Title
Mechanisms of Online Control of Speech

Author
Raharjo, Inez

Publication Date
2020

Peer reviewed|Thesis/dissertation

Mechanisms of Online Control of Speech

by

Inez Raharjo

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

DocuSigned by:
Srikantan Nagarajan
4C21BC1B9EB145E...

Srikantan Nagarajan

Chair

DocuSigned by:
John Houde

John Houde

DocuSigned by:
Christoph Schreiner
[illegible]

Christoph Schreiner, MD, PhD

DocuSigned by:
Michael Yartsev
91DDAD7670674EE...

Michael Yartsev

Committee Members

Acknowledgements

I am in disbelief that I ever get to this stage of writing my acknowledgement for my PhD thesis, especially during this time of world pandemic. Wow! This is truly something that I was only able to achieve through the grace of God and support from many people in my life.

Thank you, mom, dad, and oma for always loving and praying for me, and making sacrifices so that I was able to study far away from home. Thank you, Uncle Didi, for inspiring me to get a PhD (I clearly remember my grandparents showing me a portrait of him receiving his PhD and they were so proud) and helping get to the US in the first place. Thank you, papi, for always being supportive. And opa, hope you are proud of me from heaven above.

I am very grateful to have ended up working in Dr. Srikantan Nagarajan and Dr. John Houde's labs. They are always so supportive and understanding, yet able to inspire me to be a better researcher and an independent thinker. I am also thankful for Dr. Lydia Sohn and Dr. Chris Diederich for giving me a chance to explore my research curiosities during my lab rotations. The students in their lab at the time also made the experience enjoyable and educative.

Surviving my day-to-day work as graduate student wouldn't have been possible without the support from very loving lab members. I was very lucky to have been 'paired' to do my rotation project with Hardik, whom I've worked with very closely since then throughout all my graduate years. I call him my 'human Wikipedia' because he just always knows the answer to all my research questions, or at least point to whoever might know the answer. He's full of information about other things he's passionate about, too, and is always down for good fun. Megan was the preceding graduate student whose work attitude I really admire – she really portrays the Wonder Woman in research. I definitely strived to be at least half as hard working as her. Speaking of

hard working, Chang was an exemplary model on that front, too. He seems to be always in lab, working his butt off, even far off from his family in China and later from his newborn son. His determination was really inspiring to many of us in the lab.

So many others in the lab played such foundational roles that I owe tons of gratitude towards because otherwise it would be impossible to do research without them. Danielle, I don't even know your actual official position, but to me you are the manager and the godmother of the labs, you know all the systems and functions that make the lab run. Susan preceded her place and though I only had a chance to get to know her briefly, she was wonderful and was a steppingstone to a lot of things in the lab; I'm happy that she's finally enjoying retirement. Leighton, thank you for being so knowledgeable with all things MEG and so full of wisdom. Corby helped me understand the imaging analysis in which I was so naïve to. Anne helped tremendously with all the consent forms, which was so important especially when I was trying to do my experiment over in Berkeley. Kamalini, to me, is an expert of all things statistics and I always strive to be half as good as her in my understanding of statistics. Karuna was inspirational in her research ideas and I am also grateful for her warmth in welcoming me to the lab. Zarinah, you've really opened my eyes to what a researcher can look like; your involvement in so many things outside of research is truly inspiring and commendable. Carly has always been supportive whenever I asked for help and I really admired her work ethic despite having two little adorable angels. Nayara, I miss your warm approach to human connection, I hope you're doing well. Thank you, Abhi, Ashley, Nina and Jessie for helping me recruit subjects and run my experiments with them. Thank you to many of the interns that came to work only for a bit with the lab, but special thanks to David for helping out tremendously with Lidocaine and for being the master of memes. Sarah was basically a 'sister (graduate student) from another mother (lab)'

and her words of encouragements in my early years were a big reason why I was able to gain confidence in myself and push through. Carrie was the postdoc that you've always heard about even though she wasn't in the lab anymore during my time there because her work was so impactful. Kwang, though you joined the lab right as I was leaving, I am very glad to have crossed paths with you and thankful for your help with the last stretch of FUSP.

I am certainly thankful for the members of both my qualifying exam and thesis committee: Dr. Chris Diederich, Dr. Keith Johnson, Dr. Michael Yartsev and Dr. Christopher Scheriner, for giving very useful input in the framing of my projects that was able to propel me to the next steps in my research. I am also very thankful for Dr. Tony Keaveny whose writing class has been very fundamental in any type of writing I do now. I am also thankful for SarahJane and Kristin for always assisting me in any admin-related issues. Special thanks also to Victoria whose bias for action in my last few months really helped me push through some admin challenges.

I am also thankful for all friends I've made in the program and outside during my time as a graduate student, which I can't possibly list one by one here. Special thanks to Hikaru, who I've luckily met very early on in graduate school and has been a kind friend and wonderful roommate throughout most of my time afterwards. Kaew has also been a dependable friend whose work ethic as a graduate student was helpful to motivate me as well when we were roommates. Marcia was a wonderful landlord that was always so welcoming and warm, making home feels easy to be in during stressful times.

I've also made lifelong friends being involved in St. Anne, which helped me maintain my relationship with God during very stressful times. Father Dan Nascimento is truly one of a kind and his relatable life story makes it easy for me to open up my worries; I never really feel judged by him. I'm sad that I have left a wonderful team of young adult musicians: Gabriella, Cesar and

Anthony. They were truly the reason I still have hope for the church, and I'm so thankful to have shared prayers together. Special thanks also to Kuan, Sharon, Lynda and everyone in the IGSM community also for being supportive spiritually and as good friends when I was still living at Berkeley.

Last but not least, I really don't know how I would be able to survive with my now husband, Charles. I've started graduate school with so many worries and insecurities in my mind, and he always believed in me and pushed me to take action to overcome those. I really think all my confidence was rooted from your belief in me. Thank you for always being so patient with all my emotional breakdowns whenever I'm facing hard times in my research, and thank you for pushing me to get the internship I needed which eventually led me to the job I had hoped for. You are my rock, I love you.

Mechanisms of Online Control of Speech

by

Inez Raharjo

Abstract

Speech is central to human communication. One of the features of speech we control to make distinguishable speech is formants, which are the spectral frequencies in our speech that we vary to make different vowel sounds. It is well-understood that the control of speech features, including formants, depends heavily on feedback information. However, the exact mechanisms of feedback control of formants are not fully understood. The body of this dissertation investigates the feedback control of formants in different scenarios to explore its relationship with other speech control mechanisms. We first investigated the relationship of feedback control and the adaptation of feedforward control of formants and found that they have distinct mechanisms. We then explored the neural substrate of feedback processing for formants as compared to pitch using MEG, which allowed us to investigate network activity changes with a high temporal resolution. We found similar neural regions involved in feedback processing for both pitch and formants, though we observed opposite lateralization for pitch and formants across different frequency bands. Finally, we explored the interaction between auditory and somatosensory feedback information in the control of formants by attenuating oral somatosensory feedback using lidocaine while perturbing auditory feedback. We found a direction-dependent interaction between the two feedback modalities, specifically where we

observed an increase in compensation responses to downward auditory feedback perturbations but a decrease in compensation responses to upward auditory feedback perturbations when the oral somatosensory feedback was attenuated.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Overview

Speech is an important skill for humans to communicate with each other. Producing distinguishable speech requires a surprisingly complex motor behavior and a wide range of articulators as well as fine motor control movement. Humans are able to swiftly execute these speech movements from memory, but they also depend on feedback information, such as auditory and somatosensory feedback, to correct for possible errors in speech. Speech motor control models have acknowledged two speech control mechanisms: the feedforward control, which allows us to produce speech from memory, and for the feedback control, which allows us to use feedback to correct for speech errors (Guenther, 2016; Houde & Nagarajan, 2011; Kearney et al., 2020; Parrell, Ramanarayanan, Nagarajan, & Houde, 2019). However, how the feedback and feedforward control mechanisms are implemented varies among speech motor control models.

1.2 Feedback and feedforward control of speech

As theorized by speech motor control models, the feedback control is responsible for maintaining clear speech production, specifically by utilizing feedback information to correct for errors in speech. Evidence for the role of feedback control in different speech features, such as loudness, pitch and formants, came from feedback perturbation studies which showed compensation responses to perturbed feedback (Bauer, Hain, Mittal, Larson, & Hain, 2006; Burnett, Freedland, Larson, & Hain, 1998; Cai et al., 2012; Hawco & Jones, 2009; Heinks-Maldonado & Houde,

2005; Keough, Hawco, & Jones, 2013; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Larson, Burnett, Kiran, & Hain, 1999; Larson, Sun, & Hain, 2007; Purcell & Munhall, 2006b; Reilly & Dougherty, 2013; Scheerer & Jones, 2018). Moreover, studies have also shown that multiple feedback information plays a role in the control of the different speech features, including auditory (what we hear) and somatosensory feedback (what we feel) (Casserly, Rowley, Marino, & Pollack, 2016; Lametti, Nasir, & Ostry, 2012; Larson, Altman, Liu, & Hain, 2008; Nasir & Ostry, 2008; Tremblay, Shiller, & Ostry, 2003). On the other hand, the feedforward control is responsible for generating correct motor commands for an intended speech, which are learned from past experiences and are not dependent on feedback information. Studies have shown the existence of this feedforward control from the fact that humans are still able to produce speech when feedback information is removed, though with less integrity (Burke, 1969; Jacks & Haley, 2015; Jones & Munhall, 2003; Maas, Mailend, & Guenther, 2015; Niemi, Laaksonen, Ojala, Aaltonen, & Happonen, 2006; Scott & Ringel, 1971; Smith, Stepp, Guenther, & Kearney, 2020). Furthermore, studies have also shown that this feedforward control can adapt and learn a new speech motor mapping if it consistently detects a mismatch between expected and actual feedback information over a long period of time, which leads to a sensorimotor adaptation response (Houde, J. F., & Jordan, 1998; Jones & Munhall, 2000; Katseff, Houde, & Johnson, 2012; Purcell & Munhall, 2006a). Despite the fact that both feedback and feedforward controls have been well studied, the method by which they are implemented differ among speech motor control models. Therefore, an investigation on how the feedback and feedforward control are related to each other is needed in order to settle the differences among speech motor control models.

1.3 Neural mechanisms of feedback control

Some speech motor control models, such as DIVA and SFC, have also theorized the neural mechanisms for feedback control (Guenther, 2016; Houde & Nagarajan, 2011). Multiple studies have investigated the neural network involved in feedback processing for pitch using fMRI and MEG imaging modalities (Kort, Cuesta, Houde, & Nagarajan, 2016; Parkinson et al., 2012; Ranasinghe et al., 2019). However, the neural network involved in the feedback processing of formants have only been studied with fMRI (Niziolek & Guenther, 2013; Tourville, Reilly, & Guenther, 2008). The lack of temporal resolution of fMRI as compared to MEG did not allow for an investigation of the neural network involved in feedback processing of formants in a millisecond timescale (Hari, Levänen, & Raij, 2000; Singh, 2014), which is crucial given the rapid processing for movements of the speech articulators. Furthermore, separate neural-imaging based studies for pitch and formants doesn't allow for a direct within-subject comparison of the neural feedback mechanisms between pitch and formants. Therefore, it would be beneficial to investigate formant feedback processing using MEG imaging, which would help better understand the overall neural feedback mechanisms.

1.4 Auditory and somatosensory feedback in speech

The feedback control mechanism for speech depends on the available feedback information. Two most notable feedback information in speech production are auditory feedback, which is the sound we hear from ourselves as we speak, and somatosensory feedback, which is the somatosensation of the different articulators that we can feel when the speech articulators such as the tongue or jaw move during speaking. It has been observed that compensation responses to auditory feedback perturbations often lead to incomplete compensation responses, where the

responses opposing the perturbation does not equal to the magnitude of the perturbation (Katseff et al., 2012; MacDonald, Goldberg, & Munhall, 2010). This evidence, along with other evidences, indicate a possible competing interplay between auditory and somatosensory feedback information in the production of speech. One study has shown that removing somatosensory feedback affected compensation responses to pitch perturbations (Larson et al., 2008), but whether such effect would be observed in responses to formant perturbations is still unclear.

1.5 Hypotheses

Using auditory feedback perturbation paradigms, we explored short- and long-term responses to auditory feedback perturbations to investigate the relationship between feedback and feedforward control mechanisms. We hypothesized that feedback and feedforward controls are governed by distinct mechanisms. We then explored the neural network involved in the short-term responses to pitch and formant perturbations. We hypothesized that there will be overlapping neural regions in the feedback control of pitch and formants, with potential differences along the somatosensory cortex. Finally, we explored the interplay between auditory and somatosensory feedback information in the production of speech and hypothesized an increased reliance on auditory feedback when somatosensory feedback is removed as would be reflected by larger compensation responses to auditory feedback perturbations. Altogether, these will give us a better understanding on the role of various feedback in speech responses and the underlying neural mechanisms for the control of speech.

1.6 References

Barnes, G. R., Hillebrand, A., Fawcett, I. P., & Singh, K. D. (2004). Realistic spatial sampling for MEG beamformer images. *Human Brain Mapping*, *23*(2), 120–127. https://doi.org/https://doi.org/10.1002/hbm.20047

Bauer, J. J., Hain, T. C., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, *119*(4), 2363–2371. https://doi.org/10.1121/1.2173513

Bearelly, S., & Cheung, S. W. (2017). Sensory Topography of Oral Structures. *JAMA Otolaryngology–Head & Neck Surgery*, *143*(1), 73–80. https://doi.org/10.1001/jamaoto.2016.2772

Burke, B. D. (1969). Reduced auditory feedback and stuttering. *Behaviour Research and Therapy*, *7*(3), 303–308. https://doi.org/https://doi.org/10.1016/0005-7967(69)90011-4

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE*, *7*(7), 1–13. https://doi.org/10.1371/journal.pone.0041830

Casserly, E. D., Rowley, M. C., Marino, F. R., & Pollack, E. A. (2016). Acoustic analysis of speech produced with degradation of acoustic and somatosensory feedback. *Proceedings of Meetings on Acoustics*, *29*(1), 60016. https://doi.org/10.1121/2.0000651

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7). https://doi.org/10.1073/pnas.1216827110

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

Dalal, S. S., Guggisberg, A. G., Edwards, E., Sekihara, K., Findlay, A. M., Canolty, R. T., Berger, M. S., Knight, R. T., Barbaro, N. M., Kirsch, H. E., Nagarajan, S. S. (2008). Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. *NeuroImage*, *40*(4), 1686–1700. https://doi.org/10.1016/j.neuroimage.2008.01.023

Dalal, S. S., Zumer, J. M., Guggisberg, A. G., Trumpis, M., Wong, D. D. E., Sekihara, K., & Nagarajan, S. S. (2011). MEG/EEG Source Reconstruction, Statistical Evaluation, and Visualization with NUTMEG. *Computational Intelligence and Neuroscience*, *2011*, 758973. https://doi.org/10.1155/2011/758973

Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679. https://doi.org/10.1152/jn.00638.2010

Guenther, F. H. (2016). *Neural Control of Speech*. The MIT Press.

Hall, E. L., Robson, S. E., Morris, P. G., & Brookes, M. J. (2014). The relationship between MEG and fMRI. *NeuroImage*, *102*, 80–91. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.11.005

Hari, R., Levänen, S., & Raij, T. (2000). Timing of human cortical functions during cognition: role of MEG. *Trends in Cognitive Sciences*, *4*(12), 455–462.

https://doi.org/https://doi.org/10.1016/S1364-6613(00)01549-7

Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain Research*. https://doi.org/10.1016/j.brainres.2009.04.033

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*, *6*(3), 131–137. https://doi.org/10.1121/1.1931747

Ho, A. M. H., Chung, D. C., To, E. W. H., & Karmakar, M. K. (2004). Total airway obstruction during local anesthesia in a non-sedated patient with a compromised airway. *Canadian Journal of Anesthesia*, *51*(8), 838. https://doi.org/10.1007/BF03018461

Houde, J. F., & Jordan, M. I. (1998). *Sensorimotor adaptation in speech production. Science, 279(5354), 1213-1216..pdf. 1*.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*(October), 1–14. https://doi.org/10.3389/fnhum.2011.00082

Jacks, A., & Haley, K. L. (2015). Auditory Masking Effects on Speech Fluency in Apraxia of Speech and Aphasia: Comparison to Altered Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1670–1686. https://doi.org/10.1044/2015_JSLHR-S-14-0277

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *J. Acoust. Soc. Am.*, *108*(3 Pt 1), 1246–1251. https://doi.org/10.1121/1.1288414

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, *113*(1),

532–543. https://doi.org/10.1121/1.1529670

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308. https://doi.org/10.1177/0023830911417802

Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., Ballard, K. J., Chang, S., Chao, S., Murrray, E. S. H., Scott, T. L., Guenther, F. H. (2020). A Simple 3-Parameter Model for Examining Adaptation in Speech and Voice Production . *Frontiers in Psychology* , Vol. 10, p. 2995. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.02995

Keough, D., Hawco, C. S., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience*, *14*(1), 25. https://doi.org/10.1186/1471-2202-14-25

Kim, K. S., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research : JSLHR*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_jslhr-19-00419

Kort, N. S., Cuesta, P., Houde, J. F., & Nagarajan, S. S. (2016). Bihemispheric network dynamics coordinating vocal feedback control. *Human Brain Mapping*, *37*(4), 1474–1485. https://doi.org/10.1002/hbm.23114

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The*

Journal of the Acoustical Society of America, *89*(2), 859–866. https://doi.org/10.1121/1.1894647

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research*, *187*(4), 613–621. https://doi.org/10.1007/s00221-008-1330-z

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, *110*(6), 2845–2848. https://doi.org/10.1121/1.1417527

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (1999). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America*, *107*(1), 559–564. https://doi.org/10.1121/1.428323

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862–2872. https://doi.org/10.1121/1.2715657

Maas, E., Mailend, M.-L., & Guenther, F. H. (2015). Feedforward and Feedback Control in Apraxia of Speech: Effects of Noise Masking on Vowel Production. *Journal of Speech, Language, and Hearing Research*, *58*(2), 185–200. https://doi.org/10.1044/2014_JSLHR-S-13-0300

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068. https://doi.org/10.1121/1.3278606

Mitsuya, T., MacDonald, E. N., Munhall, K. G., & Purcell, D. W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of*

America, *138*(1), 413–424. https://doi.org/10.1121/1.4923154

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, *11*(10), 1217–1222. https://doi.org/10.1038/nn.2193

Niemi, M., Laaksonen, J. P., Ojala, S., Aaltonen, O., & Happonen, R. P. (2006). Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/. *International Journal of Oral and Maxillofacial Surgery*, *35*(10), 920–923. https://doi.org/10.1016/j.ijom.2006.06.002

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *Journal of Neuroscience*, *33*(29). https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology*, *15*(9), e1007321.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(January 2006), 2288–2297. https://doi.org/10.1121/1.2173514

Ranasinghe, K. G., Kothare, H., Kort, N., Hinkley, L. B., Beagle, A. J., Mizuiri, D., Honma, S.

