understanding of the auditory-motor integration system - specifically the auditory-vocal system and the singing network - by showing that brain networks in place for vocal pitch regulation can be reused by tasks with learned, arbitrary auditory-motor associations like those required to play the cello.

5.2 **Recapitulation of results**

5.2.1 Experiment 1 – main results

In Experiment 1 we used fMRI to address the open question of whether or not playing a musical instrument completely independent of the vocal apparatus, but one that enabled continuous mapping between action and pitch output (as opposed to discrete mapping, such as on a keyboard instrument), would make use of brain areas within the singing network. In order to test this hypothesis, we used a specially constructed MR-compatible cello interface and had cellists play back individual target tones, or sing back target tones, and contrasted the resulting brain activity within individuals. We report three major findings. First, on a behavioral level, we showed that cellists were able to accurately play target tones on the MR-Compatible cello and were also able to accurately sing the target notes. This validates the use of the MR-compatible cello as a research tool and further validates the comparison of cello playing and singing in the same participants. Second, we replicated previous work done to characterize the singing network by showing that our single-note listen/sing results recruited motor (M1, dPMC, SMA) and auditory regions (STG), as well as auditory-motor integration regions (IPS, SMG) and subcortical regions (GP, Thalamus) consistent with those recruited for similar tasks in previous studies. Third, we found that cello playing also elicited brain activity in all of those regions, and that the brain activity for cello playing directly overlapped with the brain activity for singing. The only differences observed were that singing elicited additional activation in the brainstem, and that overall, cello playing elicited more, and more widespread, activation throughout auditory and motor regions of the brain.

5.2.2 Experiment 2 – main results

In Experiment 2, we expanded upon the previous results by introducing a sudden onset pitch perturbation to specifically elicit brain activity associated with feedback monitoring and error correction mechanisms important for auditory-motor integration. Participants were given two different instructions in different conditions – either to compensate for introduced pitch perturbations or to ignore them. These instructions were designed to mirror those used a in similar paradigm for singing (Zarate and Zatorre 2005, 2008). First, behaviorally we found that participants were equally able to compensate for or to ignore a sudden onset 100 cent pitch feedback perturbation on the cello and when singing individual target tones. Second, we replicated many of the fMRI findings from our first experiment showing that, when no pitch perturbation is introduced, cello playing and singing recruit overlapping brain areas in auditory and dorsal-motor regions. Third, when participants were instructed to compensate for an introduced pitch perturbation in order to go back to hearing themselves singing or playing the correct note, they recruited overlapping brain regions in areas specifically related to auditory-motor integration, namely the IPS and SMG. When asked to ignore the incorrect pitch feedback and continue playing the original target tone, participants again recruited some directly overlapping regions, this time in dorsal M1 and in posterior STG. Differences in recruited brain regions were also observed within M1 with cello playing recruiting hand/arm areas of M1 compared to singing, whereas singing showed more activation ventrally in larynx area of M1.

5.2.3 Experiment 3 – main results

In Experiment 3, we investigated the EEG timecourse and scalp distributions of either compensating for or ignoring pitch feedback perturbations. As in the previous two studies, expert cellists were asked to reproduce individual target tones by either singing or playing them on the cello. There was also a perturbed feedback condition similar to Experiment 2 in which cellists were instructed to compensate for or to ignore a sudden onset pitch feedback perturbation of 100 cents. First, we replicated our behavioural finding from Experiments 1 and 2 that cellists could sing and play back individual target tones, and our finding from Experiment 2 that they could also compensate for and ignore introduced pitch perturbations. Second, we identified a 3-component ERP complex differs in latency from the P1-N1-P2 complex reported in the literature for pitch regulation during speech vocalizations (Liu and Larson 2007; Behroozmand, Liu, and Larson 2010). Third, we found preliminary evidence that the time course of the brain response to perturbed pitch feedback is similar for singing and playing insofar as there were no significant differences between the evoked potentials in the two conditions.

5.3 **Interpretations**

5.3.1 Similar behaviour in singing and cello playing

Our first main finding is that our participants, who were all expert cellists, could sing, play back tones on the MR-Compatible cello, and play back tones on an electric cello with an equally high degree of pitch accuracy. Moreover, they were able to compensate for introduced pitch perturbations equally well when singing and playing the cello, and also to ignore them. While this finding was expected, it is also extremely important for interpreting the significance of any of the brain imaging results that follow. This finding validates the use of cello playing as a valid instrument against which to compare previous research done in singing, and the use of the MR-compatible cello to do so.

Although the cellists we tested could not be considered as professional singers, as compared to other studies that explicitly recruited such people (Zarate, Wood, and Zatorre 2010; Kleber et al. 2007), one could argue that their high accuracy in singing may be in part attributable to the fact that most of them had had some training in voice as part of their musical education. However, even people lacking any training have still been shown to perform well on a pitch matching task (Zarate and Zatorre 2005, 2008), and the general population has been shown to sing reasonably on pitch (within 50 cents, ie. a quarter tone) (Pfordresher and Brown 2017).

Sudden-onset pitch feedback perturbations have been used throughout the auditory-vocal literature as a way to demonstrate that auditory feedback is directly influencing motor production (Elman 1981; Houde and Jordan 1998; Burnett et al. 1998; Liu and Larson 2007; Toyomura et al. 2007). Without this perturbation, one could argue that auditory and motor regions are coincidently active but not necessarily interacting. Research done on perception-action coupling in piano playing has shown that auditory feedback is not necessarily monitored during production (Finney 1997; Pfordresher 2005), and that performance is not affected by a lack of auditory feedback , even for novice piano players (Pfordresher 2005). However, by showing that the pitch feedback perturbation elicits a proportional behavioural response, and that the resulting brain activity is directly correlated with the accuracy of the response, we have ruled out this potential alternative interpretation.

Previous studies done with singing have shown that people are able to compensate for pitch feedback perturbations even without vocal training, which is consistent with previous research done using non-speech vocalizations. A wealth of non-speech vocalization research has shown that auditory and motor perturbations during speech production result in compensatory movements (Houde and Jordan 1998; Burnett et al. 1998; Jones and Munhall 2000), solidifying the link between perception and action during vocal production. In contrast, ignoring pitch feedback perturbations during singing has been shown to rely on expertise (Zarate, Wood, and Zatorre 2010). In a study comparing expert singers to non-musicians, only expert singers were able to ignore pitch feedback perturbations of 25 cents, suggesting that compensating is the default action and that only through extensive training can the association between perception and production be decoupled. Our singing findings replicate those of the auditory-vocal literature (Zarate and Zatorre 2008; Zarate, Wood, and Zatorre 2010) across all 3 experiments, with expert cellists being able to fully ignore and compensate for pitch perturbations of 200 cents (Experiment 1) and 100 cents (Experiments 2,3).

We had initially hypothesized that cellists might show a reflexive compensation for the perturbation when singing during Experiments 2 and 3 because they were not expert singers. However, we did not find that this was the case. This is likely attributable to the magnitude of our selected pitch perturbation (100 cents), which was quite large compared to those that have been shown to elicit a reflexive compensation in previous studies. As previously mentioned, only expert singers were able to ignore a pitch feedback perturbation of 25 cents in a study comparing expert singers to non-singers(Zarate, Wood, and Zatorre 2010), and during speech vocalization people showed a compensatory response for F0 shifts of as little as 9 cents despite not perceiving perturbations below 15 cents (Scheerer and Jones 2018). The main conclusion of the work done on ignoring perturbed pitch feedback, is that the ability to ignore hinges on specific training of feedforward control mechanisms.

While we did not see a compensatory reflex, we can still speculate about what might differ between singing and cello playing if smaller pitch perturbations were to be used in future experiments. Our first prediction would be that expert cellists would be able to ignore smaller pitch shifts on the cello due to a highly developed forward model. If we were to compare expert cellists to novice cellists, we would predict that the experts would be able to ignore to a greater extent than the novices. However, we have no reason to predict that expert cellists would show an advantage when ignoring small perturbations during singing. Despite showing that similar neural circuitry is recruited by both instruments to accomplish pitch regulation, the M1 regions found to be active in our studies are largely distinct, which suggests separate representations and that training of the feedforward model in cello playing would be unlikely to generalize to the context of singing. This conclusion is supported by research showing enhanced interoception (awareness of internal states) in singers relative to string musicians and non-musicians (Schirmer-Mokwa et al. 2015). This point will be elaborated on in the section of this discussion related to the role of expertise (5.4.4)

5.3.2 Similar patterns of activation

The central neuroimaging finding of Experiment 1 and Experiment 2 presented in this thesis is that the brain regions involved in pitch regulation during cello playing are very similar, perhaps identical, to those used for pitch regulation in singing. Other studies show that the network involved is also very similar to that involved in pitch regulation in speech (Ozdemir, Norton, and Schlaug 2006; Guenther and Hickok 2016; Tourville and Guenther 2011). Our interpretation of this finding is that when we learn a new auditory-motor task (like cello playing), we make use of brain networks that are in place to accomplish a similar task. This general concept has been demonstrated previously in other domains; for example in the case of mathematical operations, where it has been proposed that structures in place for spatial processing may be recycled or co-opted because they require similar computations (Dehaene and Cohen 2007).

With fMRI, we showed that singing and cello playing recruit many of the same areas to accomplish three performance tasks: simple perception and reproduction, compensating for pitch perturbations, and ignoring them. During simple performance, cello playing and singing recruited a network of brain regions in motor and auditory cortices, and in the dorsal stream, including M1, STG, dPMC, and SMA/Pre-SMA. These regions have been shown to be recruited not only for singing across a number of studies (Kleber and Zarate 2014; Perry et al. 1999; Zarate and Zatorre 2005, 2008; Kleber et al. 2007) , but also for a variety of musical instrument performance tasks like playing the piano (Parsons et al. 2005; Lotze et al. 2003; R. M. Brown et al. 2012) and guitar (Buccino et al. 2004; Vogt, Higuchi, and Roberts 2007; Vogt et al. 2004). This suggests that the engagement of these structures is in no way unique to any one instrument, but rather that they act as domain-general substrates for accomplishing auditory-motor tasks.

Compensating for pitch perturbations additionally recruited the SMG and IPS. These brain regions have been shown to contribute to auditory-motor integration in a variety of contexts including singing (Zarate and Zatorre 2005, 2008), but also keyboard playing (R. M. Brown et al. 2012; Pfordresher et al. 2014) and guitar playing (Vogt, Higuchi, and Roberts 2007; Vogt et al. 2004; Buccino et al. 2004) so this finding is very much in line with previous research. These regions are likely serving multiple functions in the context of the pitch regulatory process from estimating the magnitude of the pitch perturbation to planning the appropriate corrective movement. Previous work has shown that one of the main roles of the IPS is to enable the recognition of transposed musical pitch patterns (Foster and Zatorre 2010), while other studies have pointed to an extensive role in calibrating reaching and grasping movements (Valyear et al. 2007; Culham and Valyear 2006). Work done on the topography of the IPL has shown that the middle and rostral portions of the IPL, which is where the SMG is located, are related to speech and, more generally, to motor planning (Binkofski, Klann, and Caspers 2016). We therefore propose that auditory feedback from STG is being sent to IPS, transformed from an auditory to motor signal, which is then planned in the SMG.

When ignoring pitch perturbations, we found that participants recruited STG and dPMC to a greater extent than just playing a note in the absence of any pitch perturbation. Here our findings deviate from what we expected to see based on the literature. Previous research using this task in singing showed that expert singers recruited STG and the putamen when ignoring perturbations compared to singing a note in the absence of a feedback manipulation, whereas non-singers recruited the left SMG and M1 (Zarate and Zatorre 2008). Our results fit somewhere between the two, perhaps because participants all had musical training, which could account for the increased auditory activity (STG), but not specifically vocal training, which could account for the recruitment of auditory-motor integration areas (in this case dPMC).

The dPMC has been shown to relate to both the formation of the link between movements and their associated sounds, and also to storing this auditory-motor association. An fMRI study done on cello learning found that the dPMC is one of the brain areas that was recruited following training, that is, after an association between action and sound had been established (Wollman et al. 2018), and a separate study found that rTMS of dPMC inhibited the formation of auditory-motor associations during learning in a piano context (Lega et al. 2016). In the case of our finding that dPMC is recruited to a greater extent when ignoring a pitch perturbation than when playing a note in the absence of any perturbation likely relates to both the retrieval and updating of an existing auditory-motor association. When ignoring, participants may have been holding a mnemonic representation of the target tone in memory for the first part of the trial, as has been described in the visuomotor domain for delayed decision tasks where sensory feedback cannot be used to predict future actions (Constantinidis, Franowicz, and Goldman-Rakic 2001). Once the perturbation was applied, participants may have continued holding the target tone in memory while also using incoming sensory information to re-map their sound-to-movement association in order to accurately predict the sensory consequences of future actions. While this interpretation may not generalize to self-initiated errors, for instance playing a wrong note during the performance of a musical piece, it makes sense in the context of this experiment. Participants were aware of the fact that auditory feedback was unreliable, thereby leading them to rely on a mnemonic representation as opposed to a sensory prediction, and also that the pitch perturbation, once applied, would be maintained for the duration of the trial, making it advantageous for them to update the movement to sound mapping in order to minimize the error of sensory predictions for the remainder of the trial. We would therefor expect to see a similar response if, for example, a string went out of tune during a cello performance and the cellist had to consistently play a semitone higher for the remainder of the piece.

The preliminary results of the EEG study show a robust, 3-component ERP response to pitch-perturbed feedback that is similar for both cello planning and singing. This result differs from our prediction that these tasks would elicit the P1-N1-P2 response seen in the literature for sudden onset pitch perturbations during vocalization. Nevertheless, the MMN and P3 have also been well characterized in the auditory-motor integration literature and merit discussion here. The component that was in line with our prediction was the P100. The P100 is thought to reflect a general error detection process due to the fact that it is elicited when participants perceive an error and that, unlike the N1 and P2, the magnitude of the P100 response is not linked to the magnitude of the introduced pitch perturbation nor is it linked to the associated behavioural compensation (Scheerer and Jones 2018). In light of this previous finding, it is not surprising that this ERP component was elicited during our task, and that the response is similar for both instruments.

The N100, conversely, is very closely linked to both the behavioural response and to the magnitude and direction of the introduced shift. The larger the pitch shift, the larger the N100 response (Liu and Larson 2007). This effect is enhanced in auditory-motor conditions compared to auditory only or motor only conditions, which implies that it has a role in linking auditory and motor information as opposed to simply in error detection. The N100 is thought to reflect a feed-forward signal pertaining to the anticipated sensory outcome of a motor pattern since it has been shown to increase in magnitude as a function of the difference between the produced and perceived output (Liu and Larson 2007). Moreover, this component is thought to relate to the attenuation of auditory cortex for self-initiated vocalizations since it is most robust when the difference is the result of a production error compared to an externally generated feedback manipulation (Waszak and Herwig 2007). The fact that the response we see occurs much later than a typical N100 response could be related to the fact that our manipulation was externally generated, but could also relate to the listen/play trial structure that we chose for this experiment.

What we see is a negative deflection around 200 ms after the introduction of the pitch perturbation. Unlike the N100, which indexes a comparison of a sensory prediction to a sensory outcome, the MMN is thought to index the comparison of a sensory memory to a sensory outcome (Waszak and Herwig 2007). Though what we present cannot be considered an MMN, which is tied to the use of the oddball paradigm, our experimental design may have elicited a similar response. In the work presented here, participants had to reproduce a specific pitch. Doing so likely involved holding the pre-shift pitch in memory for comparison after the shift was introduced. This finding can help us interpret some of the fMRI results as well. In a preceding paragraph we discussed the role of dPMC in maintaining a sensory memory and also in context updating. We proposed that our participants may be comparing incoming sensory information to a mnemonic representation of the target tone, as opposed to a sensory prediction of the target tone, since they were placed in a context where sensory feedback was not a reliable predictor of the sensory consequences of future actions. Here, the presence of a later negative deflection as opposed to a classic N100 supports this hypothesis.

For the third component, we predicted that we would see a P200 response, but what we saw instead was more in line with a P300 response. The P300 relates to both of the proposed functions of the dPMC mentioned above: context updating and memory. Research done on context updating has shown that a P300 response follows an MMN in cases where sensory feedback indicates that the current sensory prediction needs to be updated to accurately predict the future sensory consequences of actions (Polich 2007). The P200, conversely, has been shown to follow an N100 in cases where no context update is required (Polich 2007), and is more likely to be elicited by incorrect feedback caused by self initiated errors (Ruiz et al. 2014; Ruiz, Jabusch, and Altenmüller 2009) than by external feedback manipulations (Nittono 2006; Waszak and Herwig 2007). This interpretation is complementary to that of maintaining a sensory mnemonic in memory while selecting an appropriate motor response given that such a working memory task would likely also contribute to the observed P300 response (Polich 2007)

In addition to the components themselves being relevant to the pitch regulation process, our finding that the singing/cello playing response is similar for both the ignore and compensate conditions supports our central hypothesis that singing and cello playing not only make use of the same brain structures, but also that the temporal dynamics of the underlying pitch regulation process is the same as well.

5.3.3 Different patterns of activation

While we have mainly focused on what is common between singing and cello playing, we did also observe some differences between the two. The main differential finding was that singing and cello playing recruited distinct regions in primary motor cortex centered around larynx and hand areas, respectively. This difference, while expected, is important as it shows that our manipulation was sensitive enough to detect differences between the two instruments when they exist, thus allowing us to more confidently assert that the observed similarities reflect overlapping processes as opposed to simply a lack of spatial resolution (but see below).

We also observed more, and more extensive brain activation for cello playing relative to singing throughout auditory, motor, and auditory-motor regions of the brain. This difference is difficult to interpret as it could be related to a physical feature of the sound like timbre or loudness, despite our best efforts to match the cello playing to the singing. Some previous research has shown that participants show greater BOLD response in STG when playing their own instrument compared to another instrument (Gebel et al. 2013), but despite our participants not being expert singers, it is still difficult to argue that they have more experience with the cello than with their own voice.

5.4 **Distillation and integration**

5.4.1 Position relative to models of auditory motor integration

The present set of findings helps to unify findings from the musical instrument literature with those of the singing and auditory-vocal integration literature. Our results support previously established models of auditory-vocal integration and broaden the scope of their applicability. There are two main models that we have focused on throughout the three studies presented here. The first is the Diva Model (see: Figure 5-1) (Tourville and Guenther 2011), and the second is the Singing Network described in Kleber and Zarate (2014) (see: Figure 5-2).

The DIVA model is a comprehensive, computationally based model of auditory-motor integration for speech. While the model was designed to account for speech phenomena, the idea of a feedforward system that carries out actions based on an internal model of the desired outcome, and the feedback system that updates future movements accordingly, is also directly applicable non-speech auditory-motor tasks (singing, cello playing) as well. In the DIVA model, auditory and somatosensory feedback are compared to internal 'error maps' and 'target maps,' which are hypothesized to be located in pSTG, and SMG, respectively. In our research, we found evidence supporting this hypothesis, and evidence showing that the same feedback system exists for cello playing as well as for singing. When we introduced a sudden-onset pitch perturbation in Experiment 2 (chapter 3), we found that participants recruited pSTG and SMG when compensating for the pitch perturbation, and that greater recruitment of these areas was associated with better compensation. The fact that the same regions were engaged during error correction for the two instruments suggests that common internal representations and error maps encode high-level, abstract information as opposed to effector-specific sets of movements and sounds. These results also support theories of efference copies, where the predicted sensory consequences of a movement are sent forward to the relevant sensory systems. If pSTG and SMG are the locations where incoming auditory and somatosensory feedback are compared against this efference copy, then it could account for the direct overlap between the two instrument modalities.

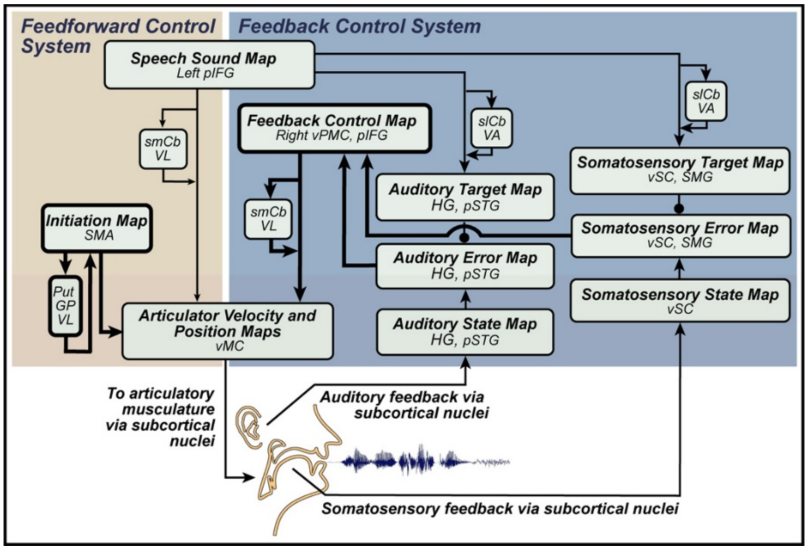

Figure 5-1 The DIVA Model | Diagram of the DIVA model of speech describes a control system account of auditory-motor integration (Tourville and Guenther 2011). Network encompasses brain regions related to movement execution (M1, PAG), movement planning and initiation (SMA, ACC), auditory feedback processing (STG), and auditory-motor integration (IPS, aINS).

Building on or complementing the DIVA model is the detailed description of the singing network put forward by Kleber and Zarate (2014). This work adds to the DIVA model by expanding on the brain structures associated with vocal pitch regulation. Speech, at least for non-tonal languages, does not require tightly controlled pitch. As a result, singing is a much better model to use for exploring the brain correlates of pitch regulation. Specifically, the singing network expands on the role of the IPS for sensorimotor transformation, and the aINS for mediating In this model, motor signals planned in dPMC and SMA are initiated by the ACC and ultimately sent to the vocal articulators by way of the PAG. The ACC and the PAG are both brain structures whose stimulation directly elicits vocalization in non-human primates, but the PAG, a midbrain nucleus, directly innervates the vocal folds, whereas the ACC is thought to relate to conscious control of movement initiation.

This distinction is important for interpreting our finding that singing and cello playing show overlapping activity in the ACC, but not in the brainstem or midbrain. It could be the case that the ACC plays a more domain-general role in initiating movements that produce sounds, whereas the PAG is related to the vocal fold movement itself and therefor has no role in cello playing. In the singing network, the cerebellum and basal ganglia (BG) provide modifying inputs over fine-grained features of the sound, a component that is shared with the DIVA model. The auditory and somatosensory consequences of these sounds are then processed by the STG and aINS, respectively, and any sensorimotor transformations required in order to update the forward model are mediated by the IPS, which serves as an auditory-motor interface. Our findings are largely consistent with this account of auditory-motor integration. When singing and when playing the cello participants recruited a network of brain regions encompassing motor (M1, SMA, ACC), auditory (STG), and auditory-motor integration (IPS) regions. Both tasks also recruited the cerebellum. We therefor propose that the singing network is applicable to more than just singing and can be used to accomplish pitch regulation in other contexts as well. We propose that the majority of the singing network is more of a pitch regulation network, and that singing happens to be one of the more common use cases of pitch regulation.