M., Lee, R., Miller, B. L., Gorno-Tempini, M. L., Vossel, K. A., Houde, J. F., Nagarajan, S. S. (2019). Neural correlates of abnormal auditory feedback processing during speech production in Alzheimer's disease. *Scientific Reports*, *9*(1), 5686. https://doi.org/10.1038/s41598-019-41794-x

Reilly, K. J., & Dougherty, K. E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4812763

Scheerer, N. E., & Jones, J. A. (2018). The Role of Auditory Feedback at Vocalization Onset and Mid-Utterance. *Frontiers in Psychology*, *9*, 2019. https://doi.org/10.3389/fpsyg.2018.02019

Scott, C. M., & Ringel, R. L. (1971). Articulation Without Oral Sensory Control. *Journal of Speech and Hearing Research*, *14*(4), 804–818. https://doi.org/10.1044/jshr.1404.804

Singh, S. P. (2014). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*(Suppl 1), S107–S112. https://doi.org/10.4103/0972-2327.128676

Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of Auditory and Somatosensory Feedback to Vocal Motor Control. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2039–2053. https://doi.org/10.1044/2020_JSLHR-19-00296

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/J.NEUROIMAGE.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869. https://doi.org/10.1038/nature01710

Chapter 2: Speech compensation responses and sensorimotor adaptation to formant feedback perturbations

2.1 Abstract

Control of speech formants is important for the production of distinguishable speech sounds and is achieved with both feedback and learned feedforward control. However, it is unclear whether the learning of feedforward control involves the mechanisms of feedback control. Speakers have been shown to compensate for unpredictable transient mid-utterance perturbations of pitch and loudness feedback, demonstrating online feedback control of these speech features. To determine whether similar feedback control mechanisms exist in the production of formants, responses to unpredictable vowel formant feedback perturbations were examined. Results showed similar within-trial compensatory responses to formant perturbations that were presented at utterance onset and mid-utterance. The relationship between online feedback compensation to unpredictable formant perturbations and sensorimotor adaptation to consistent formant perturbations was further examined. Within-trial online compensation responses were not correlated with across-trial sensorimotor adaptation. A detailed analysis of within-trial time course dynamics across trials during sensorimotor adaptation revealed that across-trial sensorimotor adaptation responses did not result from an incorporation of within-trial compensation response. These findings suggest that online feedback compensation and sensorimotor adaptation are governed by distinct neural mechanisms. These findings have

important implications for models of speech motor control in terms of how feedback and feedforward control mechanisms are implemented.

2.2 Introduction

Speaking is unique among motor behaviors for human communication – it is the prime conveyor of linguistic intent – and it depends upon incredibly precise timing and coordination of many independent articulators. Speaking's most basic defining feature that sets it apart from other motor actions is that speaking produces sound. Yet in spite of its importance to speech, the role of auditory feedback in normal speech production and in the control of speech remains unclear. One plausible way for speech to be maintained and controlled is through the monitoring of auditory feedback and using that feedback information to correct for errors in speech output. Although the sensory processing delays (50-150 ms), seen in human and non-human primates (Bendor & Wang, 2005; Houde & Nagarajan, 2011; Jenkins 3rd, Idsardi, & Poeppel, 2010; Poeppel & Hickok, 2015), inherent to this correction process rule out the possibility that speech is controlled solely and directly by auditory feedback, research has shown that the control of natural speech is nevertheless responsive to changes in auditory feedback (Lametti, Smith, Watkins, & Shiller, 2018; Liu, Xu, & Larson, 2009). Auditory feedback thus plays a modulatory role in ongoing speech production, even in dynamically changing natural speech. Numerous studies have investigated the role of auditory feedback in speech production through various auditory feedback perturbation experiments, which usually examine responses to either unpredictable or consistent, predictable perturbations (Burnett, Freedland, Larson, & Hain, 1998; Caudrelier & Rochet-Capellan, 2019, p. 15-75; Chen, Liu, Xu, & Larson, 2007; Houde & Jordan, 1998; Houde & Nagarajan, 2015, p. 267-298).

Within-trial online compensation has been observed in response to unpredictable auditory feedback perturbations, i.e. perturbations that are randomly and unexpectedly applied within an utterance. Speech production models (Guenther, 2016; Houde & Nagarajan, 2011; Kearney et al., 2020; Parrell, Ramanarayanan, Nagarajan, & Houde, 2019) have theorized that within-trial online compensatory responses are generated by a feedback control mechanism, where auditory feedback is compared to an internal representation of expected auditory feedback. If a mismatch is detected, motor correction commands are sent to speech articulators to correct for the mismatch. Studies have investigated the within-trial online compensatory responses to unpredictable pitch and loudness perturbations, both at mid-utterance and at utterance onset (Bauer, Hain, Mittal, Larson, & Hain, 2006; Burnett et al., 1998; Hawco & Jones, 2009; Heinks-Maldonado & Houde, 2005; Keough, Hawco, & Jones, 2013; Larson, Burnett, Bauer, Kiran, & Hain, 2001; Larson, Burnett, Kiran, & Hain, 1999; Larson, Sun, & Hain, 2007; Scheerer & Jones, 2018). In the control of pitch, some differences have been observed in online compensatory responses to pitch feedback perturbations that occur at mid-utterance or at utterance onset, suggesting possible different mechanisms governing the two (Hawco & Jones, 2009; Scheerer & Jones, 2018). In the control of formants, responses to unpredictable perturbations applied at utterance onset have been studied (Cai et al., 2012; Reilly & Dougherty, 2013), but only one study has investigated the responses to unpredictable formant feedback perturbations at mid-utterance (Purcell & Munhall, 2006b). However, in that study the perturbation was implemented via a cross-fading technique, and thus was not directly comparable to the way perturbations are usually implemented in pitch and loudness feedback

perturbation studies. Thus, it remains unclear whether shared online compensation mechanisms exist for formant control at utterance onset and at mid-utterance.

Understanding the mechanisms of within-trial compensatory responses would help us determine how these mechanisms relate to another type of speech response to perturbed feedback – across-trial sensorimotor adaptation. Across-trial sensorimotor adaptation has been observed in the production of both pitch and formants (Houde & Jordan, 1998; Jones & Munhall, 2000; Katseff, Houde, & Johnson, 2012; Purcell & Munhall, 2006a) in response to consistent, predictable auditory feedback perturbations, i.e. perturbations that are consistently applied over many trials, has been observed. There is a general agreement among speech production models (Guenther, 2016; Houde & Nagarajan, 2011; Kearney et al., 2020; Parrell et al., 2019) that this sensorimotor adaptation involves learning long-term changes in feedforward control that gradually anticipate the effects of consistent, predictable auditory feedback perturbations. The mechanism that accomplishes this sensorimotor adaptation of feedforward control, however, is less clear. The DIVA model and its simpler version the SimpleDIVA model assume a close relationship between feedback and feedforward control, wherein sensorimotor adaptation arises from feedforward control being learned via incorporation of online feedback compensations (Guenther, 2016; Kearney et al., 2020). As an alternative DIVA, SFC models have long been used in non-speech motor behaviors (Shadmehr, Smith, & Krakauer, 2010), and more recently have also been applied to speech behaviors (Houde & Nagarajan, 2011; Parrell et al., 2019). SFC models, on the other hand, can accommodate adaptation resulting directly from sensory prediction errors rather than necessarily being derived from the incorporation of corrective movements as is assumed in DIVA. Data consistent with SFC models have shown that online

compensation and sensorimotor adaptation can be differently affected when comparing patient and control groups (Abur et al., 2018; Chen et al., 2013; Demopoulos et al., 2018; Mollaei, Shiller, Baum, & Gracco, 2016; Mollaei, Shiller, & Gracco, 2013; Parrell, Agnew, Nagarajan, Houde, & Ivry, 2017), suggesting a potential underlying difference between the control mechanisms of online compensation and sensorimotor adaptation. Given that both online compensation and sensorimotor adaptation responses are observed in the production of formants, question arises as to whether both are controlled with the same underlying neural mechanism.

The current study investigated the within-trial online compensation and across-trial sensorimotor adaptation responses to formant feedback perturbations during speaking to elucidate the relationship between the feedback and feedforward control mechanisms of speech. We first examined online compensation responses to two types of unpredictable formant feedback perturbations: (1) whole-trial perturbations applied at utterance onset, and (2) transient perturbations applied at mid-utterance. We applied these perturbations at varying magnitudes and directions, similar to prior feedback perturbation studies (Bauer et al., 2006; Burnett, Senner, & Larson, 1997; Cai et al., 2012; Hawco & Jones, 2009; Katseff et al., 2012; Larson et al., 2001; Parrell et al., 2017; Purcell & Munhall, 2006b), to increase the unpredictability of the perturbations and to investigate whether the magnitude and direction of perturbations have an effect on the online compensation responses. Similar to that seen in responses to pitch feedback perturbations, we hypothesized we would see within-trial online compensatory responses to both the utterance-onset and mid-utterance types of formant feedback perturbations. We further examined whether responses to these two types of feedback perturbations are governed by different mechanisms (as has been suggested in a prior study (Hawco & Jones, 2009)) using

regression analyses, which has commonly used to study the relationship of speech responses and underlying mechanisms (Lametti, Nasir, & Ostry, 2012; Lametti, Rochet-Capellan, Neufeld, Shiller, & Ostry, 2014; Mollaei, Shiller, Baum, & Gracco, 2019). If the within-trial online compensatory responses to utterance-onset and mid-utterance types of feedback perturbations are governed by different mechanisms, as has been suggested to be the case for responses to pitch feedback perturbations, we would predict that the responses to formant perturbations at utterance onset and at mid-utterance will not be correlated with each other. Alternately, if these two within-trial online compensatory responses are governed by similar mechanisms, we would expect a correlation between the respective compensatory responses.

We also examined sensorimotor adaptation responses to consistent, predictable formant feedback perturbations. First, we examined the relationship between within-trial online feedback compensation to unpredictable formant perturbations and sensorimotor adaptation to consistent, predictable formant perturbations. If sensorimotor adaptation processes depend on within-trial online feedback compensation processes, we would expect a correlation between within-trial online compensation and across-trial sensorimotor adaptation responses. Alternatively, if they are governed by distinct neural mechanisms, we would predict that sensorimotor adaptation responses will not be correlated with within-trial online compensation responses to either types of unpredictable formant feedback perturbations.

Finally, we investigated the relationship between across-trial sensorimotor adaptation and within-trial online compensation by examining how the within-trial responses evolved during sensorimotor adaptation. If the mechanisms for sensorimotor adaptation depend on within-trial

online compensation, we would at the very least expect that throughout the trials of the sensorimotor adaptation, the within-trial online compensation response would gradually be subsumed by a feedforward sensorimotor adaptation response that starts at the beginning of each trial. Alternatively, if online compensation and sensorimotor adaptation are governed by distinct neural mechanisms, then we might expect to see no association between within-trial responses and the growth of across-trial sensorimotor adaptation.

2.3 Methods

2.3.1 Participants

Healthy participants were recruited for the study (n=23, 7 females) through UC Berkeley classes announcements, pamphlets and online platforms. Participants' ages ranged from 18 to 43 years (mean ± standard deviation of 21.7 ± 5.5 years). All participants were native English speakers and the majority were bilingual/multilingual (n=22). One multilingual participant's data was taken out because their sensorimotor adaptation response was an outlier that far exceeded three standard deviations from the median sensorimotor adaptation across participants. As a result, data from 22 participants was included in the analysis. Participants had no deficits in learning, motor, or speech and language abilities and gave written informed consent to participate. The study was approved by the UCSF Institutional Review Board for human research.

2.3.2 Apparatus

The experiments were performed in a quiet room equipped with sound booth. While inside the sound booth, participants sat in front of a laptop (Thinkpad W530, Lenovo Group Limited) while wearing Beyerdynamic DT 770 Pro 250 Ohm headphones and a head-mounted AKG Pro Audio

C520 condenser microphone. Participant's speech from the microphone was fed into a Focusrite Scarlett 2i2 USB Recording Audio Interface and processed and recorded using MATLAB (which also displayed the word prompts) paired with the Feedback Utility for Speech Production (FUSP) software. FUSP repeatedly analyzed 3ms frames of speech input from the microphone into separate pitch and formant representations that were, at times, altered (depending on the experiment) and used to synthesize the next 3ms of speech output to the participant's headphones. The speech data was recorded at a rate of 11025Hz and this feedback processing, along with hardware delays, introduced an imperceptible ~21ms delay in the auditory feedback, as measured following the methods outlined by Kim, et al. (Kim, Wang, & Max, 2020).

2.3.3 Experimental design and procedures

The current study consisted of 5 sessions of 165 trials each with formant feedback perturbations. Trials in sessions 1, 3 and 5 included *unpredictable formant perturbations*, applied either at utterance onset or at mid-utterance, performed to examine *within-trial online compensation responses*. Trials in sessions 2 and 4 included *consistent, predictable formant perturbations*, performed to examine *across-trial sensorimotor adaptation responses*. For any given trial, participants were instructed to say either the word 'head' (/hɛd/) or 'hid' (/hId/), extending the vowel portion of the utterance for as long as the prompt word was displayed on the screen (approximately 2 seconds). The 'hid' catch trials were infrequent (number of trials discussed below). These catch trials were added to (1) keep the experiment more engaging so as to prevent participants from becoming bored and inattentive to the speech production task, and to (2) encourage participants to make contrastive productions of these vowels and pay more attention to the acoustics of their production. Before the first session, a short practice session with

experimenter (about 5-10 trials as needed) was done to ensure that the participant could hold out their word steadily as instructed. Within each session, there were random-length breaks (1.7-2.7 seconds) between trials as well as self-paced breaks after every 15 trials. Five-minute breaks were also administered in between sessions in an attempt to wash out possible carryover effects from previous sessions, during which the experimenter would verbally engage with the participant, for example by asking questions relating to the current experiment.

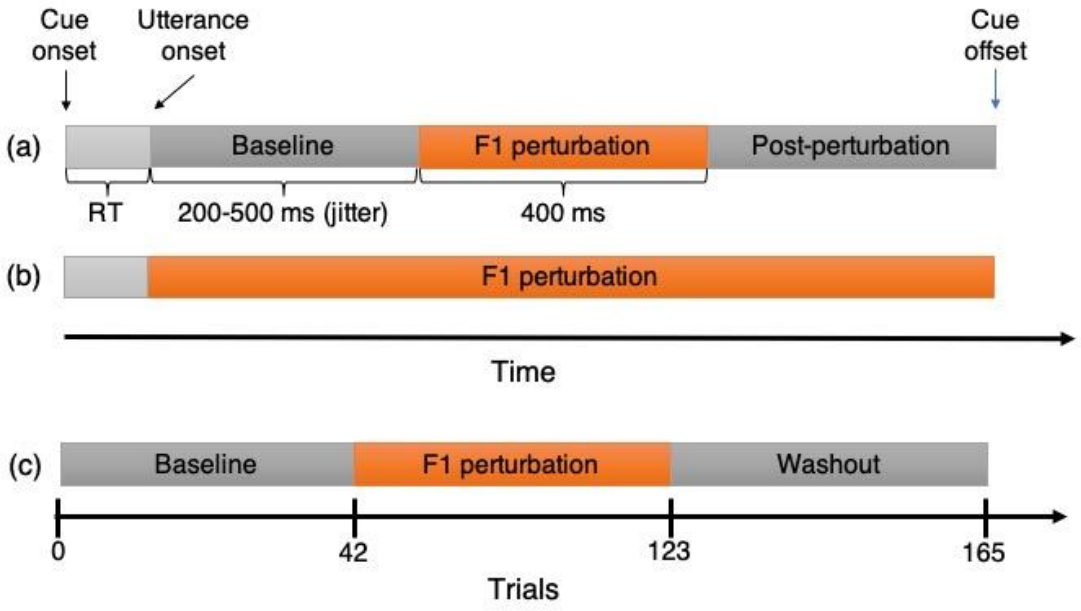

Figure 2.1: Experimental paradigm. F1 perturbation applied at different timescales: (**a**) unpredictable transient mid-utterance perturbation (a 400ms perturbation initiated after a 200-500ms jitter delay from utterance onset), (**b**) unpredictable whole utterance perturbation, initiated at utterance onset and sustained for the whole utterance, and (**c**) consistent, predictable utterance-onset whole utterance perturbation applied over many trials (sensorimotor adaptation). RT = Reaction time (light grey bars). Grey (light and dark) bars indicate where no perturbation was applied, and orange bars indicate where perturbation was applied.