Part of our main hypothesis for Experiment 2 was that the aINS and BG might be more active for singing than for cello playing because of their pivotal roles in compensating for pitch perturbations (Zarate and Zatorre 2008; Kleber and Zarate 2014) and in pitch regulation with attenuated sensory feedback (Kleber et al. 2013, 2017). We were not able to provide conclusive evidence regarding the role of the BG or the aINS in cello playing since neither region was consistently recruited during either Experiment 1 or Experiment 2. However, based on the fact that the only regions shown to differ thus far are for primary motor control, we expect that in future studies designed to have a higher SNR in subcortical regions would show overlap in these regions as well.

Vocal or instrument performance in the real world rarely just involves one note. Most cases involve sequences of notes strung together into musical phrases. One of the frameworks for understanding musical sequences is that of Hierarchical Perception Action Coordination (H-PAC) proposed by Pfordresher (2019). This model proposes that an intention (comprising both memory and an internal model) leads to the development of a sensorimotor representation of a sequence. This sequence is then planned and executed, and auditory feedback is used to adjust the representation. This component of the model interacts with a 'metacognition' component that is hypothesized to influence error judgements based on 'self-agency.' In other words, that the extent to which someone attributes an error to themselves (i.e. a consequence of their actions rather than a consequence of an external agent) will influence how they respond to it. This model will be discussed in more depth below as part of our discussion of single notes vs. melodies (Section 5.4.3).

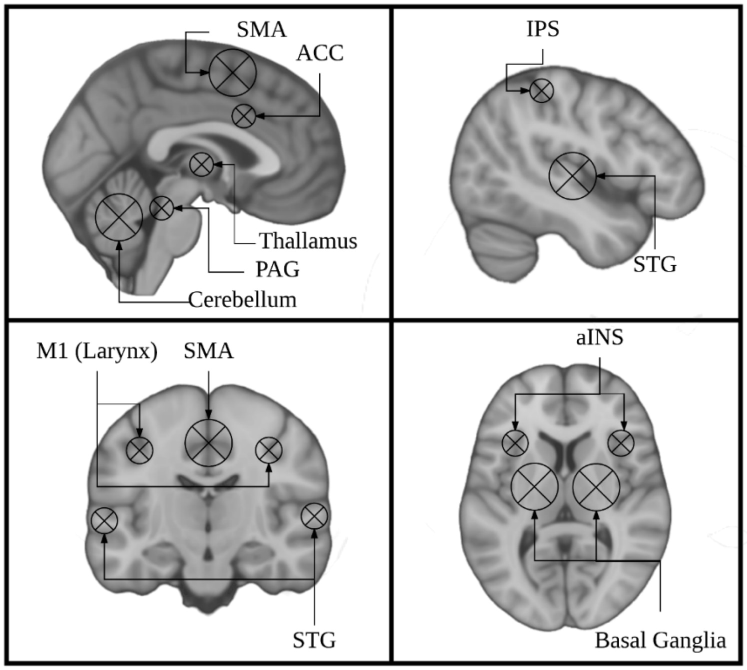

Figure 5-2 The Singing Network | Schematic representation of the singing network adapted from The Neuroscience of Singing (Kleber and Zarate 2014). Network encompasses brain regions related to movement execution (M1, PAG), movement planning and initiation (SMA, ACC), auditory feedback processing (STG), and auditory-motor integration (IPS, aINS).

Another framework in which to examine the role of feedback in action planning is Bayesian Decision Theory (Genewein and Braun 2012; Hahnloser and Narula 2017; Körding and Wolpert 2004). Bayesian decision theory has been used extensively to study sensorimotor adaptation in visually guided movements (Körding and Wolpert 2004; Yin et al. 2019) and may be useful for interpreting auditory-motor integration as well (Genewein and Braun 2012). Based on a Bayesian framework, actions are decided based on what is expected from previous experience (priors) and on feedback regarding the current success of recent predictions (likelihood). This statistical likelihood is generated from multiple sources of sensory feedback, and the priors are updated accordingly. For instance, in the case of cello playing, this feedback encompasses auditory, somatosensory, proprioceptive, and sometimes visual feedback whereas in singing, feedback is almost exclusively auditory and proprioceptive. The degree to which these various sources of feedback influence the priors is weighted according to their decision-making utility. This is in no way fully distinct from the H-PAC model, but rather proposes a possible rule set for weighting errors and updating the sensorimotor representation.

5.4.2 Continuous vs. Discrete Pitch Instruments

One of the central premises of the experiments presented in this thesis is that continuous pitch instruments like the cello bear a unique similarity to the human voice relative to keyed or fretted instruments like the piano, keyboard, or guitar. Below we will expand on the similarities and differences between these two classes of instruments. We will use Bayesian decision theory as a framework for interpreting the significance of these differences and how they may impact brain activity. For instance, when playing a continuous pitch instrument, or when singing, out of tune notes can (and should) be modified once they are produced whereas on the piano or guitar, no corrective action can be taken. Therefore, the utility of pitch feedback during performance is much higher for the continuous pitch instruments than for the discrete pitch instruments and the brain activity associated with pitch judgments (for instance in STG, IPS, SMG) is likely to be more prominent as a result.

Another difference between continuous and discrete pitch instruments is that movements on the cello must be much more precise than those on the piano or guitar – keys and frets are relatively large compared the space in which a note is considered correct on the neck of the cello. Moreover, the necessary degree of precision during cello playing changes as a function of where on the instrument you are whereas on a keyboard, 100 cents is always 1-key apart. On the cello, 100 cents can be a full centimeter apart close to the neck, as little as a few millimeters apart close to the bridge, or as much as a few strings and many centimeters away depending on the piece of music. If we again consider this in terms of utility, hand position in piano playing is almost as reliable an indicator of what pitch is produced as the pitch itself (Pfordresher 2005), whereas in cello playing pitch feedback is a more reliable index of which note you are playing than is hand position. It is also true that any note on the piano can be played with any of the fingers, and that the movement trajectory to reach those keys can vary a great deal depending on where the hand was previously and where it needs to go next. In that sense, continuous adjustments to the trajectory are necessary in the same way that continuous adjustments are necessary when reaching to push a button. However, 1 specific note only has 1 corresponding key, so the mapping between action and sound is fixed. On the cello, the same note can be played in multiple locations, with any of the fingers of the left hand, so there is a many to one mapping of action to sound. Moreover, both singing and unfretted instruments often use a 'just intonation system,' compared to keyed/fretted instruments that are 'equally tempered.' Using a just intonation system, different intervals can be defined by different frequency ratios depending on context. Conversely, with equally tempered instruments, there is a one to one correspondence between a musical interval and a specific frequency ratio. This would again lead to a much higher utility for pitch feedback in the case of continuous compared to discrete pitch instruments.

That being said, continuous and discrete pitch instruments still bear many similarities to one another in that they both produced timed, pitch sequences, and it is worth discussing how the findings from this study could inform our understanding of musical instrument production, and the relationship between musical instrument playing and singing. We found that many of the regions recruited for singing and cello playing in the studies presented here are in line with those reported previously for piano playing and guitar playing (R. M. Brown et al. 2012; Vogt, Higuchi, and Roberts 2007; Buccino et al. 2004; Vogt et al. 2004; Parsons et al. 2005; Bangert et al. 2006; Zatorre, Chen, and Penhune 2007). For instance, in Parsons 2005 study of piano playing, expert pianists recruited a network of brain regions including M1, ACC, SMA, ACC, IPL (Parsons et al. 2005)– all of which we found to have overlapping activation for singing and cello playing across Experiments 1 and 2. Based on this similarity, it is reasonable to suggest that the feedback mechanisms for keyed and continuous pitch instruments may not be all that different, and that the difference between them lies in the nature of the corrective motor action that needs to be planned. Research on action planning and its relationship to auditory feedback has been studied extensively in piano playing and may be instructive for exploring this interpretation. These studies have mostly looked at sequence production and will be discussed in the next section.

5.4.3 Single notes vs. Melodies

It is possible to argue that the listen/play single-note paradigm used for all three experiments bias our findings away from something that would generalize to melodies. Research on goal-directed movements has helped to characterize open-loop, and closed-loop mechanisms for error correction depending on the nature of the task and the feedback available (Gaveau et al. 2014). In this framework, tasks where feedback is informative for planning future actions are considered to be closed loop, whereas tasks where feedback is unavailable or cannot inform future actions are considered open loop. The single-note task may bias our results in the direction of the open loop component since, apart from executing a singular corrective action, there is no need to update the forward model to plan for future actions. However, despite this, there is evidence to support the claim that these results can generalize to sequence production as well. Research looking at the production of sequences of notes has found that regions similar to those we report are recruited (Parsons et al. 2005; Bangert et al. 2006; Kleber and Zarate 2014).

Much of the auditory-motor integration literature related to pitch feedback errors has focused on action sequencing and the role pitch errors play on planned future actions. For instance, research done on pitch errors during keyboard performance has shown that incorrect pitch feedback results in more pitch errors in the upcoming sequence (Pfordresher 2019; Pfordresher et al. 2014; Pfordresher 2003). Moreover, this research found that pitch errors representing either previously played notes, or notes that are part of the upcoming sequence, disrupted performance to a much greater extent than random notes. The significance that the authors attributed to this finding was that notes are not just being planned and executed as isolated events strung together, but rather that they are part of a higher-order organizational structure. Pitch errors are, therefore, processed within the context of this structure and not on an individual basis.

This interpretation raises interesting questions about how online pitch corrections during singing or during cello performance fit into the H-PAC model, especially when considering the vocal pitch regulatory reflex that has been studied in singing (Zarate, Wood, and Zatorre 2010). Some of the single note pitch correction findings presented here can help to extend the applicability of the H-PAC to better account for behaviours associated with playing a continuous pitch instrument. Our finding that singers and cellists perform just as well when either compensating or ignoring pitch perturbations is in line with the interpretation that music production is governed by higher-order mechanisms that are not linked to any one instrument or specific set of effectors. However, the model's proposal that errors are used primarily to update the sensorimotor representation is incomplete when considering rapid online adjustments during performance.

We propose two possible amendments that can extend the scope of this model. One is a control loop for rapidly updating ongoing actions without updating the higher order sensorimotor plan. This could account for behaviours like adding vibrato to mask a note that's a little bit out of tune, or sliding to the last note at the end of a long run. This loop would likely be highly contingent on the speed of musical performance. Based on previous work done in timing showing that timing errors have a detrimental effect on performance (Pfordresher 2003), any corrective actions would need to be taken very quickly and the time window should decrease as the temporal distance between subsequent notes decreases. For instance, if a performer knows that they are going to be holding a note for a few seconds they should be more likely to compensate for a pitch error than if they have many fast notes to move on to.

The second component that we propose to expand the H-PAC model to properly account for continuous pitch instruments is an additional feature in the metacognive model that keeps track of feedback errors and determines if an update to the sensorimotor plan is required, as was suggested in the book chapter developing this model (Pfordresher 2019). For instance, if several notes in a row are flat, then a performer should likely adjust their motor plan accordingly and play slightly higher on the neck (ie. towards the bridge) since the instrument has likely gone out of tune. This would interact with the self-agency module to determine if the incorrect feedback is coming from an external source, from a self-generated error, or from a change in the known auditory-to-sound map. A possible model for this will be discussed in Future Directions (Section 5.6.2)

One noteworthy feature of singing and cello playing that sets them apart from piano playing is the decoupling of pitch and timing. As has been pointed out in numerous studies, musical phrases require correctly timed pitched outcomes. On the piano hitting a key both initiates the sound and also determines the pitch. Conversely, in singing, pitch is determined by the tension of the vocal cords and sound initiation is determined by the diaphragm/glottis. Similarly, on cello, pitch is determined by the position of the left hand on the strings, and sound initiation is determined by the right (bowing) hand. The timing of notes following sound initiation in singing can be controlled either by the diaphragm/glottis, or by varying the vocal cord tension to transition between pitches on one breath. The same can be done on the cello, where timing can either result from bow movements or from pitch changes executed by the left hand during one long continuous bow. The right hand can also control pitch in a limited way by moving from string to string. While studies have shown that pitch errors and timing errors are largely independent during piano playing (Pfordresher 2003), we can also speculate that they may not be as distinct in continuous pitch instruments.

Regarding the possible role that these differences may have on brain activity in a cello playing task or a singing task compared to a keyboard task, we would predict that the IPS would be preferentially recruited for conditions where participants have to execute online corrections, whereas longer term adjustments to the auditory-motor map may be mediated by pre-motor and supplementary motor regions.

5.4.4 The role of expertise

Another factor that needs to be considered when interpreting the comparison between singing and cello playing is the role that expertise might play. Musical training is known to have an impact on the brain mechanisms that are recruited for auditory-motor integration tasks both in musical instrument playing and in singing (Kleber and Zarate 2014; Zatorre, Chen, and Penhune 2007; R. M. Brown, Zatorre, and Penhune 2015). For this reason, it is important to discuss the degree to which musical training may have influenced our findings, and also how our findings can contribute to the scientific community's understanding of the brain mechanisms underlying novel skill acquisition. Our participants, who were all expert cellists, were specifically chosen to not be expert singers. That is to say, they did not have formal classical vocal training like the participants in previous studies on expert vocal performance did (Kleber et al. 2007). Nevertheless, musical instrument training requires some degree of solfege (ear training), which typically has a vocal component. Because of this, it would not be entirely accurate to say that our participants were non-singers in the same way that someone lacking musical training altogether is a non-singer.

One could therefor argue that this vocal training biased our experiments towards finding shared brain mechanisms for singing and cello playing, and this overlap would not occur if our participants had no vocal training at all. However, we are confident that this is not the case. First, if the overlapping activation was related to vocal training, then we would predict that the people who did better on the singing task would also do better on the cello task. What we found was that there was no correlation between participants' performance accuracy when singing and when playing the cello. This is in line with one of our central interpretations of our findings, which is that forward control over the motor effectors is independent. Another point of evidence against the interpretation that the shared representation results from vocal training comes from a separate study where we looked at the brain correlates of learning to play the cello in participants that had no musical training whatsoever (Wollman et al. 2018). Participants were taught to play short 5-note melodies on the MR-Compatible cello and performed these sequences during fMRI after 1 week and after 4 weeks of training. We found that participants quickly formed an association between sounds and movements as indexed by co-activation of auditory and motor structures as well as the SMA after even a short amount of cello training. Several studies on instrument learning have also shown rapid formation of auditory-motor associates in dPMC, SMA, and STG, and none of these studies were contingent on vocal expertise (Wollman et al. 2018; Lega et al. 2016; Lappe et al. 2008; J. L. Chen, Rae, and Watkins 2012). As such, the overlapping brain regions recruited for singing and cello playing in the work presented here are unlikely to be attributable to any vocal expertise in our cellist participants and are more likely to reflect a domain general pitch regulatory mechanism.

We expected participants to be equally able to compensate for the perturbations in both singing/playing and this is what we observed across all three experiments. This finding is expected given that, in previous studies done on singing, even non-musicians were able to compensate for pitch feedback perturbations (Zarate and Zatorre 2005, 2008) with a comparable degree of accuracy to expert singers. However, because our participants were not expert signers, we had hypothesized that they might be less able to ignore the pitch perturbation during singing than cello playing because ignoring incorrect pitch feedback on the cello just requires the cellist to not move their hand once they have found the correct note whereas ignoring incorrect feedback during singing has been shown to involve supressing a pitch regulatory reflex, which only expert singers can do (Zarate, Wood, and Zatorre 2010). Research has shown that expert singers are able to ignore small (<25 cent) pitch perturbed feedback that non-singers compensate for reflexively (Zarate, Wood, and Zatorre 2010) but that even non-singers can ignore large (200 cent) pitch shifts.

This research suggests that the ability to ignore small pitch feedback perturbations is contingent on a highly developed forward model. It may also be related to an enhanced ability to rely on proprioceptive cues. Research on singing has shown that singers can rely on proprioceptive feedback to a greater extent than non-singers in the absence of auditory feedback (Kleber and Zarate 2014). Based on this, we predict that cellists may not have sufficient information to ignore very small pitch feedback perturbations since proprioceptive feedback is insufficient to know if a note is in tune. Research done on pitch regulation during cello performance (J. Chen et al. 2013, 2008) has found that, while cellists can initially rely on a purely forward model to shift between two notes in the absence of auditory feedback, they require ongoing auditory feedback in order to maintain pitch accuracy (J. Chen et al. 2013). The reason for this is that the vibrotactile feedback cellists receive from their instrument during performance is insufficient to derive pitch information and they must instead rely on periodic reference to physical landmarks like the neck or scroll to reset what has been characterized as a stochastic renewal process (J. Chen et al. 2013). Future research could directly address the hypothesis that smaller pitch feedback perturbations will be difficult to ignore during cello playing by looking at participants' ability to ignore pitch feedback perturbations ranging from very small (25 cents) to large (200 cents).

5.5 **Limitations**

5.5.1 Subvocalization

One possible explanation that has been suggested for the recruitment of areas within the singing network during our cello task is that participants may have been subvocalizing during cello playing. We were able to rule out overt vocalization by analysing the microphone recording to verify that participants were only playing the cello during cello performance and not humming for instance, however, sub-vocalization can be difficult to detect (Pruitt, Halpern, and Pfordresher 2019) and we cannot rule this out entirely. Future studies could address this potential confound by either measuring vocal cord activation during the cello playing task using EMG or by preventing participants from using the vocal system by asking them to hold their breath or to suck on a candy, both of which have been used in the speech-motor literature (Smith, Wilson, and Reisberg 1995).

It should also be noted that a common pedagogical technique when learning to play a musical instrument is to have students sing the melodies they are going to be playing before learning to play them on the instrument (Watkins and Scott 2012). In stringed instrument playing, it is also a common practice to have students explicitly form an association between bowing and breathing (Watkins and Scott 2012). Musical scores may even have breath marks written in to help guide bowing gestures and expressivity. While this pedagogical strategy would certainly bias our findings in the direction of activation overlap, it could also be the case that this pedagogical strategy developed because forming a link between gestures and vocalization is beneficial to musical skill acquisition. Future studies could address this issue by regulating the extent to which participants can use vocal strategies during learning and assessing the degree of learning success and the degree of overlap between the neural responses for a singing and instrument performance task. It would also be valuable to measure breathing concurrently with cello playing and singing in the scanner to determine if there is any correlation between the relative timing of breathing and tone production in the two domains.

5.5.2 fMRI Resolution

We cannot rule out the possibility that what we are interpreting as direct overlap between singing and cello playing is partially attributable to the spatial resolution limitations of fMRI, especially when using only a univariate GLM analysis approach and applying spatial smoothing. It is possible that these two instruments recruited brain areas that are extremely close to one another, but still distinct, or that there are intermixed but distinct neuronal populations representing each instrument within the same general brain region.

This limitation could be addressed a few different ways. One way would be to use a repetition suppression approach, as used for example by Brown et al. (2012) in a piano production study, which allows activity elicited by different conditions within the same region of interest to be dissociated. Future analyses could also address this limitation by using a multivoxel pattern analysis (MVPA) approach, which is known to be more sensitive to differing patterns of activation within one region. For instance, MVPA has previously been used to characterize some of the brain networks involved in goal-directed movements (Gallivan and Culham 2015).

5.6 Future Directions

5.6.1 Non-human models

Even using a multi-voxel pattern analysis approach, it is not possible to conclude that the same populations of neurons are handling both the vocal and non-vocal information in the same way. A large part of why the vocal system is so well understood is because of the large body of neurophysiological studies of non-human primates, and other vocalizing mammals like cats and bats. While it would most likely be impossible to train any non-human animal to play a cello (even a very special cello), it may be possible to train some on an auditory-motor association task that is independent of the vocal apparatus using, for instance, eye or hand movements to generate sounds with a continuous pitch distribution. Recording from populations of cells within the IPS could provide direct evidence that vocal and non-vocal tasks make use of the same cells for high-order auditory-motor processes.

5.6.2 Pitch regulation and sensorimotor planning

One of the components of the H-PAC model that remains to be characterized is the nature of the error control mechanism for online pitch corrections. In Section 5.4.1 we proposed that the process by which errors lead to updates in the sensorimotor plan may be governed by Bayesian decision theory as opposed to a thresholding process as was suggested by the author (Pfordresher 2019). This approach, which is common throughout the motor control and visuomotor integration literature, would better account for complexities like resolving multiple sources of conflicting feedback. For instance, let’s take the case of ignoring small pitch feedback perturbations. In this case, auditory feedback indicates that a note is incorrect but somatosensory feedback may still indicate that the note is correct. Based on Bayesian model, the degree to which a participant successfully ignores the incorrect feedback would depend on the strength of the prior, which is what was observed experimentally in the studies of experts compared to novices (Zarate and Zatorre 2008).

A few additional manipulations could help us test the theory that error correction behaviour during musical performance is best modelled as a Bayesian process. One such paradigm would be the gradual onset, sustained perturbation paradigms used to look at the effects of formant shifting on speech (Tremblay, Shiller, and Ostry 2003; Houde and Jordan 1998). In these paradigms auditory-to-motor mapping changes very slowly, and participants typically alter their produced output in a direction opposite to the introduced perturbation. For instance, after repeating the word ‘head’ (a middle formant /e/) under a perturbation condition that makes it sound more like ‘hid’ (a high formant /i/) participants will begin producing the word ‘had’ (a low formant /a/) which is in the opposite direction to the introduced perturbation. After producing sounds under these conditions, the perturbation is removed and an after effect is usually observed where participants continue using the adapted sound-to-movement map. Again, if we build a prediction for how our results would generalize to this context based on a Bayesian framework, accumulating evidence from auditory feedback that the position of the hand or the articulators is incorrect may elicit a compensatory response that the sudden onset pitch perturbation did not. We could then test the hypothesis that more evidence would be required during cello playing than during singing (participants would start compensating after a greater number of trials), and that cello playing would be less prone to an after effect.

Another perturbation paradigm that could be used is one that more closely parallels work done in reaching, but with the goal of producing a specific pitch as opposed to reaching a specific visually guided location. Participants could be asked to execute a large shift, the type heard in dramatic concertos, but the end target location could be systematically altered between the time they play the first note and the time they land on the end note. The same could be done in singing. We could then model the compensatory strategy and see if it is in line with what we would expect from a Bayesian process, which is to choose a strategy that minimizes the average error (Yin et al. 2019; Körding and Wolpert 2004).