Within-trial online compensation response data were collected from 495 trials evenly distributed across sessions 1, 3, and 5. Session 1 began with 15 familiarization phase trials included to acquaint participants to the experimental task and pace, with each trial being randomly chosen

from any one of the trial types described below. The remaining 480 trials were distributed across the remainder of session 1 as well as sessions 3 and 5. These 480 trials consisted of 432 perturbation trials (360 'head' and 72 'hid' trials) and 48 unperturbed trials (30 'head' and 18 'hid' trials). Perturbation trials were unpredictable perturbations of the first formant (F1) of subjects' auditory feedback. Two types of unpredictable formant perturbations were applied: (1) unpredictable formant perturbations applied at mid-utterance, transiently for 400ms with a 200-500ms jittered delay from utterance onset (*unpredictable mid-utterance perturbations*, Figure 1a), and (2) unpredictable formant perturbations at utterance onset applied for the entire trial duration (*unpredictable whole utterance perturbations*, Figure 1b). The 400ms duration for the transient perturbations was chosen as it has often been used in previous mid-utterance pitch and loudness perturbation studies (Heinks-Maldonado & Houde, 2005; Kort, Cuesta, Houde, & Nagarajan, 2016). Four different F1 feedback perturbations were applied: -50Hz, +50Hz, -200Hz and +200Hz for each mid- and whole-utterance perturbation, totaling to 8 formant feedback perturbation conditions. These 8 conditions were randomly distributed across the perturbation trials. The choice of perturbation magnitudes of 50Hz and 200Hz are well within the range of values that have been used in prior formant sensorimotor adaptation studies (Katseff et al., 2012; MacDonald, Goldberg, & Munhall, 2010).

Across-trial sensorimotor adaptation response data were collected in Sessions 2 and 4. In each of these sessions, the F1 perturbation was consistently applied over many trials arranged as a sequence of phases (Figure 1.3): (1) a baseline phase of 42 trials (30 'head' trials and 12 'hid' trials), where feedback was unperturbed, (2) a hold phase of 81 trials (60 'head' trials and 21 'hid' trials), where F1 feedback was consistently perturbed by whole-trial perturbations of either

+200 or -200Hz (depending on the session), and (3) a washout phase of 42 trials (30 'head' trials and 12 'hid' trials), where feedback was again unperturbed. The session order for the two F1 perturbation directions was counterbalanced among participants, i.e. half of the participants received +200 Hz F1 feedback perturbation for session 2 and -200 Hz F1 feedback perturbation for session 4; for the other half, this order was reversed. The 'hid' trials were again included as catch trials and not included in the analysis. The 'head' and 'hid' trials were randomized within each phase of each adaptation session.

2.3.4 Data processing and statistical analysis

All acoustic speech data was analyzed using Wave Viewer, a custom-built MATLAB-based speech analysis software (https://github.com/SpeechNeuroscienceLab/Wave-Viewer). In each trial, formants were tracked using linear predictive coding (LPC). The tracking for the first formant was further refined by manual screening, as needed, to exclude bad trials (e.g. trials with no speech response, interruption in speech production/recording, and poor formant tracking) and to occasionally fix the voice onset and offset time markings automatically detected by FUSP. The analysis focused on all good 'head' trials (including those from the familiarization phase). The 'hid' trials were excluded from analysis because they were designed as infrequent catch trials and were thus unsuitable for reliable statistical analysis. On average, less than 3% of 'head' trials were excluded from analysis across all subjects. Using the raw, F1 formant track trial data extracted, the following six analyses were performed:

2.3.4.1 Within-trial Online Compensation Responses to Unpredictable Formant Perturbations

To obtain an F1 response time-course that could highlight each participant's F1 response changes elicited by the perturbations, we performed three linear normalization steps that eliminated within- and across-trial variance. First, each trial, aligned at voice onset, was normalized by subtracting the participant's F1 unperturbed response trend (average F1 response in the unperturbed trials, normalized at voice onset such that F1(t=0) = 0Hz). This was done to reduce the within-trial variations in F1 responses of each participant. Second, each trial was then aligned at perturbation onset and normalized to the first 50ms of the post perturbation onset data – a time before subjects could detect and begin responding to the perturbation (Houde & Nagarajan, 2011; Jenkins 3rd et al., 2010; Poeppel & Hickok, 2015). This was done to reduce across-trials variations in F1 responses. Third, each trial was smoothed by averaging the F1 responses within non-overlapping 25ms windows. This was done to reduce the formant tracking variations across frames. Finally, the trials from session 1, 3, and 5 were grouped together by conditions and averaged to obtain a F1 response time-course for each participant. The average and standard error of the F1 response time-courses were then calculated across participants.

The group onset response latency was calculated as the time point where the response exceeded two standard deviations from the mean of the onset response data, which is the responses in the first 50ms after perturbation onset within perturbation conditions of the same type. For example, the group onset response latency for the -200Hz mid-utterance perturbation condition was calculated as the time point where the response to the -200 mid-utterance perturbation exceeded two standard deviations from the mean of the first 50ms responses in all (+200Hz, -200Hz, +50Hz and -50Hz) mid-utterance perturbation conditions. Using the first 50ms responses across

all perturbations within the same type was done to reduce noise in the baseline data and thus increase the sensitivity of detecting the response onset latency. To obtain the distribution of peak compensation response percentages for each of the 8 formant perturbation conditions, we first calculated each participant's peak response by averaging their normalized F1 time-course response in a 200ms time window around the group peak response latency for each condition. The calculated peak responses were then converted into compensation percentages by using the formula $\frac{Peak\ Normalized\ Response\ in\ Hz}{Perturbation\ Magnitude\ in\ Hz} \times 100\% \times compMult$ where $compMult$ (compensation multiplier) was -1 for positive perturbations and +1 for negative perturbations, in order to make all compensatory responses positive regardless of the sign of the perturbation. Violin plots were created using the peak compensation response percentages, and a one-sample two-tailed t-test for each condition was calculated to test for significance different from zero. A linear mixed effects (LME) model was run in SAS 9.4 (SAS Institute Inc., Cary, NC) using the proc mixed procedure to evaluate the main effect of perturbation magnitude (50Hz vs 200Hz), direction (positive vs negative), and type of perturbation onset (mid- vs whole-utterance) on the individual peak percent compensation responses with participant as a random factor. To account for covariation of participant's age and baseline F1 production (non-normalized F1 production at perturbation onset, calculated for each perturbation condition), we added these factors as nuisance covariates to the LME model.

2.3.4.2 Relationship Between Responses to Unpredictable Perturbations at Different Onsets

To evaluate the relationship between responses to unpredictable perturbations at mid-utterance and at utterance onset, we compared the peak compensation response percentages for the unpredictable mid-utterance perturbations and to unpredictable whole-utterance perturbations.

Scatterplots were created for the responses to each of the four perturbation values (+50Hz, -50Hz, +200Hz and -200Hz) as well as for responses to all perturbation values combined. For each scatter plot, we fitted a linear mixed-effects model using the fitlme function in MATLAB. The formula used was for a random intercept model with a fixed slope, $whole \sim 1 + mid + (1|participants)$, where $whole$ and mid represent peak compensation response percentages for unpredictable whole- and mid-utterance perturbations respectively.

2.3.4.3 Sensorimotor Adaptation Responses to Consistent, Predictable Formant Perturbations

To calculate the sensorimotor adaptation response over the course of the trials in the adaptation sessions, trials from sessions 2 and 4 were analyzed by averaging the first 75ms of F1 data points in each trial. The first 75ms of F1 data was examined to isolate the initial feedforward adaptation responses from subsequent feedback-based within-trial responses (75ms is a time before subjects could detect and begin any within-trial online compensatory response to the perturbation) (Houde & Nagarajan, 2011; Jenkins 3rd et al., 2010; Poeppel & Hickok, 2015). Each participant's F1 response across trials was then normalized using their average F1 responses in the baseline phase (first 30 'head' trials) to highlight each participants' F1 sensorimotor adaptation response changes and smoothed by averaging this F1 normalized response within non-overlapping five-trial windows to reduce formant tracking variations across trials. The mean and standard error of the F1 across-trial sensorimotor adaptation response were then calculated across participants.

To obtain the distribution of sensorimotor adaptation response percentages for +200Hz and -200Hz F1 perturbations, we first obtained sensorimotor adaptation responses by averaging each

participant's normalized F1 trial-course response in last 15 'head' trials of the hold phase (trials 76-90). The calculated sensorimotor adaptation responses were then similarly converted into compensation percentages as described in the unpredictable perturbation section above. Violin plots were created using these sensorimotor adaptation response percentages, and a one-sample two-tailed t-test for each condition was calculated to test for significance different from zero. A linear mixed effects model was run in SAS to evaluate the main effect of perturbation direction (positive vs negative) on the sensorimotor adaptation response percentages with participant as a random factor. To account for covariation of participant's age and baseline F1 production (average first 75ms F1 production in all trials of the baseline phase, calculated for each perturbation direction), we added these factors as nuisance covariates to the LME model.

2.3.4.4 Relationship Between Responses to Unpredictable and Consistent, Predictable Perturbations

To evaluate the relationship between responses to unpredictable perturbations and to consistent, predictable perturbations, we compared the sensorimotor adaptation responses to consistent, predictable perturbations to the peak compensation responses to both types of unpredictable perturbations. Scatterplots were created for the comparisons of sensorimotor adaptation responses to responses to each of the unpredictable perturbation types (mid- and whole utterance). For each scatter plot, we fitted a linear mixed-effects model using the formula $adapt \sim 1 + unpredict + (1|participants)$, where $adapt$ represents sensorimotor adaptation response percentages and $unpredict$ represents peak compensation response percentages for unpredictable (either mid- or whole utterance) perturbations.

2.3.4.5 Within-trial responses in adaptation experiments

We investigated whether within-trial compensation drives sensorimotor adaptation by examining the within-trial responses in the different phases of the adaptation experiments. To do this, we selected a subset of participants who showed both a within-trial online compensation response to the unpredictable perturbations as well as a sensorimotor adaptation response to consistent, predictable perturbations. There were 15 participants who positively compensated to the +200Hz unpredictable whole-utterance perturbations and positively adapted to the +200Hz consistent, predictable perturbations. There were 12 participants who positively compensated to the -200Hz unpredictable whole-utterance perturbations and positively adapted to the -200Hz consistent, predictable perturbations.

To obtain the within-trial responses in the different phases of the adaptation experiments, each selected participant's within-trial F1 response time-courses in 'head' trials were averaged in each of the following analysis phases: late baseline phase (last 15 'head' trials in baseline phase, trials 16-30), early adaptation phase (first 15 'head' trials in hold phase, trials 31-45), late adaptation phase (last 15 'head' trials in hold phase, trials 76-90), and late washout phase (last 15 'head' trials in washout phase, trials 106-120). The individual averaged within-trial F1 response time-courses were then normalized by subtracting the individual's average F1 time-course response in the baseline phase (first 30 trials). This was done reduce the within-trial variations in F1 responses of each participant. The average and standard error of the time-course within-trial F1 responses in the different phases of the adaptation experiments were then calculated across all selected participants.

To highlight the change in responses in the initial feedforward responses across the different phases, we plotted boxplots showing the distribution of the selected participants' response onsets (0-75ms) in terms of percent compensation for each analysis phase in the adaptation experiment (Figure 6c-d). One-sample t-test for each boxplot distribution was calculated to test for significance different from zero. Additionally, to highlight the change in within-trial responses, we plotted pairwise scatterplot overlaid over boxplots showing the distribution of the selected participants' responses at onset ('O', 0-75ms) and at mid-utterance ('M', 600-800ms) in terms of percent compensation for each phase in the adaptation experiment (Figure 6e-f). A linear mixed effects model was run in SAS to evaluate the main effect of within-trial time window (0-75ms vs 600-800ms), adaptation analysis phase (late baseline, early adaptation, late adaptation and late washout), and perturbation direction (+200Hz vs -200Hz) on the individual within-trial percent compensation responses with participant as a random factor.

2.3.4.6 Evaluating Session Order Effects

In our experimental design, 2 out of the 3 sessions consisting of unpredictable formant perturbations (i.e., sessions 3 and 5) followed sessions with adaptation experiments (sessions 2 and 4). Given that previous studies have shown short-term changes in speech motor control following adaptation experiments (Ito, Coppola, & Ostry, 2016; Shiller, Sato, Gracco, & Baum, 2009), this experimental structure may have had carryover effects from the consistent, predictable perturbation sessions to the following unpredictable perturbation sessions. To examine whether any such carryover effects took place in our study, we performed two analyses on F1 production in the unpredictable perturbation sessions. The first analysis was done to investigate whether participant's baseline F1 production shifted across the 3 sessions. To do this,

we examined the first 50ms of the F1 formant tracks in each trial in session 1, 3, and 5. This time window was selected to avoid including any possible within-trial responses to the unpredictable perturbations. ANOVA was then performed on the median baseline F1 production to evaluate the effect of session order with participants as a random factor. A second analysis was done to investigate whether participant's peak compensation response percentages shifted across the 3 sessions. To do this, we calculated the peak percent compensation response percentages for each perturbation condition, similar to what was done above, but separately for each session normalized response average. A linear mixed effects model was then run to evaluate the main effect of session order (session 1, 3 and 5), perturbation magnitude (50Hz vs 200Hz), perturbation direction (positive vs negative) and type of perturbation onset (mid- vs whole-utterance) on the peak percent compensation response percentages with participant as a random factor.

2.4 Results

2.4.1 Participants compensated to unpredictable mid- and whole-utterance formant perturbations

Participants exhibited within-trial online compensation not only for F1 perturbations applied unpredictably for the whole-utterance (Figure 2b), but also for F1 perturbations applied unpredictably at mid-utterance (Figure 2a). Response onset latencies for -50Hz, +50Hz, -200Hz and +200Hz perturbations are shown in Table 1. The distributions of peak compensation responses are shown as violin plots in Figure 2c (mid-utterance) and 2d (whole utterance), and significance levels, when compared to zero, are shown in Table 1. Responses to mid- and whole-utterance perturbations were significant compared to zero with the exception of responses to -

50Hz mid-utterance perturbations. A linear mixed effects model of peak compensation response percentages showed significant main effects for perturbation magnitude (200Hz vs 50Hz: $F(1,146)=8.32$, $p=4.50\times10^{-3}$) and direction (positive vs negative: $F(1,146)=5.10$, $p=0.03$), but not for type of perturbation (mid- vs whole-utterance: $F(1,146)=3.72$, $p=0.06$). There are no significant interaction between perturbation magnitude and direction ($F(1,146)=4.52$, $p=0.11$), perturbation magnitude and type of perturbation ($F(1,146)=0.72$, $p=0.40$), perturbation direction and type of perturbation ($F(1,146)=0.02$, $p=0.90$) and between perturbation magnitude, direction and type of perturbation ($F(1,146)=0$, $p=0.99$). No significant effects for the covariates of age or baseline F1 production was found ($F(1,146)=0.23$, $p=0.63$ and $F(1,146)=1.51$, $p=0.22$ respectively). Overall, participants significantly compensated for both unpredictable mid- and whole-utterance F1 perturbations.

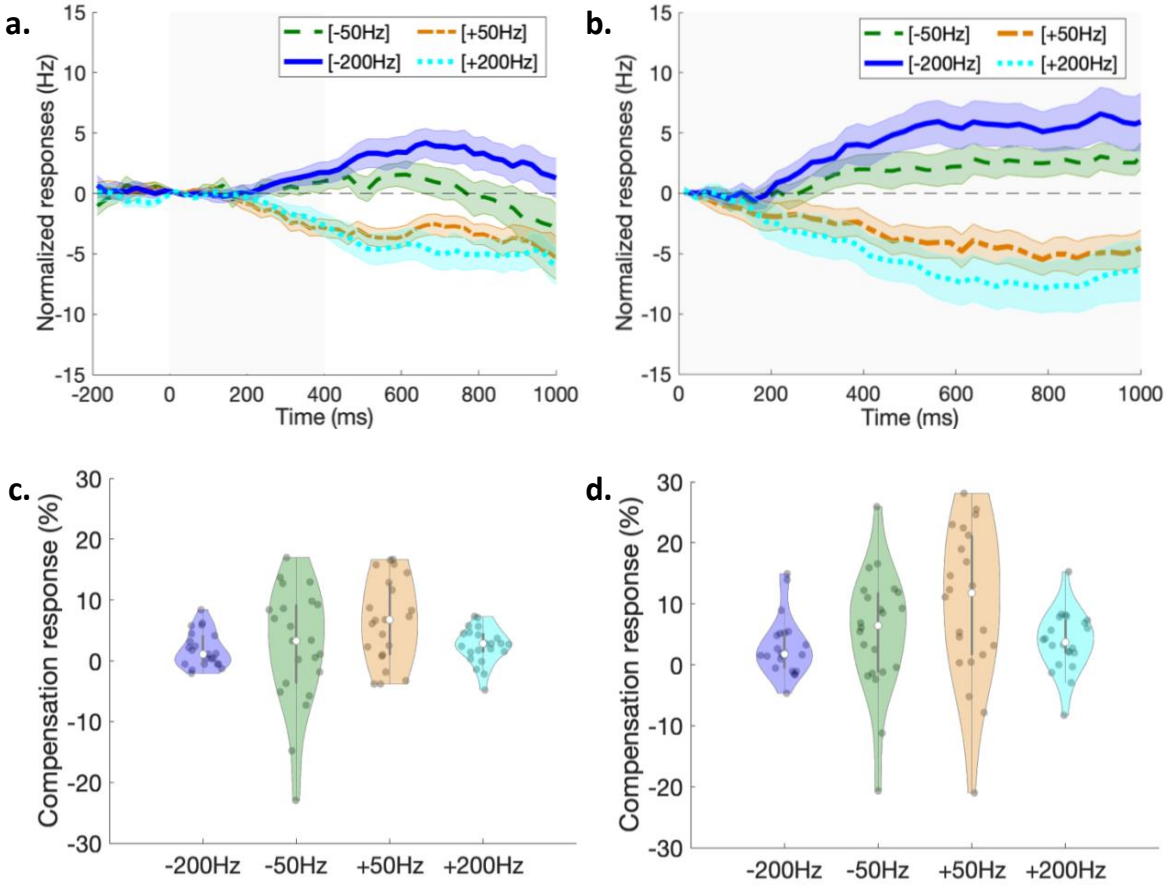

Figure 2.2: Online formant response to unpredictable mid-utterance and whole utterance formant perturbations. Normalized F1 responses averaged across participants within each condition are shown for (**a**) unpredictable transient mid-utterance and (**b**) unpredictable whole-utterance F1 perturbations. Time-range of perturbation are shown as shaded grey area (perturbation onset at t=0). Mean responses (lines) and SEM (shaded colored region) are shown. Responses to -50Hz perturbations are indicated with dashed line (green), +50Hz perturbations with dash-dot line (orange), -200Hz perturbations with solid line (blue) and +200Hz perturbations with dotted line (cyan). The distribution of peak compensation responses for each condition are shown as violin plots for (**c**) unpredictable transient mid-utterance and (**d**) unpredictable whole-utterance F1 perturbations. The grey bar indicates the range from the 1st to 3rd quartile and the white dot indicates the median. The shape of the violin plot reflects the kernel density estimate of the data, and the colored dots are actual individual response data points. Table 1 lists the p-values for significant differences from zero of the peak compensation responses.

Table 2.1: Latency and magnitude of responses to unpredictable perturbations for each condition. Response onset latency was calculated from the across-participant averaged responses. Peak compensations across participants (mean and SEM) were calculated in the 200ms window around the peak latency of the averaged group response. T-values (two-tailed) and p-values indicate the significant difference from zero of the peak compensations.

Perturbation type	Perturbation value (Hz)	Onset latency (ms)	Peak compensation (%) (Mean ± SEM)	T-values (DoF=21)	Significance level (p-value)
Mid	-50	450	2.36 ± 2.07	1.14	0.27
	+50	200	6.63 ± 1.47	3.97	1.94×10^{-4}
	-200	275	1.86 ± 0.62	3.01	6.66×10^{-3}
	+200	275	2.50± 0.63	3.97	6.92×10^{-4}
Whole	-50	625	5.34± 2.12	2.52	0.02
	+50	325	9.94 ± 2.66	3.73	1.23×10^{-3}
	-200	275	2.92 ± 1.02	2.88	9.02×10^{-3}
	+200	200	3.82 ± 1.03	3.70	1.33×10^{-3}

2.4.2 Within-trial compensations for unpredictable mid- and whole utterance F1 perturbations were significantly correlated

We found a positive correlation between responses to unpredictable mid-utterance and whole-utterance perturbations. Correlation was significantly positive when all four perturbation values were aggregated with participant as a random factor (Figure 3a: Regression coefficient estimate=0.34, p=0.02). When the correlation analysis was separately performed for each perturbation value, mid- and whole- utterance perturbation responses were significantly correlated only for responses to +200Hz and -200Hz perturbations (Figure 3b: Regression coefficient estimate=1.08, $p=6.23 \times 10^{-4}$ for -200Hz perturbations; Figure 3c: Regression coefficient estimate=0.35, p=0.10 for -50Hz perturbations; Figure 3d: Regression coefficient estimate=0.71, p=0.03 for +200Hz perturbations; Figure 3e: Regression coefficient estimate=-0.21, p=0.59 for +50Hz perturbations).

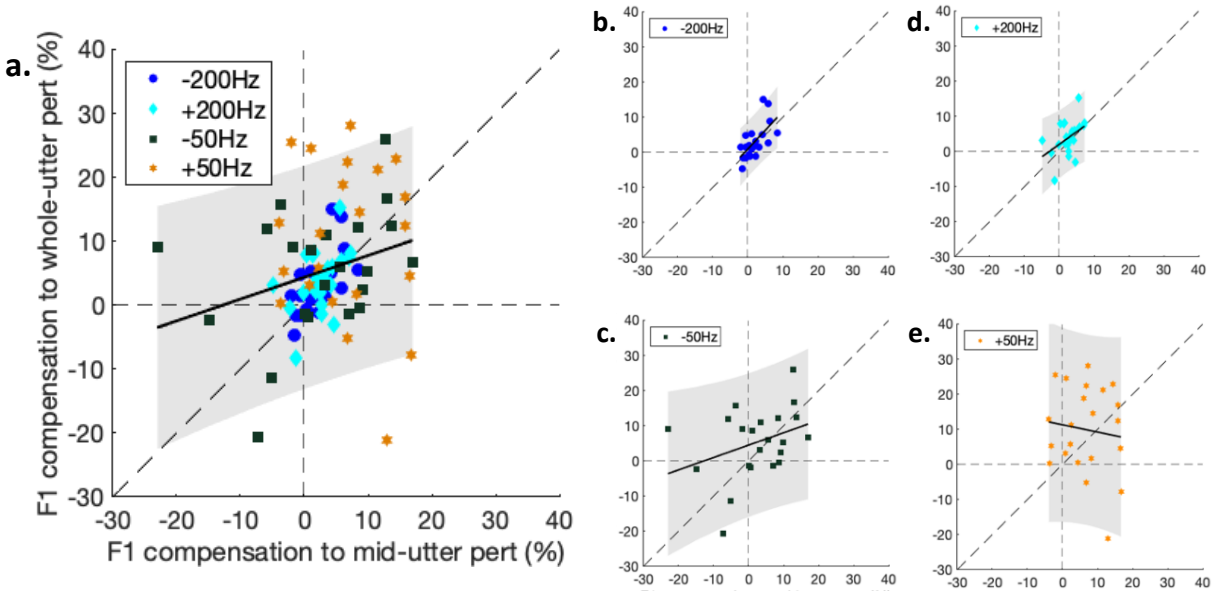

Figure 2.3: Compensatory responses to mid-utterance perturbations are correlated with responses to whole utterance perturbations. Scatterplot of peak compensation responses to transient mid-utterance and to whole-utterance formant perturbations for (**a**) all conditions and separately for (**b-e**) -200Hz, -50Hz, +200Hz and +50Hz perturbation conditions respectively.

Responses to -200Hz perturbations are indicated with circles, +200Hz perturbations with diamonds, -50Hz perturbations with squares, and +50Hz perturbations with hexagrams. The slope and 95% confidence interval of the correlations are indicated by the solid black line and grey shaded area, respectively. Dashed lines represent the coordinate axes and the diagonal with slope=1.

2.4.3. Participants adapted to consistent, predictable F1 perturbations

Consistent with what has been seen in earlier studies, participants adapted their initial feedforward responses to a consistent, predictable F1 perturbation across trials in the adaptation experiments. Sensorimotor adaptation response was quantified here as the initial 75ms F1 response in each trial that was then averaged for every 5 successive trials. Figure 4a shows the normalized sensorimotor adaptation response across trials of the adaptation experiments for +200 Hz (blue) and -200 Hz (cyan). The distributions of individual sensorimotor adaptation response percentages quantified in the late adaptation phase are shown as violin plots in Figure 4b and significance levels, when compared to zero, are shown in Table 2. A linear mixed effects model of sensorimotor adaptation responses showed no significant main effect for perturbation direction, or for the covariates of age or baseline F1 production ($F(1,20)=0.26$, $p=0.62$; $F(1,20)=0.03$, $p=0.86$; and $F(1,20)=0.23$, $p=0.64$ for direction, age and baseline F1 respectively). These findings show that the sensorimotor adaptation response to consistent, predictable F1 feedback perturbations can be observed within the initial feedforward response.

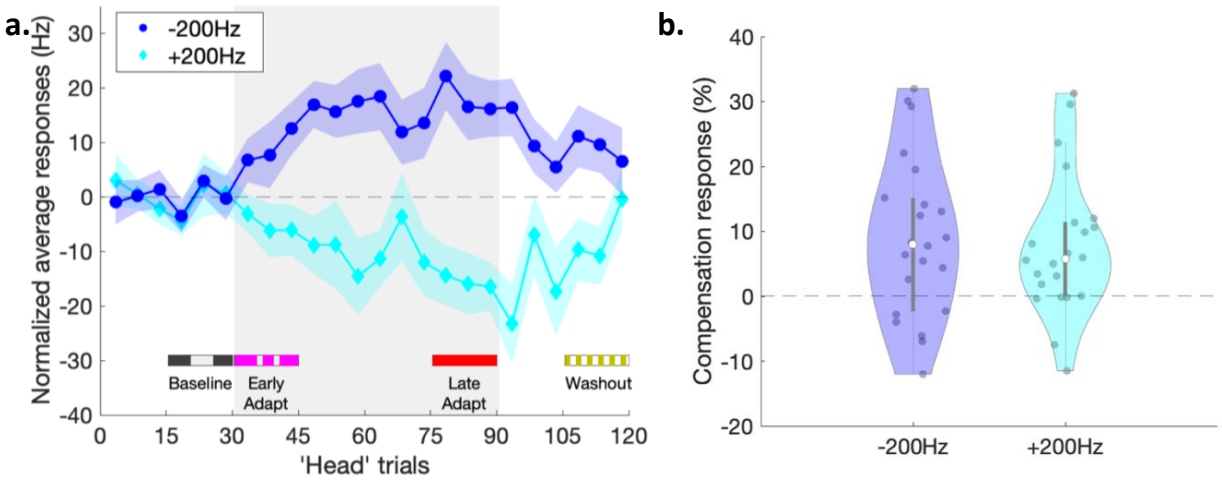

Figure 2.4: Sensorimotor adaptation results to consistent, predictable formant perturbations. (**a**) Average responses in the first 75ms of each trial to consistent, predictable formant perturbations across trials; average over 5 'head' trials (points joined by solid line) and standard error (shaded colored region) are shown. Responses to -200Hz and +200Hz are shown with circles and diamonds respectively. Shaded grey region represents trials where whole-utterance perturbation was consistently applied. Colored bars at the bottom indicate phases of trials in the adaptation experiment used for the within-trial time course analysis presented in Figure 6: baseline phase ('head' trials 16-30, dashed grey line), early adaptation phase ('head' trials 31-45, dash-dot magenta line), late adaptation phase ('head' trials 76-90, solid red line), and washout phase ('head' trials 106-120, dotted gold line). (**b**) The distributions of adaptation responses in the late adaptation phase are shown as violin plots. Table 2 lists the p-values for significant differences from zero of the sensorimotor adaptation responses.

2.4.4 Online compensation for unpredictable feedback perturbations and sensorimotor adaptation to consistent, predictable feedback perturbations were not correlated

Responses to unpredictable F1 feedback perturbations (online compensation) and to consistent, predictable F1 feedback perturbations (sensorimotor adaptation) were generally uncorrelated (Figure 5). Correlations between responses to mid-utterance perturbations and sensorimotor adaptation responses were all non-significant (Figure 5a: Regression coefficient estimate=-0.12, p=0.80 for all conditions aggregated with participants as a random effect; Figure 5b: Regression coefficient estimate=-0.53, p=0.56 for -200Hz perturbations; and Figure 5c: Regression

coefficient estimate=0.09, p=0.90 for +200Hz perturbations). Similarly, correlations between responses to whole-utterance perturbations and sensorimotor adaptation responses were also non-significant (Figure 5d: Regression coefficient estimate=-0.12, p=0.67 for all conditions aggregated with participant as a random effect; Figure 5e: Regression coefficient estimate=-0.04, p=0.94 for -200Hz perturbations; and Figure 5f: Regression coefficient estimate=0.33, p=0.48 for +200Hz perturbations).

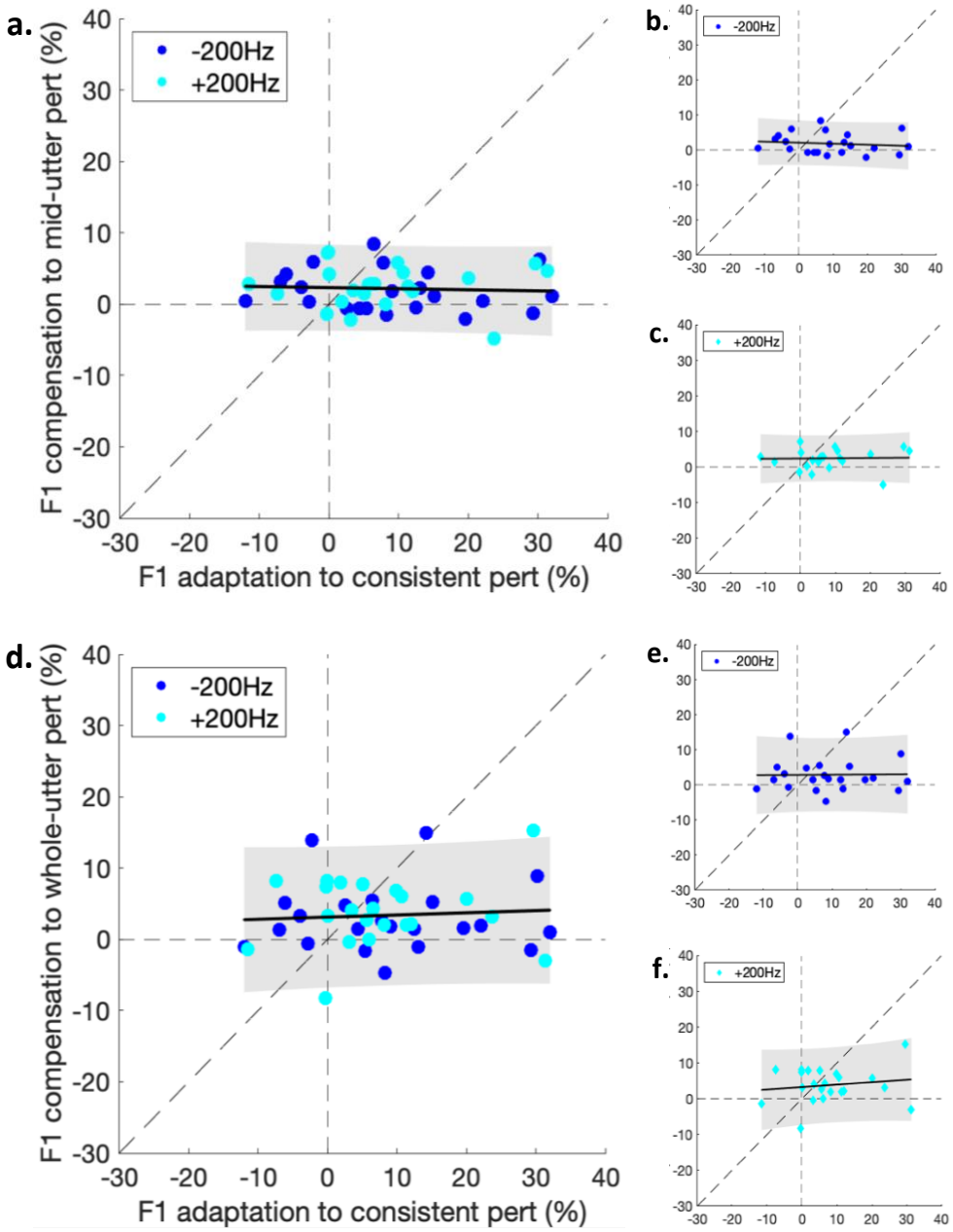

Figure 2.5: Online formant compensation is not associated with formant sensorimotor adaptation. Scatterplot of sensorimotor adaptation responses to consistent, predictable formant perturbations and compensation responses to unpredictable transient mid-utterance perturbations for (**a**) both conditions and separately for (**b-c**) -200Hz and +200Hz perturbation conditions respectively. Scatterplot of sensorimotor adaptation responses to consistent, predictable formant perturbations and online compensation responses to whole-utterance perturbations for (**d**) for both conditions and separately for (**e-f**) -200Hz and +200Hz perturbation conditions respectively. The slope and 95% confidence interval of the correlations are indicated by the solid black line and grey shaded area, respectively. Dashed lines represent the coordinate axes and the diagonal with slope=1.