In addition to helping characterize the nature of auditory-motor integration, these paradigms could also help inform the motor control literature. An ongoing debate in this literature is the extent to which such Bayesian models are effector independent. A recent study had participants learn to control a cursor on a screen with their finger or their wrist and, on some trials, applied a visual perturbation that made the cursor shift left or right (Yin et al. 2019). They found that participants required more trials to learn to compensate for the perturbation when using their fingers than when using their wrists to guide the cursor and hypothesized that the typical reaching movement led participants to overly rely on their prior learned relationship between finger movement and cursor location (Yin et al. 2019). Here we show data that even across two completely discrete effectors (hands, larynx) participants can compensate equally well. If it is the case that error compensation strategies are Bayesian in nature, then it would support the hypothesis that, with sufficient training, action plans are consolidated into a higher-order sensorimotor representation that is largely independent of the end effector. This interpretation is consistent with the H-PAC model, and with the work presented in this thesis.

Bibliography

Askenfelt, Anders, and Erik V. Jansson. 1992. "On Vibration Sensation and Finger Touch in Stringed Instrument Playing." *Music Perception: An Interdisciplinary Journal* 9 (3): 311–49. https://doi.org/10.2307/40285555.

Bangert, Marc, Thomas Peschel, Gottfried Schlaug, Michael Rotte, Dieter Drescher, Hermann Hinrichs, Hans-Jochen Heinze, and Eckart O. Altenmüller. 2006. "Shared Networks for Auditory and Motor Processing in Professional Pianists: Evidence from FMRI Conjunction." *NeuroImage* 30 (3): 917–26. https://doi.org/10.1016/j.neuroimage.2005.10.044.

Baumann, Simon, Susan Koeneke, Martin Meyer, Kai Lutz, and Lutz Jäncke. 2005. "A Network for Sensory-Motor Integration." *Annals of the New York Academy of Sciences* 1060 (1): 186–188. https://doi.org/10.1196/annals.1360.038.

Behroozmand, Roozbeh, Nadine Ibrahim, Oleg Korzyukov, Donald A. Robin, and Charles R. Larson. 2015. "Functional Role of Delta and Theta Band Oscillations for Auditory Feedback Processing during Vocal Pitch Motor Control." *Frontiers in Neuroscience* 9: 109. https://doi.org/10.3389/fnins.2015.00109.

Behroozmand, Roozbeh, and Charles R. Larson. 2011. "Error-Dependent Modulation of Speech-Induced Auditory Suppression for Pitch-Shifted Voice Feedback." *BMC Neuroscience* 12 (1): 54. https://doi.org/10.1186/1471-2202-12-54.

Behroozmand, Roozbeh, Hanjun Liu, and Charles R. Larson. 2010. "Time-Dependent Neural Processing of Auditory Feedback during Voice Pitch Error Detection." *Journal of Cognitive Neuroscience* 23 (5): 1205–17. https://doi.org/10.1162/jocn.2010.21447.

Belin, Pascal, Robert J. Zatorre, Rick Hoge, Alan C. Evans, and Bruce Pike. 1999. "Event-Related FMRI of the Auditory Cortex." *NeuroImage* 10 (4): 417–29. https://doi.org/10.1006/nimg.1999.0480.

Bendixen, Alexandra, Iria SanMiguel, and Erich Schröger. 2012. "Early Electrophysiological Indicators for Predictive Processing in Audition: A Review." *International Journal of Psychophysiology*, Predictive information processing in the brain: Principles, neural mechanisms and models, 83 (2): 120–31. https://doi.org/10.1016/j.ijpsycho.2011.08.003.

Berkowska, Magdalena, and Simone Dalla Bella. 2009. "Acquired and Congenital Disorders of Sung Performance: A Review." *Advances in Cognitive Psychology* 5 (November): 69–83. https://doi.org/10.2478/v10053-008-0068-2.

Binkofski, Ferdinand Christoph, Juliane Klann, and Svenja Caspers. 2016. "Chapter 4 - On the Neuroanatomy and Functional Role of the Inferior Parietal Lobule and Intraparietal Sulcus." In *Neurobiology of Language*, edited by Gregory Hickok and Steven L. Small, 35–47. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00004-3.

Brossier, Paul M. 2007. "Automatic Annotation of Musical Audio for Interactive Applications." Centre for Digital Music, Queen Mary University of London. https://aubio.org/phdthesis/.

Brown, Rachel M., Joyce L. Chen, Avrum Hollinger, Virginia B. Penhune, Caroline Palmer, and Robert J. Zatorre. 2012. "Repetition Suppression in Auditory–Motor Regions to Pitch and Temporal Structure in Music." *Journal of Cognitive Neuroscience* 25 (2): 313–28. https://doi.org/10.1162/jocn_a_00322.

Brown, Rachel M., Robert J. Zatorre, and Virginia B. Penhune. 2015. "Chapter 4 - Expert Music Performance: Cognitive, Neural, and Developmental Bases." In *Progress in Brain Research*, edited by Stanley Finger and François Boller Eckart Altenmüller, 217:57–86. Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies. Elsevier. http://www.sciencedirect.com/science/article/pii/S0079612314000223.

Brown, Steven, Michael J Martinez, Donald A Hodges, Peter T Fox, and Lawrence M Parsons. 2004. "The Song System of the Human Brain." *Cognitive Brain Research* 20 (3): 363–75. https://doi.org/10.1016/j.cogbrainres.2004.03.016.

Brown, Steven, Elton Ngan, and Mario Liotti. 2008. "A Larynx Area in the Human Motor Cortex." *Cerebral Cortex (New York, N.Y.: 1991)* 18 (4): 837–45. https://doi.org/10.1093/cercor/bhm131.

Buccino, Giovanni, Stefan Vogt, Afra Ritzl, Gereon R Fink, Karl Zilles, Hans-Joachim Freund, and Giacomo Rizzolatti. 2004. "Neural Circuits Underlying Imitation Learning of Hand Actions: An Event-Related FMRI Study." *Neuron* 42 (2): 323–34. https://doi.org/10.1016/S0896-6273(04)00181-3.

Burnett, Theresa A., Marcia B. Freedland, Charles R. Larson, and Timothy C. Hain. 1998. "Voice F0 Responses to Manipulations in Pitch Feedback." *The Journal of the Acoustical Society of America* 103 (6): 3153–61. https://doi.org/10.1121/1.423073.

Burnett, Theresa A., and Charles R. Larson. 2002. "Early Pitch-Shift Response Is Active in Both Steady and Dynamic Voice Pitch Control." *The Journal of the Acoustical Society of America* 112 (3): 1058–63. https://doi.org/10.1121/1.1487844.

Chen, Jessie, Marjorie H. Woollacott, Steven Pologe, and George P. Moore. 2008. "Pitch and Space Maps of Skilled Cellists: Accuracy, Variability, and Error Correction." *Experimental Brain Research* 188 (4): 493–503. https://doi.org/10.1007/s00221-008-1380-2.

Chen, Jessie, Marjorie Woollacott, Steve Pologe, and George P. Moore. 2013. "Stochastic Aspects of Motor Behavior and Their Dependence on Auditory Feedback in Experienced Cellists." *Frontiers in Human Neuroscience* 7 (July). https://doi.org/10.3389/fnhum.2013.00419.

Chen, Joyce L., Virginia B. Penhune, and Robert J. Zatorre. 2008. "Listening to Musical Rhythms Recruits Motor Regions of the Brain." *Cerebral Cortex* 18 (12): 2844–54. https://doi.org/10.1093/cercor/bhn042.

Chen, Joyce L., Charlotte Rae, and Kate E. Watkins. 2012. "Learning to Play a Melody: An FMRI Study Examining the Formation of Auditory-Motor Associations." *NeuroImage* 59 (2): 1200–1208. https://doi.org/10.1016/j.neuroimage.2011.08.012.

Chen, Joyce L., Robert J. Zatorre, and Virginia B. Penhune. 2006. "Interactions between Auditory and Dorsal Premotor Cortex during Synchronization to Musical Rhythms." *NeuroImage* 32 (4): 1771–81. https://doi.org/10.1016/j.neuroimage.2006.04.207.

Christoffels, Ingrid K., Elia Formisano, and Niels O. Schiller. 2007. "Neural Correlates of Verbal Feedback Processing: An FMRI Study Employing Overt Speech." *Human Brain Mapping* 28 (9): 868–79. https://doi.org/10.1002/hbm.20315.

Constantinidis, Christos, Matthew N. Franowicz, and Patricia S. Goldman-Rakic. 2001. "The Sensory Nature of Mnemonic Representation in the Primate Prefrontal Cortex." *Nature Neuroscience* 4 (3): 311. https://doi.org/10.1038/85179.

Culham, Jody C, and Kenneth F Valyear. 2006. "Human Parietal Cortex in Action." *Current Opinion in Neurobiology*, Cognitive neuroscience, 16 (2): 205–12. https://doi.org/10.1016/j.conb.2006.03.005.

Dalla Bella, Simone, Jean-François Giguère, and Isabelle Peretz. 2009. "Singing in Congenital Amusia." *The Journal of the Acoustical Society of America* 126 (1): 414–24. https://doi.org/10.1121/1.3132504.

Debener, Stefan, Markus Ullsperger, Markus Siegel, Katja Fiehler, D. Yves von Cramon, and Andreas K. Engel. 2005. "Trial-by-Trial Coupling of Concurrent Electroencephalogram and Functional Magnetic Resonance Imaging Identifies the Dynamics of Performance Monitoring." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 25 (50): 11730–37. https://doi.org/10.1523/JNEUROSCI.3286-05.2005.

Dehaene, Stanislas, and Laurent Cohen. 2007. "Cultural Recycling of Cortical Maps." *Neuron* 56 (2): 384–98. https://doi.org/10.1016/j.neuron.2007.10.004.

Dujardin, Eva, and Uwe Jürgens. 2005. "Afferents of Vocalization-Controlling Periaqueductal Regions in the Squirrel Monkey." *Brain Research* 1034 (1): 114–31. https://doi.org/10.1016/j.brainres.2004.11.048.

Eliades, Steven J., and Xiaoqin Wang. 2003. "Sensory-Motor Interaction in the Primate Auditory Cortex during Self-Initiated Vocalizations." *Journal of Neurophysiology* 89 (4): 2194–2207. https://doi.org/10.1152/jn.00627.2002.

———. 2008. "Neural Substrates of Vocalization Feedback Monitoring in Primate Auditory Cortex." *Nature* 453 (7198): 1102–6. https://doi.org/10.1038/nature06910.

———. 2012. "Neural Correlates of the Lombard Effect in Primate Auditory Cortex." *The Journal of Neuroscience* 32 (31): 10737–48. https://doi.org/10.1523/JNEUROSCI.3448-11.2012.

Elman, Jeffrey L. 1981. "Effects of Frequency-Shifted Feedback on the Pitch of Vocal Productions." *The Journal of the Acoustical Society of America* 70 (1): 45–50. https://doi.org/10.1121/1.386580.

Esposito, Antonio, Guy Demeurisse, Barbara Alberti, and Franco Fabbro. 1999. "Complete Mutism after Midbrain Periaqueductal Gray Lesion." *NeuroReport* 10 (4): 681–685.