Table 2.2: Sensorimotor adaptation responses across participants. Mean and SEM were calculated from the first 75ms of each of the last 15 'head' trials of the hold phase (late adaptation phase). T-values (two-tailed) and p-values indicate the significant difference from zero of the sensorimotor adaptation responses.

Perturbation value (Hz)	Adaptation Responses (%) (Mean ± SEM)	T-values (DoF=21)	Significance level (p-value)
-200	8.98 ± 2.63	3.41	2.63×10^{-3}
+200	7.67 ± 2.29	3.45	3.04×10^{-3}

2.4.5 Within-trial response dynamics in sensorimotor adaptation further revealed independence between online compensation responses and sensorimotor adaptation

We explored how the within-trial response dynamics changed over the course of the trials in the adaptation experiment. For this analysis, we examined the within-trial responses of participants who both positively compensated to whole-utterance perturbations and also positively adapted to consistent, predictable perturbations (Figure 6). This was done to ensure that the analysis would not be biased by participants who did not show a within-trial response (in the whole-utterance

perturbations). A total of 15 participants positively compensated to the +200Hz unpredictable whole-utterance perturbations and positively adapted to the +200Hz consistent, predictable perturbations. A total of 12 participants positively compensated to the -200Hz unpredictable whole-utterance perturbations and positively adapted to the -200Hz consistent, predictable perturbations. The within-trial responses shown (Figure 6a-b) were normalized to each participant's within-trial response in the baseline phase ('head' trials 1-30).

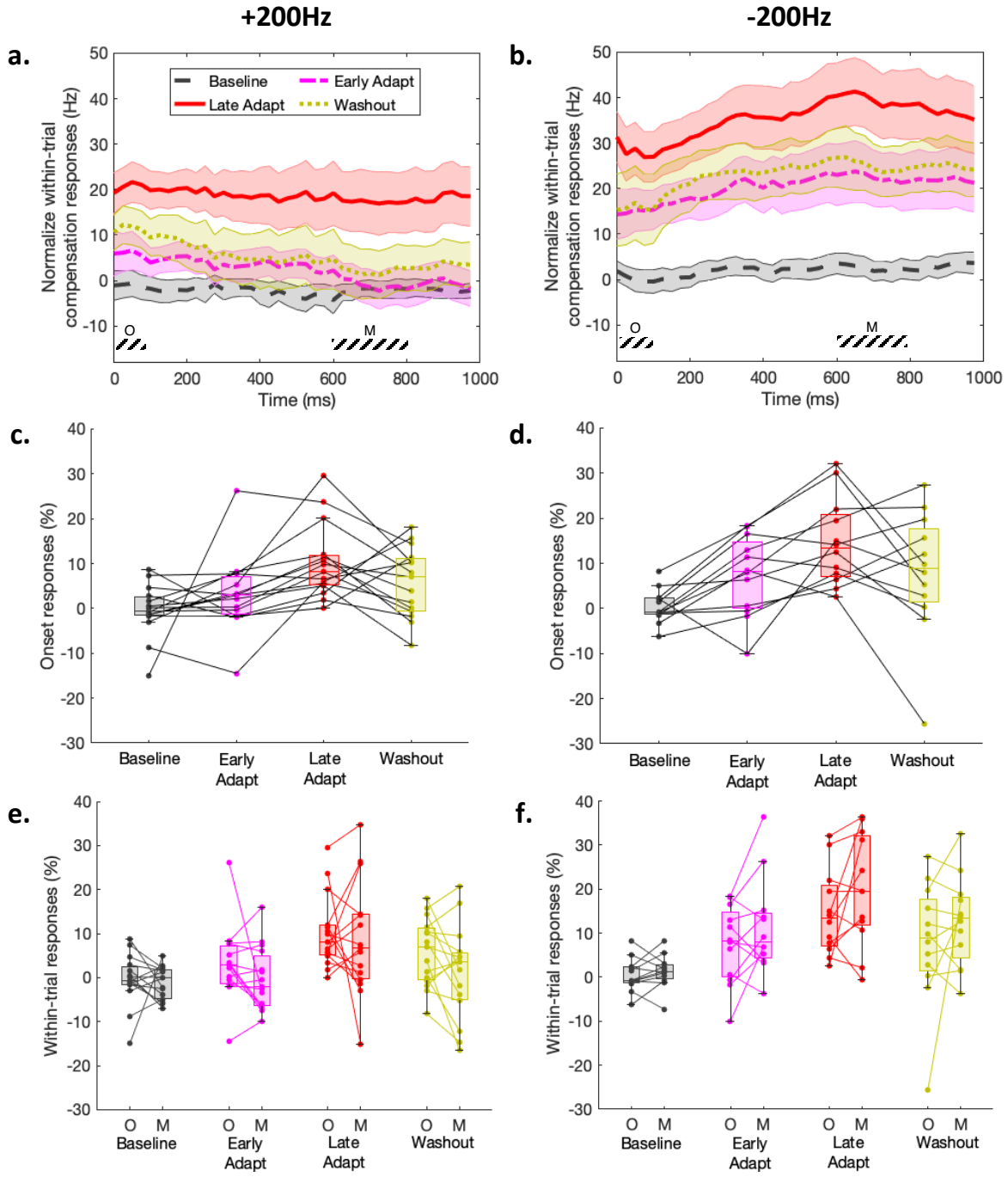

Figure 2.6: Evolution of within-trial formant time-course across the adaptation experiment. Normalized within-trial perturbation responses in different phases of the adaptation experiments with (**a**) +200Hz and (**b**) -200Hz consistent, predictable perturbations. Mean responses (lines) and SEM (shaded colored region) are shown. The colors represent responses in different ranges of 'head' trials: baseline ('head' trials 16-30, dashed grey), early adaptation ('head' trials 31-45, dash-dot magenta), late adaptation ('head' trials 76-90, solid red), and washout ('head' trials 106-120, dotted gold). These ranges are indicated as bars at the bottom of Figure 4a. The plots show both changes in initial response as well as within-trial response time course in different

phases of the adaptation experiments. Shaded bars at the bottom of the plots indicate time windows for onset ('O', 0-75ms) and mid-utterance ('M', 600-800ms) within-trial compensation response. (**c-d**) Boxplots show Onset (0-75ms) within-trial compensation responses distribution for +200Hz and -200Hz adaptation experiment respectively. Lines connect individual subjects' responses. (**e-f**) Pairwise onset ('O', 0-75ms) and mid-utterance ('M', 600-800ms) within-trial compensation response distribution for +200 Hz and -200 Hz adaptation experiment respectively. Lines connect individual subjects' responses.

This analysis showed changes even in the initial 75ms of the average within-trial time-course of each phase of the adaptation experiment, showing evidence of sensorimotor adaptation in the initial feedforward response. The distribution across participants of the initial (0-75ms) within-trial responses in each trial, expressed as percent compensation, are shown across phases of the adaptation experiment (Figure 6c-d). The initial 75ms response in the -200Hz adaptation experiment was not significantly above zero in baseline or washout phases (baseline phase: mean ± SEM=0.81 ± 2.19, $t(11)=0.37$, $p=0.72$; and washout phase: mean ± SEM=15.96 ± 8.03, $t(11)=1.99$, $p=0.07$) but was significantly above zero in the early and late adaptation phases (early adaptation phase: mean ± SEM=14.75 ± 5.12, $t(11)=2.88$, $p=0.02$; and late adaptation phase: mean ± SEM=29.25 ± 5.57, $t(11)=5.25$, $p=2.71\text{x}10^{-4}$). In contrast, in the +200Hz adaptation experiment, the initial 75ms response was not significantly above zero in baseline and early adaptation phases (baseline phase: mean ± SEM=0.86 ± 3.01, $t(14)=0.29$, $p=0.78$; and early adaptation phase: mean ± SEM=-6.22 ± 4.39, $t(14)=-1.42$, $p=0.18$) but was significantly above in late adaptation and washout phases (late adaptation phase: mean ± SEM=-20.51 ± 4.28, $t(14)=-4.79$, $p=2.87\text{x}10^{-4}$; and washout phase: mean ± SEM=-11.58 ± 3.97, $t(14)=-2.91$, $p=0.01$ respectively).

Comparisons of the late (600-800ms) within-trial responses to the initial 75ms within-trial responses over the phases of the adaptation experiment except the baseline phase showed that the

sensorimotor adaptation response occurred regardless of the existence of a within-trial compensatory response. The baseline phase was not included in this analysis because no perturbations had yet been applied in this phase, so within-trial responses were not expected. A linear mixed effects model for within-trial responses showed a significant main effect of phases (early adaptation, late adaptation and washout; $F(2,133)=11.91$, $p<0.0001$) and perturbation direction (-200Hz vs +200Hz; $F(1,133)=29.68$, $p<0.0001$) but not of within-trial time window (0-75ms vs 600-800ms from voice onset; $F(1,133)=0.25$, $p=0.62$), with a significant interaction between within-trial time window and perturbation direction ($F(1,133)=7.43$, $p=7.30\text{x}10^{-3}$). No significant interaction was found between phases and within-trial time window ($F(2,133)=0.17$, $p=0.84$), between phases and perturbation direction ($F(2,133)=0.06$, $p=0.94$) or between phases, within-trial time window and perturbation direction ($F(2,133)=0.07$), $p=0.93$). Given the significant main effect of perturbation direction, we then analyzed the within-trial responses to -200Hz and +200Hz consistent, predictable perturbations separately. Linear mixed effects model for -200Hz showed significant main effects of phases ($F(2,55)=5.42$, $p=7.10\text{x}10^{-3}$) and within-trial time window ($F(1,55)=4.59$, $p=0.04$) but no significant interaction between phases and within-trial time window ($F(2,55)=0.03$, $p=0.97$). Linear mixed effects model for +200Hz showed a significant main effect of phase ($F(2,70)=7.05$, $p=1.6\text{x}10^{-3}$), no significant main effect of within-trial time window ($F(1,70)=3.00$, $p=0.09$) and no significant interaction between phases and within-trial time window ($F(2,70)=0.25$, $p=0.78$).

2.4.6 F1 production in unpredictable perturbation sessions were not affected by experimental session order

We found no significant main effect of session order on the median baseline F1 production across the unpredictable perturbation sessions ($F(2,21)=0.41$, $p=0.67$). The second analysis was done on peak compensation responses in each of the 3 unpredictable perturbation sessions. Once again, we found no significant main effect of session order on peak compensation responses ($F(1,498)=0.02$, $p=0.90$). We also found no significant main effect of perturbation magnitude ($F(1,498)=1.29$, $p=0.26$) and type of perturbation onset ($F(1,498)=2.87$, $p=0.09$) but found a significant main effect of perturbation direction ($F(1,498)=6.84$, $p=9.2\text{x}10^{-3}$). These findings indicate that there were no significant carryover effects from the adaptation experiments to the following unpredictable perturbation sessions.

2.5 Discussion

In this study, we showed clear evidence for online auditory feedback control of formants. We observed that similar online auditory feedback compensation mechanisms are evident both at utterance onset and at mid-utterance. We also observed that auditory feedback compensation responses were not associated with sensorimotor adaptation of formants. These results have important implications for speech motor control models.

2.5.1 Transient auditory feedback perturbations induce similar responses in control of formants, pitch and loudness

We show that speakers produce short-latency compensatory responses to transient mid-utterance perturbations of F1 of the auditory feedback of their speech. The magnitude of these compensatory responses was significantly affected by the perturbation magnitude, where responses to 50Hz perturbations were larger than to 200Hz perturbations, consistent to what has been observed in other studies of formant perturbations (Daliri, Chao, & Fitzgerald, 2020; Katseff et al., 2012). The formant feedback perturbations we used had a similar time-course (abrupt with jittered onset) to the transient mid-utterance perturbations used in studies showing compensatory responses to pitch and loudness perturbations (Burnett et al., 1998; Heinks-Maldonado & Houde, 2005; Keough et al., 2013; Kort et al., 2016; Larson et al., 2001, 1999). Although mid-utterance formant perturbations have been investigated in a previous study, in that study the perturbation was always applied 300ms after voice onset and was cross-faded in linearly over 500ms (Purcell & Munhall, 2006b). Therefore, the perturbations used in the current study allowed for more direct comparison with pitch and loudness feedback perturbation responses. A striking similarity in responses to the transient mid-utterance perturbations across the speech features is the latency of response. In pitch and loudness perturbations, studies have found a response latency ranging around 150-250ms (Burnett et al., 1998, 1997; Heinks-Maldonado & Houde, 2005). In formant perturbations, the response latency ranges around 200-325ms. The response onset time similarity suggests a similar timescale for the auditory feedback processing of these different speech features.

While this may be considered surprisingly fast given the greater mass of the tongue and upper vocal tract when compared to the vocal folds, a recent study done by Bakst, S. (2017) that used ultrasound to track tongue movement during the production of varying CVC English words found that amount of tongue movement needed to make an acoustic change can be very small. Even though the tongue is a big muscle organ, it needs to be moved only by a very small amount to achieve a compensation response. The jaw may not technically be involved/needed for this compensation response to occur. Therefore, the size of the system may appear to be not an important factor in determining the reaction time which refers to only the starting point of response, representing how responsive the system is to the motor command. Consistent with this notion, a study by Pfister, et al. (2014) shows how larger organs like hand and foot even have a reaction time of about 300ms.

2.5.2 Feedback control of formants at utterance onset and at mid-utterance share a similar mechanism

We explored the possible differences in feedback control of formants at utterance onset and at mid-utterance by comparing online compensation responses to formant feedback perturbations applied for the whole utterance and transiently at mid-utterance. Previously, Hawco and Jones (2009) found a large difference (~50% vs ~17%) between the peak response to pitch feedback perturbations at utterance onset and at mid-utterance, and they suggested that the difference is due to different control mechanisms operating at utterance onset and at mid-utterance. In our study, we found slightly larger peak responses to formant perturbations at utterance onset (~5%) than at mid-utterance (~3%), and this difference was not statistically significant.

What can explain this difference in utterance onset and mid-utterance perturbation responses between pitch and formants? First, it should be noted that in the Hawco and Jones (2009) study, participants were given an external pitch reference at every trial, whereas in this study no such reference was provided. Second, other studies have shown the control of pitch may be special in its heightened reliance on auditory feedback for its control. In particular, studies of the speech of post-lingually deafened adults have shown that such deafness quickly degrades the control of pitch, while speech intelligibility can remain intact for years (Cowie & Douglas-Cowie, 1992; Houde &Nagarajan, 2015, p. 267-298; Lane & Webster, 1991). Third, it is possible that underlying mechanisms for the control of pitch and formants are different. Previous studies have found that individual with Parkinson's disease produced significantly larger compensation responses to pitch perturbations (X. Chen et al., 2013; Liu, Wang, Metman, & Larson, 2012) but smaller compensation responses to formant perturbations (Mollaei et al., 2016). These findings may reflect differences in the control of formant and pitch, perhaps where recruitment of sensorimotor regions may be greater for the control of formants that involve a larger set of muscles and associated sources of somatosensory feedback in the vocal tract than for pitch (Bouchard, Mesgarani, Johnson, & Chang, 2013; Huang et al., 2019; Mollaei et al., 2016).

In addition to responses to formant perturbations at onset and at mid-utterance not being significantly different in magnitude, we further showed a strong correlation between whole utterance and mid-utterance perturbation responses, suggesting a similar mechanism governing these two responses. The idea that onset and mid-utterance formant control are governed by a similar mechanism can be explained by the SFC model (Parrell & Houde, 2019) by making assumptions about the Kalman gain. In SFC, the Kalman gain is responsible for modulating the

corrective response to a mismatch between actual and expected feedback (in this case, brought upon by perturbations). Our results and those of Hawco and Jones (2009) could be explained by (1) the Kalman gain for pitch control being larger than it is for formant control, and (2) the Kalman gain being larger at utterance onset, where there is more production uncertainty than during continuation of the utterance. Therefore, our results suggest that the control mechanism throughout an utterance, whether it be at onset or at mid-utterance, is the same but potentially with varying magnitude of Kalman gain at different times (onset vs mid-utterance) and for different speech features (formants vs pitch).

2.5.3 Online compensation and sensorimotor adaptation are governed by different control mechanisms

Many speech motor control models acknowledge the existence of two speech control mechanisms: feedback and feedforward control (Guenther, 2016; Kearney et al., 2020; Parrell et al., 2019). Feedback control is responsible for maintaining speech production within an utterance, which is exhibited by the online compensation responses to unpredictable perturbations. In contrast, feedforward control is responsible for generating correct motor commands for an intended speech sound; feedforward motor commands are learned from past experiences and their execution does not depend on sensory feedback during production. However, the feedforward system can learn a new speech motor mapping if it consistently detects a mismatch between the actual and expected speech feedback over a long period of time, which leads to a sensorimotor adaptation response seen in numerous studies (Houde & Jordan, 1998; Jones & Munhall, 2000; Katseff et al., 2012; Purcell & Munhall, 2006a).