Finney, Steven A. 1997. "Auditory Feedback and Musical Keyboard Performance." *Music Perception: An Interdisciplinary Journal* 15 (2): 153–74. https://doi.org/10.2307/40285747.

Foster, Nicholas E. V., and Robert J. Zatorre. 2010. "A Role for the Intraparietal Sulcus in Transforming Musical Pitch Information." *Cerebral Cortex* 20 (6): 1350–59. https://doi.org/10.1093/cercor/bhp199.

Gaab, Nadine, Christian Gaser, Tino Zaehle, Lutz Jancke, and Gottfried Schlaug. 2003. "Functional Anatomy of Pitch Memory—an FMRI Study with Sparse Temporal Sampling." *Neuroimage* 19 (4): 1417–1426.

Gallivan, Jason P, and Jody C Culham. 2015. "Neural Coding within Human Brain Areas Involved in Actions." *Current Opinion in Neurobiology*, Motor circuits and action, 33 (August): 141–49. https://doi.org/10.1016/j.conb.2015.03.012.

Gaveau, Valérie, Laure Pisella, Anne-Emmanuelle Priot, Takao Fukui, Yves Rossetti, Denis Pélisson, and Claude Prablanc. 2014. "Automatic Online Control of Motor Adjustments in Reaching and Grasping." *Neuropsychologia*, Special Issue in Honor of Marc Jeannerod, 55 (March): 25–40. https://doi.org/10.1016/j.neuropsychologia.2013.12.005.

Gebel, B., Ch. Braun, E. Kaza, E. Altenmüller, and M. Lotze. 2013. "Instrument Specific Brain Activation in Sensorimotor and Auditory Representation in Musicians." *NeuroImage* 74 (July): 37–44. https://doi.org/10.1016/j.neuroimage.2013.02.021.

Genewein, Tim, and Daniel A. Braun. 2012. "A Sensorimotor Paradigm for Bayesian Model Selection." *Frontiers in Human Neuroscience* 6 (October). https://doi.org/10.3389/fnhum.2012.00291.

Goebl, Werner, and Caroline Palmer. 2008. "Tactile Feedback and Timing Accuracy in Piano Performance." *Experimental Brain Research* 186 (3): 471–79. https://doi.org/10.1007/s00221-007-1252-1.

Gooler, David M., and William E. O'Neill. 1987. "Topographic Representation of Vocal Frequency Demonstrated by Microstimulation of Anterior Cingulate Cortex in the Echolocating Bat,Pteronotus Parnelli Parnelli." *Journal of Comparative Physiology A* 161 (2): 283–94. https://doi.org/10.1007/BF00615248.

Grabski, Krystyna, Laurent Lamalle, Coriandre Vilain, jean-luc Schwartz, Nathalie Vallée, Irène Troprès, Monica Baciu, Jean-François Le Bas, and Marc Sato. 2012. "Functional MRI Assessment of Orofacial Articulators: Neural Correlates of Lip, Jaw, Larynx, and Tongue Movements." *Human Brain Mapping* 33 (October): 2306–21. https://doi.org/10.1002/hbm.21363.

Grefkes, Christian, and Gereon R Fink. 2005. "The Functional Organization of the Intraparietal Sulcus in Humans and Monkeys." *Journal of Anatomy* 207 (1): 3–17. https://doi.org/10.1111/j.1469-7580.2005.00426.x.

Guenther, Frank H., and Gregory Hickok. 2016. "Chapter 58 - Neural Models of Motor Speech Control." In *Neurobiology of Language*, edited by Gregory Hickok and Steven L. Small, 725–40. San Diego: Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00058-4.

Hahnloser, Richard H. R., and Gagan Narula. 2017. "A Bayesian Account of Vocal Adaptation to Pitch-Shifted Auditory Feedback." *PLOS ONE* 12 (1): e0169795. https://doi.org/10.1371/journal.pone.0169795.

Haslinger, Bernhard, Peter Erhard, Eckart Altenmüller, Andreas Hennenlotter, Markus Schwaiger, Helga Gräfin von Einsiedel, Ernst Rummeny, Bastian Conrad, and Andrés O. Ceballos-Baumann. 2004. "Reduced Recruitment of Motor Association Areas during Bimanual Coordination in Concert Pianists." *Human Brain Mapping* 22 (3): 206–15. https://doi.org/10.1002/hbm.20028.

Heinks-Maldonado, Theda H., Daniel H. Mathalon, Max Gray, and Judith M. Ford. 2005. "Fine-Tuning of Auditory Cortex during Speech Production." *Psychophysiology* 42 (2): 180–90. https://doi.org/10.1111/j.1469-8986.2005.00272.x.

Herholz, Sibylle C., and Robert J. Zatorre. 2012. "Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure." *Neuron* 76 (3): 486–502. https://doi.org/10.1016/j.neuron.2012.10.011.

Hickok, Gregory, Bradley Buchsbaum, Colin Humphries, and Tugan Muftuler. 2003. "Auditory-Motor Interaction Revealed by FMRI: Speech, Music, and Working Memory in Area Spt." *Journal of Cognitive Neuroscience* 15 (5): 673–682.

Higuchi, Satomi, H. Holle, N. Roberts, S. B. Eickhoff, and Stefan Vogt. 2012. "Imitation and Observational Learning of Hand Actions: Prefrontal Involvement and Connectivity." *NeuroImage* 59 (2): 1668–83. https://doi.org/10.1016/j.neuroimage.2011.09.021.

Hollinger, Avrum D., and Marcelo M. Wanderley. 2013. "MRI-Compatible Optically-Sensed Cello." In *2013 IEEE Sensors*, 1–4. https://doi.org/10.1109/ICSENS.2013.6688614.

———. 2015. "The Design, Implementation, and Testing of a Sensorized MRI-Compatible Cello." *IEEE Sensors Journal* 15 (11): 6125–34. https://doi.org/10.1109/JSEN.2015.2449876.

Houde, John F., and Michael I. Jordan. 1998. "Sensorimotor Adaptation in Speech Production." *Science* 279 (5354): 1213–16. https://doi.org/10.1126/science.279.5354.1213.

Hutchins, Sean Michael, and Isabelle Peretz. 2012. “A Frog in Your Throat or in Your Ear? Searching for the Causes of Poor Singing.” *Journal of Experimental Psychology. General* 141 (1): 76–97. https://doi.org/10.1037/a0025064.

Hutchins, Sean, and Isabelle Peretz. 2012. “Amusics Can Imitate What They Cannot Discriminate.” *Brain and Language* 123 (3): 234–39. https://doi.org/10.1016/j.bandl.2012.09.011.

Hutchins, Sean, Jean Mary Zarate, Robert J. Zatorre, and Isabelle Peretz. 2010. “An Acoustical Study of Vocal Pitch Matching in Congenital Amusia.” *The Journal of the Acoustical Society of America* 127 (1): 504–12. https://doi.org/10.1121/1.3270391.

Hyde, Krista L., Jason P. Lerch, Robert J. Zatorre, Timothy D. Griffiths, Alan C. Evans, and Isabelle Peretz. 2007. “Cortical Thickness in Congenital Amusia: When Less Is Better Than More.” *Journal of Neuroscience* 27 (47): 13028–32. https://doi.org/10.1523/JNEUROSCI.3039-07.2007.

Hyde, Krista L., Robert J. Zatorre, and Isabelle Peretz. 2011. “Functional MRI Evidence of an Abnormal Neural Network for Pitch Processing in Congenital Amusia.” *Cerebral Cortex* 21 (2): 292–99. https://doi.org/10.1093/cercor/bhq094.

Jeffries, K J., Jonathan Fritz, and Allen Braun. 2003. “Words in Melody: An H215O PET Study of Brain Activation during Singing and Speaking.” *Neuroreport* 14 (April): 749–54. https://doi.org/10.1097/00001756-200304150-00018.

Jenkinson, Mark, Christian F. Beckmann, Timothy E. J. Behrens, Mark W. Woolrich, and Stephen M. Smith. 2012. “FSL.” *NeuroImage* 62 (2): 782–90. https://doi.org/10.1016/j.neuroimage.2011.09.015.

Jones, Jeffery A., and K. G. Munhall. 2000. “Perceptual Calibration of F0 Production: Evidence from Feedback Perturbation.” *The Journal of the Acoustical Society of America* 108 (3): 1246–51. https://doi.org/10.1121/1.1288414.

Jones, Jeffery A., Nichole Scheerer, and Anupreet Tumber. 2013. “The Relationship between Vocal Pitch Feedback Error and Event-Related Brain Potentials.” *Proceedings of Meetings on Acoustics* 19 (1): 060151. https://doi.org/10.1121/1.4800544.

Jürgens, U., and R. Pratt. 1979. “The Cingular Vocalization Pathway in the Squirrel Monkey.” *Experimental Brain Research* 34 (3): 499–510.

Jürgens, Uwe. 1976. “Projections from the Cortical Larynx Area in the Squirrel Monkey.” *Experimental Brain Research* 25 (4): 401–11. https://doi.org/10.1007/BF00241730.

———. 2002. “Neural Pathways Underlying Vocal Control.” *Neuroscience & Biobehavioral Reviews* 26 (2): 235–58. https://doi.org/10.1016/S0149-7634(01)00068-9.

———. 2009. “The Neural Control of Vocalization in Mammals: A Review.” *Journal of Voice* 23 (1): 1–10. https://doi.org/10.1016/j.jvoice.2007.07.005.

Jürgens, Uwe, and D. von Cramon. 1982. “On the Role of the Anterior Cingulate Cortex in Phonation: A Case Report.” *Brain and Language* 15 (2): 234–48. https://doi.org/10.1016/0093-934X(82)90058-X.

Jürgens, Uwe, and Detlev Ploog. 1970. “Cerebral Representation of Vocalization in the Squirrel Monkey.” *Experimental Brain Research* 10 (5): 532–54. https://doi.org/10.1007/BF00234269.

Jürgens, Uwe, and R. Pratt. 1979a. “The Cingular Vocalization Pathway in the Squirrel Monkey.” *Experimental Brain Research* 34 (3): 499–510.

———. 1979b. “Role of the Periaqueductal Grey in Vocal Expression of Emotion.” *Brain Research* 167 (2): 367–78.

Kirzinger, A., and U. Jürgens. 1982. “Cortical Lesion Effects and Vocalization in the Squirrel Monkey.” *Brain Research* 233 (2): 299–315. https://doi.org/10.1016/0006-8993(82)91204-5.

Kleber, Boris A., Niels Birbaumer, R. Veit, T. Trevorrow, and Martin Lotze. 2007. “Overt and Imagined Singing of an Italian Aria.” *NeuroImage* 36 (3): 889–900. https://doi.org/10.1016/j.neuroimage.2007.02.053.

Kleber, Boris A., Anders Friberg, Anthony Zeitouni, and Robert J. Zatorre. 2017. “Experience-Dependent Modulation of Right Anterior Insula and Sensorimotor Regions as a Function of Noise-Masked Auditory Feedback in Singers and Nonsingers.” *NeuroImage* 147 (February): 97–110. https://doi.org/10.1016/j.neuroimage.2016.11.059.

Kleber, Boris A., and Jean Mary Zarate. 2014. “The Neuroscience of Singing.” In *The Oxford Handbook of Singing*, edited by Graham F. Welch, David M. Howard, and John Nix, 257–76. Oxford University Press.

Kleber, Boris A., Anthony G. Zeitouni, Anders Friberg, and Robert J. Zatorre. 2013. “Experience-Dependent Modulation of Feedback Integration during Singing: Role of the Right Anterior Insula.” *Journal of Neuroscience* 33 (14): 6070–80. https://doi.org/10.1523/JNEUROSCI.4418-12.2013.