Despite being well-studied, we still do not have a clear understanding on how exactly the feedforward system is able to adapt to consistently perturbed feedback. As a result, models of speech differ in how feedback and feedforward control components are related. The DIVA model suggests that "…the feedforward control system constantly monitors the corrective commands generated by the feedback control system, gradually incorporating repeatedly occurring corrections into the feedforward command." (Guenther, 2016). Specifically, the DIVA model has four separate parameters governing: 1) responses to auditory feedback perturbations, 2) responses to somatosensory feedback perturbations, 3) incorporation of auditory feedback responses into the feedforward controller, and 4) incorporation of somatosensory feedback responses into the feedforward controller. A simpler 3-parameter version of this model, SimpleDIVA, uses a single parameter to govern the incorporation of both auditory and somatosensory feedback responses into the feedforward controller, while retaining the two feedback control parameters (Kearney et al., 2020). Thus, both DIVA and SimpleDIVA assume a close relationship between feedback and feedforward control. In contrast, the SFC model accommodates distinct mechanisms for feedback control and the learning of feedforward control, as adaptation can arise directly from sensory prediction errors rather than necessarily being derived from the incorporation of corrective movements as is assumed in DIVA (Parrell & Houde, 2019). In our study, we found three lines of evidence that the feedforward sensorimotor adaptation mechanism is distinct from the feedback control mechanism.

First, we found sensorimotor adaptation exceeded online compensation. We looked at the first 75ms formant responses for each trial in the adaptation experiments to purely isolate the feedforward responses. Our results are comparable to previous findings (Behroozmand &

Sangtian, 2018; Houde & Jordan, 1998; Katseff et al., 2012; Purcell & Munhall, 2006a), where participants on average showed significant sensorimotor adaptation to the consistent, predictable perturbation. Importantly, however, even within the first 15-20 trials, we saw that the sensorimotor adaptation response grew larger than the peak online compensation response. While this can be explained by both DIVA and SFC, the maximum asymptotic sensorimotor adaptation response could have not exceeded the maximum asymptotic online compensation response in SimpleDIVA.

Second, we found that online compensation and sensorimotor adaptation responses were not significantly correlated across participants, as was also found in a previous study (Franken, Acheson, McQueen, Hagoort, & Eisner, 2019). In fact, we found that some participants were able to adapt even though they did not compensate for the unpredictable perturbations and vice versa.

Third, we found that sensorimotor adaptation in the feedforward response was able to take place regardless of the within-trial response pattern. We were able to examine the dynamics of the adaptation response through a novel within-trial time-series analysis of the adaptation response. We found that responses to -200Hz consistent, predictable perturbations showed some evidence of within-trial compensation which could also be seen in the late phase of the adaptation experiment, superposed on a feedforward adaptation response that started at the beginning of each trial. However, for responses to +200Hz consistent, predictable perturbations, we did not find evidence for online compensation either in the early or late phases of the adaptation experiment, but we found clear evidence for the development of a feedforward sensorimotor

adaptation response that started at the beginning of each trial. Together, the evidence is more consistent with the idea that online compensation and sensorimotor adaptation are driven by distinct neural mechanisms.

It is likely that what we have learnt in this study is also applicable to dynamic speech. This study examined the effects of feedback perturbations on the production of static vowels, which lack the temporal dynamics found in natural speech. However, even though producing static vowels and dynamically changing speech are different speech tasks, there is no evidence that mechanisms of processing and learning from auditory feedback perturbations are different between these tasks. In fact, there is clear evidence that the feedback processing phenomena we observed here are also at work during natural dynamic speech production where both online compensation and sensorimotor adaptation have been observed (Lametti, Smith, Watkins, & Shiller, 2018; Liu et al., 2009).

During the review process of the current manuscript, we became aware of a recent publication from Lester-Smith, et al. (2020) which also explored the relationship between within-trial compensation responses and across-trial sensorimotor adaptation responses to F1 perturbations. Although they reported a significant positive correlation between within-trial compensation responses and across-trial sensorimotor adaptation responses to F1 perturbations, these results were acknowledged as potentially non-significant if corrections for multiple comparisons were applied, which would be consistent with our current findings. The study also differed from the current study in terms of experimental design in many ways, including prompt words, experimental session order, study population, perturbation magnitudes, and the time windows

used in their data analysis. Therefore, it is difficult to directly compare the results of that study with our findings, and future studies are warranted to explore and reconcile these discrepancies.

2.5.4 Directional asymmetry in compensation responses

An interesting observation in our current study is the effect of perturbation direction on responses to both unpredictable within-trial and consistent, predictable across-trial perturbations. We found a significant main effect of perturbation direction in the within-trial responses to unpredictable within-trial perturbations. Specifically, we found a smaller compensation response average to negative perturbations (3.12%) than to positive perturbations (5.72%). However, we did not find a significant main effect of direction in either the within-trial or across-trial responses to consistent, predictable across-trial perturbations (adaptation experiments). Previous studies have found a directional asymmetry in responses to both unpredictable within-trial and consistent, predictable across-trial perturbations (Liu, Meshman, Behroozmand, & Larson, 2011; Mitsuya, MacDonald, Munhall, & Purcell, 2015). Liu (2011) found a directional asymmetry in responses to unpredictable transient mid-utterance pitch perturbations, where responses to negative pitch perturbations showed larger responses than to positive pitch perturbations. Mitsuya (2015) found a directional asymmetry in responses to consistent, predictable across-trial F1 perturbations, where the adaptation magnitude is smaller in the negative direction when using the vowels /i/, /I/ and /ɛ/ (but not when using the vowels /æ/, /ɔ/ or /u/). Across all these studies, there does not seem to be any consistent trend in terms of how the compensation responses were affected by the perturbation direction. Furthermore, many other factors can result in the observed asymmetries in these studies, including articulatory constraints leading to tradeoffs between auditory and somatosensory feedback and perceptual nonlinearities, as well as variation in

somatosensory feedback for different directions of articulatory change (Kothare, et al., 2020, in press). Therefore, whether there is a consistent directional asymmetry remains unclear.

2.6 Conclusions and future directions

In this study, we showed that transient mid-utterance feedback perturbations induced similar responses in formants compared to what have been shown previously in pitch and loudness. This compensation to transient formant perturbations at mid-utterance was highly correlated with compensation to whole utterance formant perturbations, suggesting that these compensations are governed by a similar mechanism. We also found evidence suggesting that online compensation and sensorimotor adaptation are governed by distinct mechanisms. Further studies need to be performed to investigate the underlying neural bases of this mechanistic difference between sensorimotor adaptation of the feedforward control and online compensation in feedback control.

2.7 References

Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLOS ONE*, *13*(1), e0191839. Retrieved from https://doi.org/10.1371/journal.pone.0191839

Bakst, S. (2017). Articulation and Altered Auditory Feedback. Ph.D. Dissertation, University of California, Berkeley. Retrieved from https://digitalassets.lib.berkeley.edu/etd/ucb/text/Bakst_berkeley_0028E_17361.pdf

Bauer, J. J., Hain, T. C., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, *119*(4), 2363–2371. https://doi.org/10.1121/1.2173513

Behroozmand, R., & Sangtian, S. (2018). Neural bases of sensorimotor adaptation in the vocal motor system. *Experimental Brain Research*, *0*(0), 0. https://doi.org/10.1007/s00221-018-5272-9

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161–1165. https://doi.org/10.1038/nature03867

Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–332. https://doi.org/10.1038/nature11911

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Burnett, T. A., Senner, J. E., & Larson, C. R. (1997). Voice F0 responses to pitch-shifted

auditory feedback: a preliminary study. *Journal of Voice*, *11*(2), 202–211. https://doi.org/10.1016/S0892-1997(97)80079-3

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE*, *7*(7), 1–13. https://doi.org/10.1371/journal.pone.0041830

Caudrelier, T., & Rochet-Capellan, A. (2019). Changes in speech production in response to formant perturbations: An overview of two decades of research. In S. Fuchs, J. Cleland, & A. Rochet-Capellan (Authors), Speech production and perception: Learning and memory (pp. 15-75). Bern, Switzerland: Peter Lang D.

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

Chen, X., Zhu, X., Wang, E. Q., Chen, L., Li, W., Chen, Z., & Liu, H. (2013). Sensorimotor control of vocal pitch production in Parkinson's disease. *Brain Research*, *1527*, 99–107. https://doi.org/10.1016/j.brainres.2013.06.030

Daliri, A., Chao, S.-C., & Fitzgerald, L. C. (2020). Compensatory Responses to Formant Perturbations Proportionally Decrease as Perturbations Increase. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3392–3407. https://doi.org/10.1044/2020_JSLHR-19-00422

Demopoulos, C., Kothare, H., Mizuiri, D., Henderson-Sabes, J., Fregeau, B., Tjernagel, J., Houde, J. F., Sherr, E. H., & Nagarajan, S. S. (2018). Abnormal speech motor control in individuals with 16p11.2 deletions. *Scientific Reports*, *8*(1), 1–10.

https://doi.org/10.1038/s41598-018-19751-x

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2019). Consistency influences altered auditory feedback processing. *Quarterly Journal of Experimental Psychology*, *72*(10), 2371–2379. https://doi.org/10.1177/1747021819838939

Guenther, F. H. (2016). *Neural Control of Speech*. The MIT Press.

Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain Research*. https://doi.org/10.1016/j.brainres.2009.04.033

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*, *6*(3), 131–137. https://doi.org/10.1121/1.1931747

Houde, J. F., & Jordan, M. I. (1998). *Sensorimotor adaptation in speech production. Science, 279(5354), 1213-1216..pdf. 1*.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*(October), 1–14. https://doi.org/10.3389/fnhum.2011.00082

Houde, J., & Nagarajan, S. S. (2015). Auditory Feedback. In M. A. Redford (Author), The handbook of speech production (pp. 267-297). Malden, MA: Wiley Blackwell.

Huang, X., Fan, H., Li, J., Jones, J. A., Wang, E. Q., Chen, L., … Liu, H. (2019). External cueing facilitates auditory-motor integration for speech control in individuals with Parkinson's disease. *Neurobiology of Aging*, *76*, 96–105. https://doi.org/https://doi.org/10.1016/j.neurobiolaging.2018.12.020

Ito, T., Coppola, J. H., & Ostry, D. J. (2016). Speech motor learning changes the neural response to both auditory and somatosensory signals. *Scientific Reports*, *6*(1), 25926.

https://doi.org/10.1038/srep25926

Jenkins 3rd, J., Idsardi, W. J., & Poeppel, D. (2010). The analysis of simple and complex auditory signals in human auditory cortex: magnetoencephalographic evidence from M100 modulation. *Ear and Hearing*, *31*(4), 515–526. https://doi.org/10.1097/AUD.0b013e3181d99a75

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *J. Acoust. Soc. Am.*, *108*(3 Pt 1), 1246–1251. https://doi.org/10.1121/1.1288414

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308. https://doi.org/10.1177/0023830911417802

Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., … Guenther, F. H. (2020). A Simple 3-Parameter Model for Examining Adaptation in Speech and Voice Production . *Frontiers in Psychology* , Vol. 10, p. 2995. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.02995

Keough, D., Hawco, C. S., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience*, *14*(1), 25. https://doi.org/10.1186/1471-2202-14-25

Kim, K. S., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research : JSLHR*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_jslhr-19-00419

Kort, N. S., Cuesta, P., Houde, J. F., & Nagarajan, S. S. (2016). Bihemispheric network

dynamics coordinating vocal feedback control. *Human Brain Mapping*, *37*(4), 1474–1485. https://doi.org/10.1002/hbm.23114

Kothare, H., Raharjo, I., Ramanarayanan, V., Ranasinghe, K., Parrell, B., Johnson, K., Houde, J. F., & Nagarajan, S. S. (2020). Sensorimotor adaptation of speech depends on the direction of auditory feedback alteration. In press.

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Current Biology*, *28*(19), 3106-3113.e2. https://doi.org/10.1016/j.cub.2018.07.030

Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., & Ostry, D. J. (2014). Plasticity in the human speech motor system drives changes in speech perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(31), 10339–10346. https://doi.org/10.1523/JNEUROSCI.0108-14.2014

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America*, *89*(2), 859–866. https://doi.org/10.1121/1.1894647

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, *110*(6), 2845–2848. https://doi.org/10.1121/1.1417527

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (1999). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America*, *107*(1), 559–564.

https://doi.org/10.1121/1.428323

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862–2872. https://doi.org/10.1121/1.2715657

Lester-Smith, R. A., Daliri, A., Enos, N., Abur, D., Lupiani, A. A., Letcher, S., & Stepp, C. E. (2020). The Relation of Articulatory and Vocal Auditory–Motor Control in Typical Speakers. Journal of Speech, Language, and Hearing Research, 63(11), 3628–3642. https://doi.org/10.1044/2020_JSLHR-20-00192

Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*, *122*(5), 951–957. https://doi.org/10.1016/j.clinph.2010.08.010

Liu, H., Wang, E. Q., Metman, L. V., & Larson, C. R. (2012). Vocal responses to perturbations in voice auditory feedback in individuals with parkinson's disease. *PLoS ONE*, *7*(3). https://doi.org/10.1371/journal.pone.0033629

Liu, H., Xu, Y., & Larson, C. R. (2009). Attenuation of vocal responses to pitch perturbations during Mandarin speech. *The Journal of the Acoustical Society of America (JASA)*, *125*(4), 2299–2306. https://doi.org/10.1121/1.3081523

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068. https://doi.org/10.1121/1.3278606

Mitsuya, T., MacDonald, E. N., Munhall, K. G., & Purcell, D. W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of America*, *138*(1), 413–424. https://doi.org/10.1121/1.4923154

Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research*, *1646*, 269–277. https://doi.org/10.1016/j.brainres.2016.06.013

Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2019). The Relationship Between Speech Perceptual Discrimination and Speech Production in Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, *62*(12), 4256–4268. https://doi.org/10.1044/2019_JSLHR-S-18-0425

Mollaei, F., Shiller, D. M., & Gracco, V. L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*, *28*(12), 1668–1674. https://doi.org/10.1002/mds.25588

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *The Journal of Neuroscience*, 3316–3363. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Parrell, B., & Houde, J. F. (2019). Modeling the Role of Sensory Feedback in Speech Motor Control and Learning. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 2963–2985. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology*, *15*(9), e1007321.

Pfister, M., Lue, J.-C. L., Stefanini, F. R., Falabella, P., Dustin, L., Koss, M. J., & Humayun, M. S. (2014). Comparison of reaction response time between hand and foot controlled devices in simulated microsurgical testing. BioMed Research International, 2014, 769296.

https://doi.org/10.1155/2014/769296

Poeppel, D., & Hickok, G. (2015). Chapter 14 - Electromagnetic recording of the auditory system. In M. J. Aminoff, F. Boller, & D. F. B. T.-H. of C. N. Swaab (Eds.), *The Human Auditory System* (Vol. 129, pp. 245–255). https://doi.org/https://doi.org/10.1016/B978-0-444-62630-1.00014-7

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(January 2006), 2288–2297. https://doi.org/10.1121/1.2173514

Reilly, K. J., & Dougherty, K. E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4812763

Scheerer, N. E., & Jones, J. A. (2018). The Role of Auditory Feedback at Vocalization Onset and Mid-Utterance. *Frontiers in Psychology*, *9*, 2019. https://doi.org/10.3389/fpsyg.2018.02019

Shadmehr, R., Smith, M. A., & Krakauer, J. W. (2010). Error Correction, Sensory Prediction, and Adaptation in Motor Control. *Annual Review of Neuroscience*, *33*(1), 89–108. https://doi.org/10.1146/annurev-neuro-060909-153135

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *The Journal of the Acoustical Society of America*, *125*(2), 1103–1113. https://doi.org/10.1121/1.3058638

Chapter 3: Neural responses to pitch and formant feedback perturbations

3.1 Abstract

The control of multiple speech features, such as pitch and formants, is important for clear and distinguishable speech. Previous studies have shown that auditory feedback information is important to control for such speech features. Imaging studies with fMRI further investigated the neural network responsible for the feedback processing of pitch and formants, while studies with MEG and ECoG have only investigated the feedback processing of pitch. No study has directly examined either the common regions in feedback processing of both pitch and formants, nor the feature-specific regions in feedback processing of pitch vs formants. Furthermore, previous MEG and ECoG studies have only looked at feedback processing in the high-gamma frequency band. The current study investigated the neural network for feedback processing of pitch and formants with high temporal resolution by applying auditory feedback perturbations to the pitch or formants of subjects' speech feedback while recording neural responses with MEG. The results showing the neural network for feedback processing of formants were then compared to the neural network for feedback processing of pitch, which was collected from the same set of subjects; this allowed for a more direct within-subject comparison of pitch vs formant feedback processing. Neuro-behavioral correlations were performed to inspect regions that are likely responsible for driving compensation responses to feedback perturbations by correlating compensation responses to the neural responses to each pitch and formant feedback perturbation.

The results revealed a wide-spread neural activation in response to pitch and formant perturbations across all frequency bands, with opposite lateralization for pitch and formants across different frequency bands. Findings from this study should be considered in the neural-based models of speech motor control, specifically regarding regions responsible for the feedback control mechanism.

3.2 Introduction

Speech production involves the control of several acoustic features. Among several acoustic features, two of the most commonly studied features include pitch and formants. In English, formant production controls for segmental features that determine phonological identity (vowels/consonants) and thus also semantic (words) content, whereas pitch production controls for suprasegmental features such as intonation and stress. Multiple speech studies have demonstrated the role of auditory feedback in the online (within-utterance) control for both pitch and formants using auditory feedback perturbations, where either pitch or formants in the auditory feedback were perturbed, leading to what's known as a compensation response (Burnett et al., 1998; Cai et al., 2012; Chen, Liu, Xu, & Larson, 2007; Hawco & Jones, 2009; Houde, J. F., & Jordan, 1998; Keough et al., 2013; Larson et al., 2001, 1999; Purcell & Munhall, 2006b; Reilly & Dougherty, 2013). Speech production models theorized that the ability to compensate for speech errors are generated by a feedback control mechanism, which 1) compares the feedback information to an internal representation of expected feedback and 2) sends necessary motor correction commands to the speech articulators when a mismatch is detected (Guenther, 2016; Houde & Nagarajan, 2011; Kearney et al., 2020; Parrell et al., 2019). Despite the general

agreement on the existence of a feedback control mechanism, the neural regions responsible for each step of the feedback control is not fully understood.

Several studies have investigated the neural network involved in the feedback control of pitch and formants in speech using fMRI and MEG. Neural responses to pitch perturbations have been studied both using fMRI and MEG (Kort et al., 2016; Parkinson et al., 2012; Ranasinghe et al., 2019), while formant perturbations have been studied only using fMRI (Niziolek & Guenther, 2013; Tourville et al., 2008). However, no study has directly examined either the common regions in feedback processing of both pitch and formants, nor the feature-specific regions for pitch vs formants feedback processing (i.e., regions dominant for pitch feedback processing and vice versa). Studies with fMRI lacked the temporal resolution MEG has to offer (Hari et al., 2000; Singh, 2014), and previous MEG and ECoG studies have only looked at feedback processing in the high-gamma frequency band (Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Kort et al., 2016). Furthermore, the perturbations applied in previous studies were not exactly comparable: the pitch perturbations used were transiently applied mid-utterance, while the formant perturbations used were applied for the whole-utterance. Therefore, investigating neural responses to both pitch and formant perturbations across all frequency bands using temporally similar perturbations in the MEG would be ideal in order to uncover similarities and differences between the different steps of neural feedback processing for pitch and formants.

In this study, we investigated the neural responses to pitch and formant perturbations to elucidate the common (shared across both speech features) regions and feature-specific (dominant for each speech feature) regions involved in feedback processing of pitch and formants. We applied

transient mid-utterance pitch and formant perturbations to participants during MEG. We expected compensation responses to be comparable as observed in previous pitch and formant perturbation studies. We then examined the neural activity shared between responses to pitch and formant perturbations in two different ways, one way by combining responses to both pitch and formant perturbation and examining the overall neural activity, and another by overlapping regions from the separate neural activity responses to pitch and formant perturbations. We then contrasted the neural activity responses to pitch and formant perturbations to investigate regions that are more involved in either pitch or formant feedback processing. Finally, we examined which neural regions are involved in driving compensation responses to pitch and formant perturbations using correlation analysis. We explored whether the regions that are likely to drive compensation responses are feature specific (involved specifically for either pitch or formant processing) or are involved in overall feedback processing.

3.3 Methods

3.3.1 Participants

Healthy participants were recruited for the study (n=16, 2 females and 1 gender non-conforming). Participants' ages ranged from 24 to 61 years at the time of data collection (mean ± standard deviation of 37.47 ± 11.42 years). All participants self-reported to be native English speakers with no deficits in learning, motor or speech and language abilities. Participants were recruited through online platforms, pamphlets posted around the UCSF community and by word of mouth, gave written informed consent to participate, and passed the audiometric testing. The study was approved by the UCSF Institutional Review Board for human research.

3.3.2 Experimental design and procedures

Participants performed a feedback perturbation experiment during whole head MEG neural recording while lying in a supine position. The experiment consisted of three blocks of 135 trials each. For every trial, participants were instructed to phonate the word 'EH' as long as the word appeared on the screen (approximately 1s). The participants phonated into an MEG-compatible optical microphone (Optimic MEG, Optoacoustics, http://www.optoacoustics.com/medical/optimic-meg/features) and received auditory feedback through an MEG-compatible earply earphones (ER-3A insert headphone, Etymotic, http://www.etymotic.com/auditory-research/accessories/er3a.html). Auditory feedback output level was set to be attenuated by 30 dB and was adjusted if participants feel uncomfortable (this was uncommon). During the phonation, participants heard either normal (unperturbed) or perturbed feedback to their speech. The perturbations were applied transiently for 400ms with a jittered onset between 200-500ms, making the perturbations unpredictable. Four types of unpredictable perturbations were applied: +100 cents pitch perturbation, -100 cents pitch perturbation, +200 Hz formant perturbation, and -200 Hz formant perturbation. The 100 cents pitch perturbation magnitude was selected as it has previously been shown to induce a robust behavioral and neural response. The 200 Hz formant perturbation was selected as it has previously been often used to induce robust behavioral response, though its neural response has yet been investigated. There were 81 trials with unperturbed feedback and 81 trials for each type of unpredictable perturbation, totaling to 405 trials that were equally randomized to the three blocks. To ensure that participants did not exhaust themselves, breaks were placed in between trials (random-length, 1.7-2.7s), after every 15 trials (self-paced), and between the experimental blocks.

3.3.3 Speech alteration apparatus

The Feedback Utility for Speech Production (FUSP) software was used to perform the auditory feedback perturbations. FUSP repeatedly analyzed 3ms frames of speech input from the microphone into separate pitch and formant representations that were, at times, altered (depending on the trial) and used to synthesize the next 3ms of speech output to the participant's headphones. The speech data was recorded at a rate of 11025Hz and this feedback processing introduced an imperceptible ~21ms delay in the auditory feedback, as measured following the methods outlined by Kim, et al. (Kim, Wang, & Max, 2020).

3.3.4 Speech data analysis

All acoustic speech data was analyzed using Wave Viewer, a custom-built MATLAB-based speech analysis software (https://github.com/SpeechNeuroscienceLab/Wave-Viewer). In each trial, pitch and formants were tracked using linear predictive coding (LPC). The tracking for pitch and the first formant was further refined by manual screening, as needed, to exclude bad trials e.g., trials with no or insufficient (<1 second) speech response, interruption in speech production/recording, and poor pitch (F0) and formant (F1) tracking.

Using the raw F0 and F1 track trial data extracted, we calculated the within-trial response time courses to each pitch and formant perturbations. We first converted each trial responses to compensation in cents using the following formula: $Compensation\ responses\ (cents) = multipler\ \times 1200 \times log_2(\frac{responses\ (Hz)}{baseline\ (Hz)})$ where $multipler$ represents a -1 multiplier for responses to positive perturbations (and +1 for responses to negative perturbations) and $baseline\ (Hz)$ represents the average response in Hz in the baseline window (-50 to +50ms

from perturbation onset). Each trial was then smoothed by averaging the F0/F1 responses within non-overlapping 25ms windows; this was done to reduce the pitch/formant tracking variations across frames. Responses to both perturbation directions for each F0 and F1 perturbations were then aggregated to reflect the imaging analysis (discussed later), which was done to reduce noise in the data, and then averaged to obtain the F0 and F1 response time-course for each participant, excluding outlier trials (trials with responses exceeding mean±3 standard deviations). Finally, the average and standard error of F0 and F1 response time-courses were calculated across participants.

To obtain the distribution of compensation responses for each F0 and F1 perturbation conditions, we first calculated each participant's peak response by averaging their F0/F1 time-course response in a 200ms time window around the group peak response latency for each condition. Violin plots were created using the peak compensation responses, and a one-sample two-tailed t-test for each condition was calculated to test for significance different from zero.

3.3.5 MEG recording

Participants performed the experiments during whole head MEG neural recording while lying in the supine position. The MEG system (CTF, Coquitlam, British Columbia, Canada) consists of 275 axial gradiometers with a data recording sampling rate of 1200 Hz. Three fiducial coils were placed on the nasion, left and right preauricular points to triangulate the participant's head position relative the MEG sensory array. These fiducial markers were then coregistered with participant's anatomical MRI, obtained in a separate session, to generate head shape.

3.3.6 MEG data preprocessing

The MEG sensor data were manually marked for speech onset and speech offset events in each trial, and automatically marked at perturbation onset by FUSP triggers. Third gradient noise correction filters and DC offset based on the whole trial were applied to the data. Data was then visually inspected to remove trials with artifacts (indicated by abnormally large and/or noisy signals due to EMG, head movement, eye blinks or saccades). Sensor data was notch filtered around 120Hz with a width of 4Hz. Only trials with a minimum of 1 second between speech onset and offset and 200ms between speech onset and perturbation onset were included. Trials with opposite direction but same type perturbation (pitch or formant) were aggregated into the same group, resulting in two main perturbation conditions (100 cent pitch and 200 Hz formant perturbations) used for the subsequent analyses.

3.3.7 MEG data analysis

The NUTMEG toolbox was the main tool used for the MEG analysis (Dalal et al 2011). We focused our analysis on the following frequency bands: theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz), low gamma (30-55 Hz) and high gamma (65-150 Hz) bands. A broad selection of frequency bands was selected given that it is still unclear which frequency band(s) reflect findings in fMRI (Hall, Robson, Morris, & Brookes, 2014). We created a total of 5 neural activation maps to reflect responses to: (1) pitch perturbations, (2) formant perturbations, (3) either pitch or formant perturbations, (4) both pitch and formant perturbations, and (5) formant compared to pitch perturbations.

To obtain the neural activation map for responses to each pitch and formant perturbations compared to pre-perturbation responses, we performed time-frequency analyses using a time-frequency optimized spatially adaptive filtering algorithm implemented in NUTMEG (Dalal et al., 2008) to localize induced activity in the different frequency bands. For each participant and each condition (pitch and formant perturbations), a multisphere lead field was first calculated for every 8mm voxel in the brain. Source localization was then calculated using both lead field and sensor covariance. Noise-corrected pseudo-F ratios were computed between the active windows (post perturbation onset) and the prestimulus control baseline (200ms window pre perturbation onset). The size of analysis window for each frequency bands is as follow: (1) theta band: 400 ms with an overlap of 100 ms, (2) alpha band: 400 ms with an overlap of 100 ms, (3) beta band: 200 ms with an overlap of 50 ms, (4) low gamma band: 200 ms with an overlap of 50 ms, and (5) high gamma band: 100 ms with an overlap of 25 ms. We examined the 50-950ms post perturbation responses with 100ms intervals across all frequencies. The time-frequency beam former results for each participant and each condition were then normalized using each participant's normalized MRI parameters to the MNI template brain using SPM8 (standard Montreal Neurological Institute (MNI) template, statistical parametric mapping (SPM8) http://www.fil.ion.ucl.ac.uk/spm/software/spm8/]. Average and variance maps were calculated for each time-frequency point and smoothed using a Gaussian kernel with a half-width of 20x20x20 mm3 (Barnes, Hillebrand, Fawcett, & Singh, 2004). Using this, group statistics for each pitch and formant perturbations were computed using the NUTMEG time-frequency statistics toolbox with statistical nonparametric mapping (SnPM) (Dalal et al., 2011). Specifically, a pseudo-t statistic was obtained for each voxel and time window using nonparametric null distribution obtained through permutations. The resulting images reflect the t-

values from said statistics for each pitch and formant perturbations and were cluster-corrected to 30 voxels, $p<0.05$ such that only clusters with 30 contiguous voxels with $p<0.05$ were shown (Dalal et al., 2011; Kort et al., 2016).

To obtain the neural activation map for the responses to either pitch and formant perturbations, we performed a similar time-frequency analysis as described above but with multisphere lead field calculated across all trials in both pitch and formant perturbation conditions. The resulting images were cluster-corrected to 40 voxels, $p<0.05$. To obtain the neural activation map for responses to both pitch and formant perturbations, we mapped regions that were overlapping and have both positive or negative t-statistics in the pitch and formant activation response maps. To obtain the neural activation map for responses to formant compared to pitch perturbations, a paired t-test was performed between pitch and formant perturbation average and variance maps. The resulting images were cluster-corrected to 30 voxels, $p<0.05$.

Finally, to obtain neurobehavioral correlations comparing individual participant's compensation responses with neural responses to each pitch and formant perturbations, we computed Pearson's correlation coefficients for activations in all voxels against the compensation responses obtained in speech data analysis. The resulting images were cluster-corrected to 40 voxels, $p<0.05$.

3.4 Results

3.4.1 Feature-specific behavioral compensation responses

Participants shows positive feature specific but negative non-feature specific compensation responses to both pitch and formant perturbations (Figure 3.1). In other words, the average F0

responses to F0 perturbations were positively larger than to F1 perturbations, and the average F1 responses to F1 perturbations were also positively larger than to F0 perturbations. The average feature-specific F0 responses peaked higher and slightly faster (Mean±SEM = 11.93±2.69 cents at 500ms) than the average feature-specific F1 responses (Mean±SEM = 6.36±2.36 cents at 525ms) (Figure 3.1a-b). On the other hand, the non-feature specific responses are in the opposite direction of the applied perturbations: peak non-feature specific F1 responses are -4.18±1.92 cents at 575ms and peak non-feature specific F0 responses are -2.42±1.55 cents at 650ms. The feature specific individual peak compensation responses are significant (Mean±SEM = 9.59±2.13 cents, $t(14)=4.51$, $p=4.898e-4$ and Mean±SEM = 4.98±2.07 cents , $t(14)=2.41$, $p=0.03$ for F0 and F1 feature-specific responses, respectively), but the non-feature specific individual peak compensation responses are not (Mean±SEM = -1.92±1.39 cents , $t(14)=-1.38$, $p=0.19$ and Mean±SEM = -3.07±1.57 cents, $t(14)=-1.95$, $p=0.07$ for F0 and F1 non-feature specific responses, respectively) (Figure 3.1c-d).

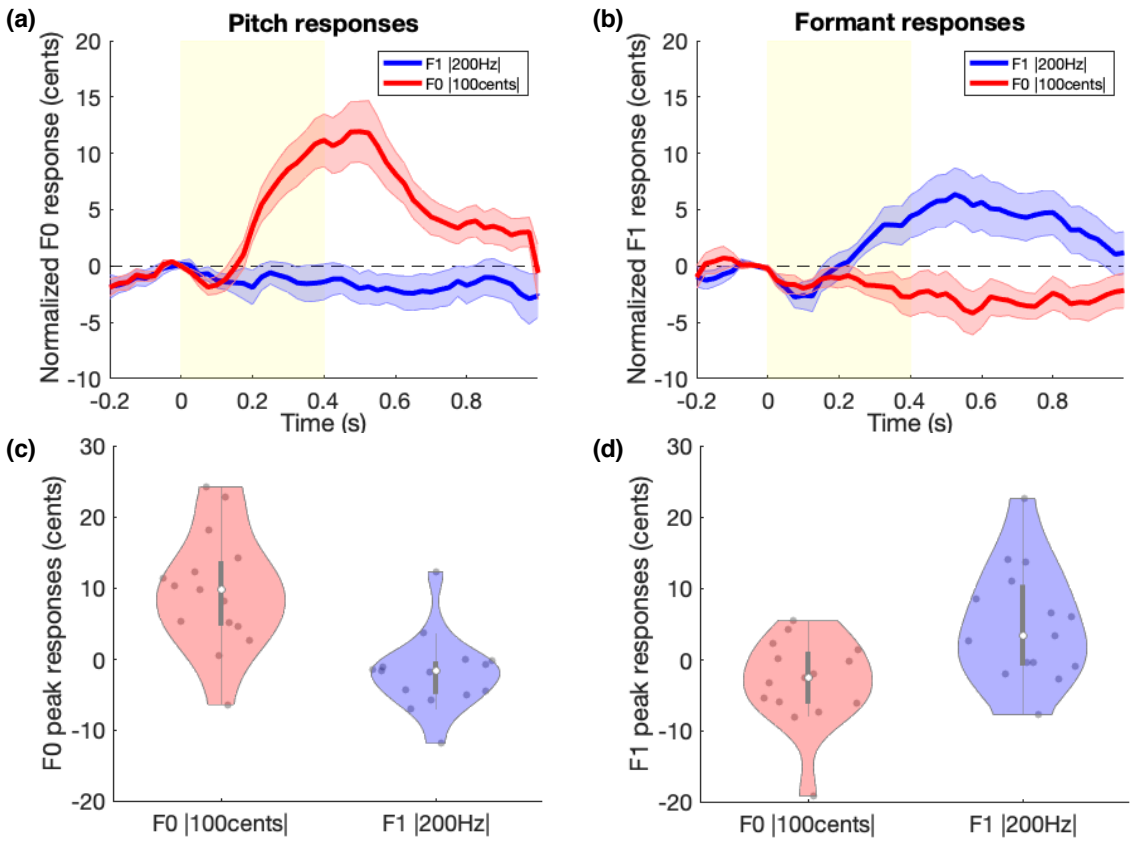

Figure 3.1: Online formant and pitch responses to unpredictable mid-utterance pitch and formant perturbations. (a) F0 and (b) F1 compensation responses in cents averaged across participants to each unpredictable mid-utterance 100 cents F0 (red) and 200Hz F1 (blue) perturbations. Compensation responses to the opposite directions were combined to parallel the neural response analysis results. Time-range of perturbation is shown as shaded yellow area (perturbation onset at t=0). Mean responses (lines) and SEM (shaded colored region) are shown. The distributions of peak (c) F0 and (d) F1 compensation responses to unpredictable mid-utterance (c) pitch and (d) formant perturbations are shown as violin plots. The grey bar indicates the range from 1st to 3rd quartile and the white dot indicates the median. The shape of the violin plot reflects the kernel density estimate of the data, and the colored dots are actual individual response data points.