Körding, Konrad P., and Daniel M. Wolpert. 2004. “Bayesian Integration in Sensorimotor Learning.” *Nature* 427 (6971): 244. https://doi.org/10.1038/nature02169.

Kung, Shu-Jen, Joyce L. Chen, Robert J. Zatorre, and Virginia B. Penhune. 2012. “Interacting Cortical and Basal Ganglia Networks Underlying Finding and Tapping to the Musical Beat.” *Journal of Cognitive Neuroscience* 25 (3): 401–20. https://doi.org/10.1162/jocn_a_00325.

Lahav, Amir, Adam Boulanger, Gottfried Schlaug, and Elliot Saltzman. 2005. “The Power of Listening: Auditory-Motor Interactions in Musical Training.” *Annals of the New York Academy of Sciences* 1060 (1): 189–94. https://doi.org/10.1196/annals.1360.042.

Lappe, Claudia, Sibylle C. Herholz, Laurel J. Trainor, and Christo Pantev. 2008. “Cortical Plasticity Induced by Short-Term Unimodal and Multimodal Musical Training.” *The Journal of Neuroscience* 28 (39): 9632–39. https://doi.org/10.1523/JNEUROSCI.2254-08.2008.

Larson, Charles R. 1985. “The Midbrain Periaqueductal Gray.” *Journal of Speech, Language, and Hearing Research* 28 (2): 241–49. https://doi.org/10.1044/jshr.2802.241.

Larson, Charles R. 1991. “On the Relation of PAG Neurons to Laryngeal and Respiratory Muscles during Vocalization in the Monkey.” *Brain Research* 552 (1): 77–86. https://doi.org/10.1016/0006-8993(91)90662-F.

Lega, Carlotta, Marianne A. Stephan, Robert J. Zatorre, and Virginia B. Penhune. 2016. “Testing the Role of Dorsal Premotor Cortex in Auditory-Motor Association Learning Using Transcranical Magnetic Stimulation (TMS).” *PLOS ONE* 11 (9): e0163380. https://doi.org/10.1371/journal.pone.0163380.

Lehéricy, Stéphane, Habib Benali, Pierre-François Van de Moortele, Mélanie Pélégrini-Issac, Tobias Waechter, Kamil Ugurbil, and Julien Doyon. 2005. “Distinct Basal Ganglia Territories Are Engaged in Early and Advanced Motor Sequence Learning.” *Proceedings of the National Academy of Sciences* 102 (35): 12566–71. https://doi.org/10.1073/pnas.0502762102.

Liu, Hanjun, and Charles R. Larson. 2007. “Effects of Perturbation Magnitude and Voice F0 Level on the Pitch-Shift Reflex.” *The Journal of the Acoustical Society of America* 122 (6): 3671–77. https://doi.org/10.1121/1.2800254.

Lotze, Martin, Pedro Montoya, Michael Erb, Ernst Hülsmann, Herta Flor, Uwe Klose, Niels Birbaumer, and Wolfgang Grodd. 1999. "Activation of Cortical and Cerebellar Motor Areas during Executed and Imagined Hand Movements: An FMRI Study." *Journal of Cognitive Neuroscience* 11 (5): 491–501.

Lotze, Martin, Gabriela Scheler, Heng-Ru May Tan, Christoph H. Braun, and Niels P. Birbaumer. 2003. "The Musician's Brain: Functional Imaging of Amateurs and Professionals during Performance and Imagery." *NeuroImage* 20 (3): 1817–29. https://doi.org/10.1016/j.neuroimage.2003.07.018.

Loui, Psyche, David Alsop, and Gottfried Schlaug. 2009. "Tone-Deafness – a New Disconnection Syndrome?" *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* 29 (33): 10215–20. https://doi.org/10.1523/JNEUROSCI.1701-09.2009.

Luo, Cheng, Zhi-wei Guo, Yong-xiu Lai, Wei Liao, Qiang Liu, Keith M. Kendrick, De-zhong Yao, and Hong Li. 2012. "Musical Training Induces Functional Plasticity in Perceptual and Motor Networks: Insights from Resting-State FMRI." *PLOS ONE* 7 (5): e36568. https://doi.org/10.1371/journal.pone.0036568.

Luo, Jinhong, Steffen R. Hage, and Cynthia F. Moss. 2018. "The Lombard Effect: From Acoustics to Neural Mechanisms." *Trends in Neurosciences* 41 (12): 938–49. https://doi.org/10.1016/j.tins.2018.07.011.

Mandell, Jake, Katrin Schulze, and Gottfried Schlaug. 2007. "Congenital Amusia: An Auditory-Motor Feedback Disorder?" *Restorative Neurology and Neuroscience* 25 (3–4): 323–34.

Mathias, Brian, William J. Gehring, and Caroline Palmer. 2017. "Auditory N1 Reveals Planning and Monitoring Processes during Music Performance." *Psychophysiology* 54 (2): 235–47. https://doi.org/10.1111/psyp.12781.

———. 2019. "Electrical Brain Responses Reveal Sequential Constraints on Planning during Music Performance." *Brain Sciences* 9 (2): 25. https://doi.org/10.3390/brainsci9020025.

Mathias, Brian, Caroline Palmer, Fabien Perrin, and Barbara Tillmann. 2014. "Sensorimotor Learning Enhances Expectations During Auditory Perception." *Cerebral Cortex*, March, bhu030. https://doi.org/10.1093/cercor/bhu030.

Metzner, Walter. 1996. "Anatomical Basis for Audio-Vocal Integration in Echolocating Horseshoe Bats." *Journal of Comparative Neurology* 368 (2): 252–69. https://doi.org/10.1002/(SICI)1096-9861(19960429)368:2<252::AID-CNE6>3.0.CO;2-2.

Moss, Cynthia F, and Shiva R Sinha. 2003. "Neurobiology of Echolocation in Bats." *Current Opinion in Neurobiology* 13 (6): 751–58. https://doi.org/10.1016/j.conb.2003.10.016.

Müller-Preuss, Peter, and Detlev W. Ploog. 1981. "Inhibition of Auditory Cortical Neurons during Phonation." *Brain Research* 215 (1): 61–76. https://doi.org/10.1016/0006-8993(81)90491-1.

Nichols, Thomas, Matthew Brett, Jesper Andersson, Tor Wager, and Jean-Baptiste Poline. 2005. "Valid Conjunction Inference with the Minimum Statistic." *NeuroImage* 25 (3): 653–60. https://doi.org/10.1016/j.neuroimage.2004.12.005.

Nittono, Hiroshi. 2006. "Voluntary Stimulus Production Enhances Deviance Processing in the Brain." *International Journal of Psychophysiology*, From Laboratory to Real World: Challenging Topics in Psychophysiology Today, 59 (1): 15–21. https://doi.org/10.1016/j.ijpsycho.2005.06.008.

Ozdemir, Elif, Andrea Norton, and Gottfried Schlaug. 2006. “Shared and Distinct Neural Correlates of Singing and Speaking.” *NeuroImage* 33 (December): 628–35. https://doi.org/10.1016/j.neuroimage.2006.07.013.

Parlitz, Dietrich, and Marc Bangert. 1999. “Short and Medium Motor Responses to Auditory Pitch Shift: Latency Measurements of the Professional Musician’s Audio-motor Loop for Intonation.” *The Journal of the Acoustical Society of America* 105 (2): 1298–1298. https://doi.org/10.1121/1.426186.

Parsons, Lawrence M., Justine Sergent, Donald A. Hodges, and Peter T. Fox. 2005. “The Brain Basis of Piano Performance.” *Neuropsychologia* 43 (2): 199–215. https://doi.org/10.1016/j.neuropsychologia.2004.11.007.

Penhune, Virginia B. 2011. “Sensitive Periods in Human Development: Evidence from Musical Training.” *Cortex* 47 (9): 1126–37. https://doi.org/10.1016/j.cortex.2011.05.010.

Penhune, Virginia B., and Christopher J. Steele. 2012. “Parallel Contributions of Cerebellar, Striatal and M1 Mechanisms to Motor Sequence Learning.” *Behavioural Brain Research* 226 (2): 579–91. https://doi.org/10.1016/j.bbr.2011.09.044.

Peretz, Isabelle, Elvira Brattico, and Mari Tervaniemi. 2005. “Abnormal Electrical Brain Responses to Pitch in Congenital Amusia.” *Annals of Neurology* 58 (3): 478–82. https://doi.org/10.1002/ana.20606.

Peretz, Isabelle, and Krista L. Hyde. 2003. “What Is Specific to Music Processing? Insights from Congenital Amusia.” *Trends in Cognitive Sciences* 7 (8): 362–67. https://doi.org/10.1016/S1364-6613(03)00150-5.

Perry, David W., Robert J. Zatorre, Michael Petrides, Bessie Alivisatos, Ernst Meyer, and Alan C. Evans. 1999. “Localization of Cerebral Activity during Simple Singing.” *NeuroReport* 10 (18): 3979–3984.

Pfordresher, Peter Q. 2003. “Auditory Feedback in Music Performance: Evidence for a Dissociation of Sequencing and Timing.” *Journal of Experimental Psychology: Human Perception and Performance* 29 (5): 949–64. https://doi.org/10.1037/0096-1523.29.5.949.

———. 2005. “Auditory Feedback in Music Performance: The Role of Melodic Structure and Musical Skill.” *Journal of Experimental Psychology. Human Perception and Performance* 31 (6): 1331–45. https://doi.org/10.1037/0096-1523.31.6.1331.

———. 2019. *Sound and Action in Music Performance*. Academic Press.

Pfordresher, Peter Q., and Steven Brown. 2007. “Poor-Pitch Singing in the Absence of ‘Tone Deafness.’” *Music Perception: An Interdisciplinary Journal* 25 (2): 95–115. https://doi.org/10.1525/mp.2007.25.2.95.

———. 2017. “Vocal Mistuning Reveals the Origin of Musical Scales.” *Journal of Cognitive Psychology* 29 (1): 35–52. https://doi.org/10.1080/20445911.2015.1132024.

Pfordresher, Peter Q., and James T. Mantell. 2014. “Singing with Yourself: Evidence for an Inverse Modeling Account of Poor-Pitch Singing.” *Cognitive Psychology* 70 (May): 31–57. https://doi.org/10.1016/j.cogpsych.2013.12.005.

Pfordresher, Peter Q., James T. Mantell, Steven Brown, Robert Zivadinov, and Jennifer L. Cox. 2014. “Brain Responses to Altered Auditory Feedback during Musical Keyboard Production: An FMRI Study.” *Brain Research* 1556 (March): 28–37. https://doi.org/10.1016/j.brainres.2014.02.004.

Pfordresher, Peter Q., and C. Palmer. 2002. “Effects of Delayed Auditory Feedback on Timing of Music Performance.” *Psychological Research* 66 (1): 71–79. https://doi.org/10.1007/s004260100075.

Pfordresher, Peter Q., and Caroline Palmer. 2006. “Effects of Hearing the Past, Present, or Future during Music Performance.” *Perception & Psychophysics* 68 (3): 362–76. https://doi.org/10.3758/BF03193683.

Polich, John. 2007. “Updating P300: An Integrative Theory of P3a and P3b.” *Clinical Neurophysiology* 118 (10): 2128–48. https://doi.org/10.1016/j.clinph.2007.04.019.

Pruitt, Tim A., Andrea R. Halpern, and Peter Q. Pfordresher. 2019. “Covert Singing in Anticipatory Auditory Imagery.” *Psychophysiology* 56 (3): e13297. https://doi.org/10.1111/psyp.13297.

Rauschecker, Josef P. 2011. “An Expanded Role for the Dorsal Auditory Pathway in Sensorimotor Control and Integration.” *Hearing Research* 271 (1–2): 16–25. https://doi.org/10.1016/j.heares.2010.09.001.

Reznik, Daniel, Yael Henkin, Noa Schadel, and Roy Mukamel. 2014. “Lateralized Enhancement of Auditory Cortex Activity and Increased Sensitivity to Self-Generated Sounds.” *Nature Communications* 5 (June): 4059. https://doi.org/10.1038/ncomms5059.

Rubens, A. B. 1975. “Aphasia with Infarction in the Territory of the Anterior Cerebral Artery.” *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior* 11 (3): 239–50.

Ruiz, María Herrojo, Julius Huebl, Thomas Schönecker, Andreas Kupsch, Kielan Yarrow, Joachim K. Krauss, Gerd-Helge Schneider, and Andrea A. Kühn. 2014. “Involvement of Human Internal Globus Pallidus in the Early Modulation of Cortical Error-Related Activity.” *Cerebral Cortex (New York, N.Y.: 1991)* 24 (6): 1502–17. https://doi.org/10.1093/cercor/bht002.

Ruiz, María Herrojo, Hans-Christian Jabusch, and Eckart O. Altenmüller. 2009. “Detecting Wrong Notes in Advance: Neuronal Correlates of Error Monitoring in Pianists.” *Cerebral Cortex* 19 (11): 2625–39. https://doi.org/10.1093/cercor/bhp021.

Sanmiguel, Iria, Juanita Todd, and Erich Schröger. 2013. “Sensory Suppression Effects to Self-Initiated Sounds Reflect the Attenuation of the Unspecific N1 Component of the Auditory ERP.” *Psychophysiology* 50 (4): 334–43. https://doi.org/10.1111/psyp.12024.

Scheerer, Nichole E., J. Behich, Hanjun Liu, and Jeffery A. Jones. 2013. “ERP Correlates of the Magnitude of Pitch Errors Detected in the Human Voice.” *Neuroscience* 240 (June): 176–85. https://doi.org/10.1016/j.neuroscience.2013.02.054.

Scheerer, Nichole E., and Jeffery A. Jones. 2014. “The Predictability of Frequency-Altered Auditory Feedback Changes the Weighting of Feedback and Feedforward Input for Speech Motor Control.” *European Journal of Neuroscience* 40 (12): 3793–3806. https://doi.org/10.1111/ejn.12734.

———. 2018. “Detecting Our Own Vocal Errors: An Event-Related Study of the Thresholds for Perceiving and Compensating for Vocal Pitch Errors.” *Neuropsychologia* 114 (June): 158–67. https://doi.org/10.1016/j.neuropsychologia.2017.12.007.

Schirmer-Mokwa, Katharina L., Pouyan R. Fard, Anna M. Zamorano, Sebastian Finkel, Niels Birbaumer, and Boris A. Kleber. 2015. “Evidence for Enhanced Interoceptive Accuracy in Professional Musicians.” *Frontiers in Behavioral Neuroscience* 9 (December). https://doi.org/10.3389/fnbeh.2015.00349.

Schuller, G., and S. Radtke-Schuller. 1990. “Neural Control of Vocalization in Bats: Mapping of Brainstem Areas with Electrical Microstimulation Eliciting Species-Specific Echolocation Calls in the Rufous Horseshoe Bat.” *Experimental Brain Research* 79 (1): 192–206. https://doi.org/10.1007/BF00228889.

Schulz, G. M., M. Varga, K. Jeffires, C. L. Ludlow, and A. R. Braun. 2005. "Functional Neuroanatomy of Human Vocalization: An H215O PET Study." *Cerebral Cortex (New York, N.Y.: 1991)* 15 (12): 1835–47. https://doi.org/10.1093/cercor/bhi061.

Segado, Melanie, Avrum D. Hollinger, Joseph Thibodeau, Virginia B. Penhune, and Robert J. Zatorre. 2018. "Partially Overlapping Brain Networks for Singing and Cello Playing." *Frontiers in Neuroscience* 12 (May). https://doi.org/10.3389/fnins.2018.00351.

Seidler, R. D., D. C. Noll, and P. Chintalapati. 2006. "Bilateral Basal Ganglia Activation Associated with Sensorimotor Adaptation." *Experimental Brain Research* 175 (3): 544–55. https://doi.org/10.1007/s00221-006-0571-y.

Smith, J. David, Margaret Wilson, and Daniel Reisberg. 1995. "The Role of Subvocalization in Auditory Imagery." *Neuropsychologia*, The Neuropsychology of Mental Imagery, 33 (11): 1433–54. https://doi.org/10.1016/0028-3932(95)00074-D.

Smotherman, Michael S. 2007. "Sensory Feedback Control of Mammalian Vocalizations." *Behavioural Brain Research*, Mammalian Vocalization: Neural, Behavioural, and Environmental Determinants, 182 (2): 315–26. https://doi.org/10.1016/j.bbr.2007.03.008.

Stephan, Marianne A., Carlotta Lega, and Virginia B. Penhune. 2018. "Auditory Prediction Cues Motor Preparation in the Absence of Movements." *NeuroImage* 174 (July): 288–96. https://doi.org/10.1016/j.neuroimage.2018.03.044.

Sutton, D., R. E. Trachy, and R. C. Lindeman. 1981. "Primate Phonation: Unilateral and Bilateral Cingulate Lesion Effects." *Behavioural Brain Research* 3 (1): 99–114. https://doi.org/10.1016/0166-4328(81)90031-0.

Tonkinson, Steven. 1994. "The Lombard Effect in Choral Singing." *Journal of Voice* 8 (1): 24–29. https://doi.org/10.1016/S0892-1997(05)80316-9.

Tourville, Jason A., and Frank H. Guenther. 2011. "The DIVA Model: A Neural Theory of Speech Acquisition and Production." *Language and Cognitive Processes* 26 (7): 952–81. https://doi.org/10.1080/01690960903498424.

Tourville, Jason A., Kevin J. Reilly, and Frank H. Guenther. 2008. "Neural Mechanisms Underlying Auditory Feedback Control of Speech." *NeuroImage* 39 (3): 1429–43. https://doi.org/10.1016/j.neuroimage.2007.09.054.

Toyomura, A., S. Koyama, T. Miyamaoto, A. Terao, T. Omori, H. Murohashi, and S. Kuriki. 2007. "Neural Correlates of Auditory Feedback Control in Human." *Neuroscience* 146 (2): 499–503.

Trachy, Ronald E., Dwight Sutton, and Roger C. Lindeman. 1981. "Primate Phonation: Anterior Cingulate Lesion Effects on Response Rate and Acoustical Structure." *American Journal of Primatology* 1 (1): 43–55. https://doi.org/10.1002/ajp.1350010106.

Tremblay, Stéphanie, Douglas M. Shiller, and David J. Ostry. 2003. "Somatosensory Basis of Speech Production." *Nature* 423 (6942): 866–69. https://doi.org/10.1038/nature01710.

Tremblay-Champoux, Alexandra, Simone Dalla Bella, Jessica Phillips-Silver, Marie-Andrée Lebrun, and Isabelle Peretz. 2010. "Singing Proficiency in Congenital Amusia: Imitation Helps." *Cognitive Neuropsychology* 27 (6): 463–76. https://doi.org/10.1080/02643294.2011.567258.

Turner, Benjamin O., Erick J. Paul, Michael B. Miller, and Aron K. Barbey. 2018. "Small Sample Sizes Reduce the Replicability of Task-Based FMRI Studies." *Communications Biology* 1 (1): 62. https://doi.org/10.1038/s42003-018-0073-z.

Valyear, Kenneth F., Cristiana Cavina-Pratesi, Amanda J. Stiglick, and Jody C. Culham. 2007. "Does Tool-Related FMRI Activity within the Intraparietal Sulcus Reflect the Plan to Grasp?" *Neuroimage* 36: T94–T108.

Vogt, Stefan, Giovanni Buccino, Afra Ritzl, Gereon Fink, Karl Zilles, Hans-Joachim Freund, and Giacomo Rizzolatti. 2004. "Neural Circuits Underlying Imitation of Novel Hand Actions: An Event-Related FMRI Study: TH 165." *Neuroimage* 22 (June). insights.ovid.com.

Vogt, Stefan, Satomi Higuchi, and N. Roberts. 2007. "Neural Circuits Involved in Physical and Observational Practice of Hand Actions: A Common Pathway to Skilled Performance?" *Journal of Sport & Exercise Psychology* 29 (July): S12–S12.

Waszak, Florian, and Arvid Herwig. 2007. "Effect Anticipation Modulates Deviance Processing in the Brain." *Brain Research* 1183 (December): 74–82. https://doi.org/10.1016/j.brainres.2007.08.082.

Watanabe, Takeshi, and Yasuji Katsuki. 1974. "Response Pattern of Single Auditory Neurons of the Cat to Species Specific Vocalization." *The Japanese Journal of Physiology* 24 (2): 135–55. https://doi.org/10.2170/jjphysiol.24.135.

Watkins, Cornelia, and Laurie Scott. 2012. *From the Stage to the Studio: How Fine Musicians Become Great Teachers*. Oxford University Press.

Welch, Graham F. 1979. "Poor Pitch Singing: A Review of the Literature." *Psychology of Music* 7 (1): 50–58. https://doi.org/10.1177/030573567971006.

Wollman, Indiana, Virginia Penhune, Melanie Segado, Thibaut Carpentier, and Robert J. Zatorre. 2018. "Neural Network Retuning and Neural Predictors of Learning Success Associated with Cello Training." *Proceedings of the National Academy of Sciences* 115 (26): E6056–64. https://doi.org/10.1073/pnas.1721414115.

Yin, Cong, Huijun Wang, Kunlin Wei, and Konrad Paul Kording. 2019. "Sensorimotor Priors Are Effector-Dependent." *Journal of Neurophysiology*, May. https://doi.org/10.1152/jn.00228.2018.

Yousry, T. A., U. D. Schmid, H. Alkadhi, D. Schmidt, A. Peraud, A. Buettner, and P. Winkler. 1997. "Localization of the Motor Hand Area to a Knob on the Precentral Gyrus. A New Landmark." *Brain* 120 (1): 141–157.

Zarate, Jean Mary, Sean Wood, and Robert J. Zatorre. 2010. "Neural Networks Involved in Voluntary and Involuntary Vocal Pitch Regulation in Experienced Singers." *Neuropsychologia* 48 (2): 607–18. https://doi.org/10.1016/j.neuropsychologia.2009.10.025.

Zarate, Jean Mary, and Robert J. Zatorre. 2005. "Neural Substrates Governing Audiovocal Integration for Vocal Pitch Regulation in Singing." *Annals of the New York Academy of Sciences* 1060 (1): 404–408.

———. 2008. "Experience-Dependent Neural Substrates Involved in Vocal Pitch Regulation during Singing." *NeuroImage* 40 (4): 1871–87. https://doi.org/10.1016/j.neuroimage.2008.01.026.

Zatorre, Robert J., Joyce L. Chen, and Virginia B. Penhune. 2007. "When the Brain Plays Music: Auditory–Motor Interactions in Music Perception and Production." *Nature Reviews Neuroscience* 8 (7): 547–558.