3.4.2 Brain regions involved in auditory feedback processing across frequency bands

A wide network of significant activation is observed in response to pitch (Figure 3.2), formant (Figure 3.3) compared to steady-state vocalization (pre-perturbation onset) across different frequency bands. Neural regions involved in response to both pitch and formant perturbations

(including both overlapped and feature-specific regions) are shown in Figure 3.4 (with peak of activations shown in Table 3.1-3.5), and regions that directionally overlapped (either both enhanced or decreased activation) in response to both pitch and formant perturbations are shown in Figure 3.5. Overall, most regions showing enhanced or decreased activation compared to baseline in Figure 3.4 are also observed in Figure 3.5, indicating that these regions are involved in the processing of both pitch and formant perturbations, though the few regions that are observed only in Figure 4 signify their higher involvement in processing for only either pitch or formants.

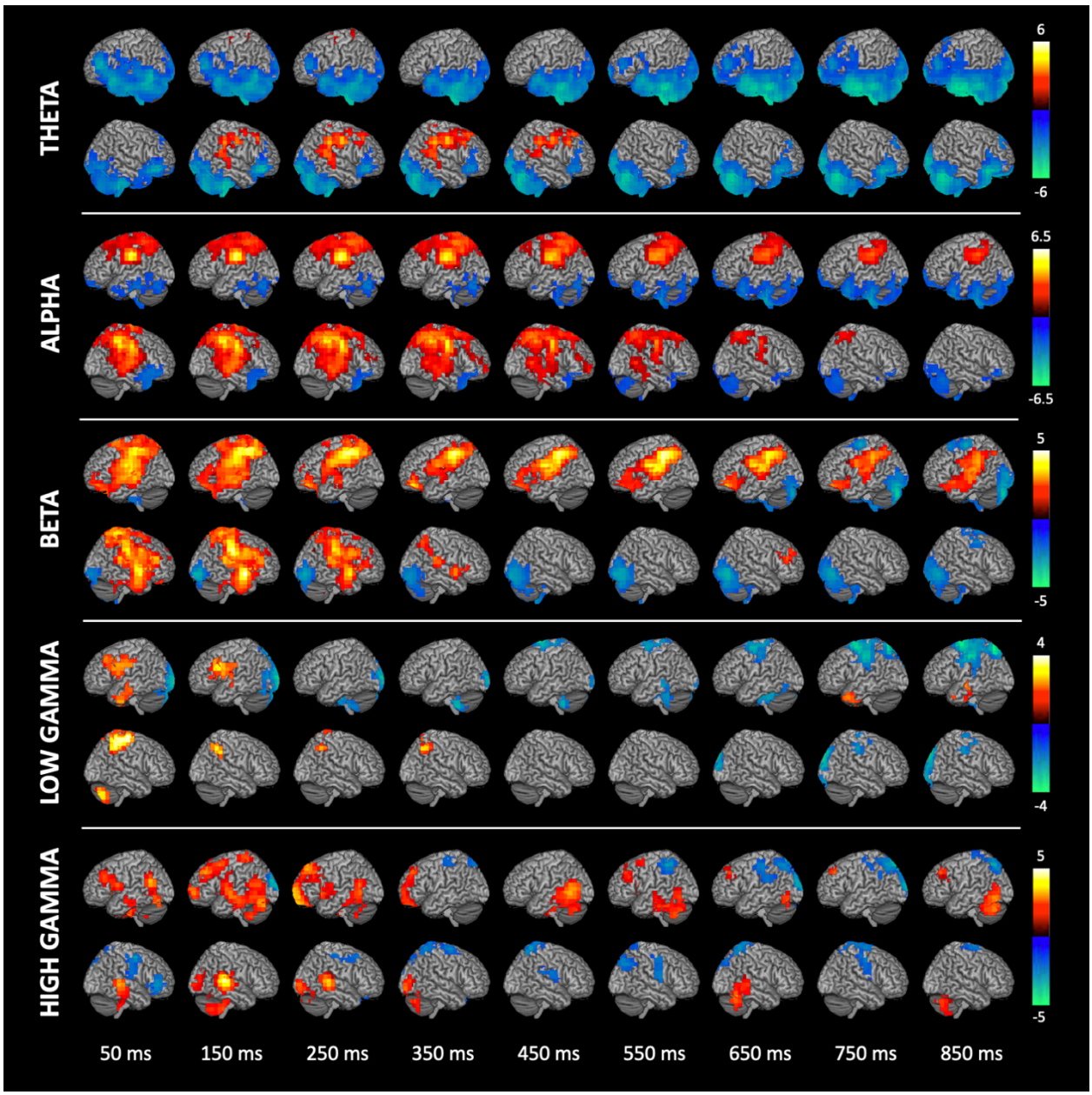

Figure 3.2: Time-frequency dynamics of neural regions involved in auditory feedback processing of pitch. Neural responses aligned to perturbation onset that are significant in response to pitch perturbations compared to steady-state vocalizations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images are cluster corrected, 30 voxels, $p<0.05$. Color scale represents t-value.

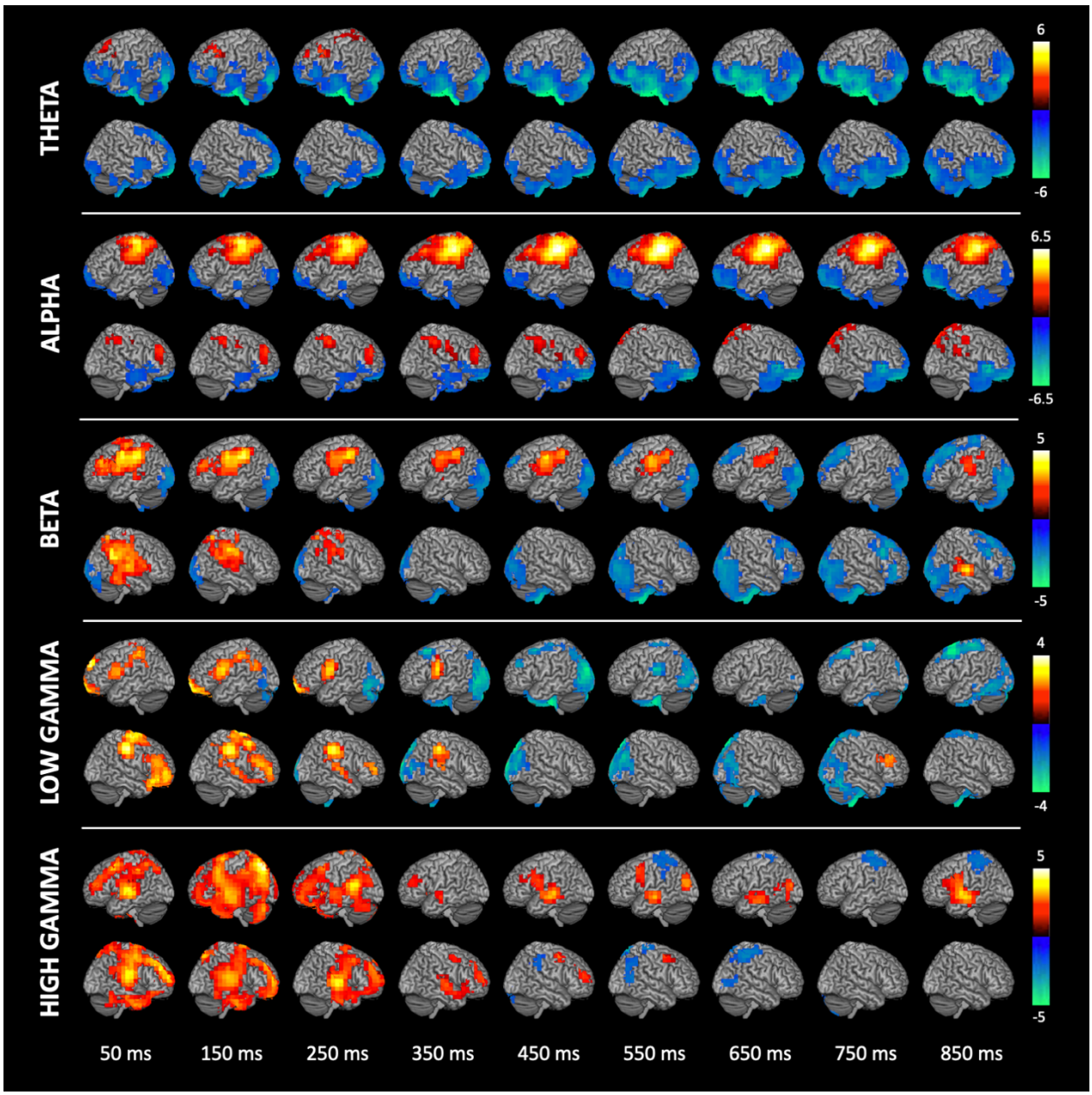

Figure 3.3: Time-frequency dynamics of neural regions involved in auditory feedback processing of formants. Neural responses aligned to perturbation onset that are significant in response to formant perturbations compared to steady-state vocalizations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images are cluster corrected, 30 voxels, $p<0.05$. Color scale represents t-value.

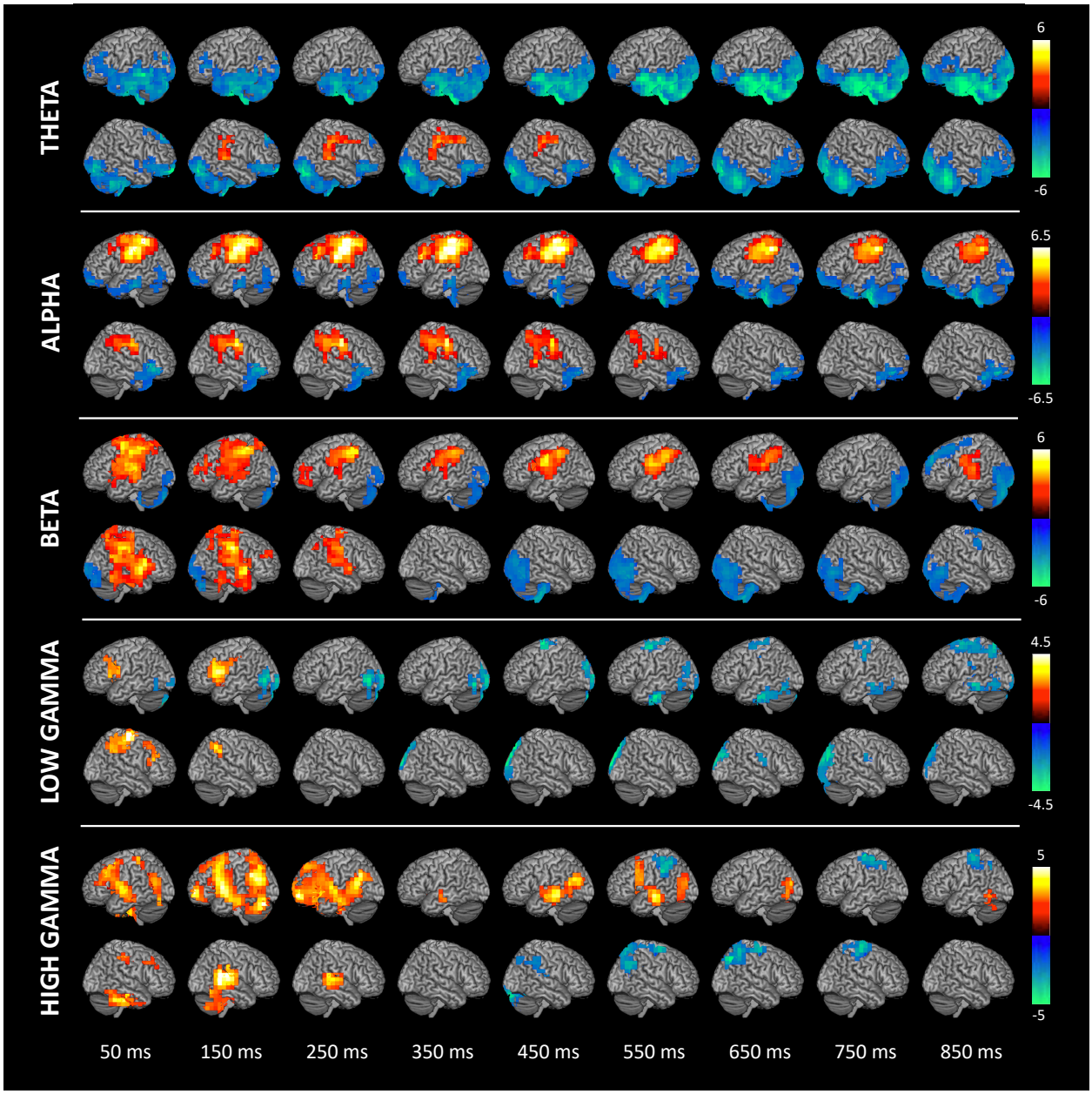

Figure 3.4: Time-frequency dynamics of neural regions involved in auditory feedback processing of either pitch or formants. Neural responses aligned to perturbation onset that are significant in response to pitch or formant perturbations compared to steady-state vocalizations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images are cluster corrected, 30 voxels, $p<0.05$. Color scale represents t-value.

Table 3.1: Regions of significant enhancement in response to both pitch and formant feedback perturbations compared to unperturbed response in the theta band (4-7Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Middle Temporal Gyrus	-50.0 4.0 -18.0	550-650 ms	50-950 ms	-6.410	1.63E-05
Left Cerebellum, Posterior Lobe, Tuber	-55.0 -50.0 -35.0	650-750 ms	50-950 ms	-6.344	1.82E-05
Left Fusiform Gyrus	-53.0 -12.0 -32.0	850-950 ms	50-950 ms	-6.746	9.38E-06
Left Inferior Occipital Gyrus	-14.0 -91.0 -16.0	850-950 ms	50-950 ms	-5.509	7.70E-05
Left Cerebellum, Posterior Lobe, Uvula	-17.0 -75.0 -34.0	850-950 ms	50-950 ms	-5.366	9.95E-05
Left Cuneus	-17.0 -93.0 5.0	850-950 ms	50-950 ms	-5.152	1.47E-04
Left Inferior Temporal Gyrus	-48.0 -5.0 -35.0	850-950 ms	50-950 ms	-6.871	9.95E-04
Left Superior Frontal Gyrus	-17.0 66.0 -8.0	850-950 ms	550-950 ms	-5.146	1.48E-04
Right Medial Frontal Gyrus	10.0 66.0 -15.0	50-150 ms	50-250 ms	-6.816	8.38E-06
Right Superior Frontal Gyrus	16.0 42.0 52.0	50-150 ms	50-350 ms	-4.557	4.48E-04
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	56.0 -53.0 -43.0	50-150 ms	50-950 ms	-4.501	4.99E-04
Right Lingual Gyrus	25.0 -82.0 -15.0	250-350 ms	50-950 ms	-7.541	2.71E-06
Right Inferior Frontal Gyrus	54.0 26.0 -9.0	350-450 ms	50-950 ms	-3.631	2.70E-03
Right Cerebellum, Posterior Lobe, Declive	16.0 -76.0 -18.0	650-750 ms	50-950 ms	-6.537	1.32E-05
Right Cerebellum, Posterior Lobe, Tuber	40.0 -60.0 -34.0	850-950 ms	50-950 ms	-6.322	1.89E-05
Right Middle Temporal Gyrus	56.0 -35.0 5.0	150-250 ms	150-450 ms	3.620	2.80E-03
Right Postcentral Gyrus	48.0 -27.0 38.0	350-450 ms	150-550 ms	3.940	0.0029
Right Precentral Gyrus	62.0 -3.0 36.0	450-550 ms	250-550 ms	3.673	2.50E-03

Table 3.2: Regions of significant enhancement in response to both pitch and formant feedback perturbations compared to unperturbed response in the alpha band (8-12Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Inferior Temporal Gyrus	-62.0 -25.0 -22.0	50-150 ms	50-850 ms	-4.211	8.71E-04
Left Postcentral Gyrus	-38.0 -27.0 45.0	150-250 ms	50-850 ms	8.254	9.88E-06
Left Superior Frontal Gyrus	-24.0 55.0 -17.0	50-150 ms	50-950 ms	-3.760	0.0021
Left Precentral Gyrus	-62.0 -12.0 28.0	250-350 ms	50-950 ms	6.539	1.31E-05
Left Middle Temporal Gyrus	-62.0 -27.0 -17.0	450-550 ms	50-950 ms	-3.987	0.0014
Left Inferior Parietal Lobule	-54.0 -44.0 47.0	550-650 ms	50-950 ms	7.504	2.86E-06
Left Cerebellum, Anterior Lobe	-16.0 -44.0 -34.0	550-650 ms	50-950 ms	-5.398	9.39E-05
Left Superior Frontal Gyrus	-17.0 66.0 -3.0	850-950 ms	50-950 ms	-3.720	0.0021
Left Middle Temporal Gyrus	-54.0 2.0 -32.0	750-850 ms	550-950 ms	-3.201	0.0064
Right Inferior Frontal Gyrus	54.0 19.0 -4.0	50-150 ms	50-450 ms	-4.758	3.06E-04
Right Precentral Gyrus	54.0 -12.0 38.0	250-350 ms	50-650 ms	6.001	3.25E-05
Right Postcentral Gyrus	54.0 -11.0 46.0	250-350 ms	50-650 ms	5.776	4.80E-05
Right Inferior Parietal Lobule	48.0 -42.0 45.0	250-350 ms	50-650 ms	3.969	1.40E-03
Right Superior Temporal Gyrus	53.0 10.0 -25.0	850-950 ms	50-950 ms	-4.222	8.53E-04
Right Middle Frontal Gyrus	53.0 42.0 -9.0	850-950 ms	50-950 ms	-4.010	0.0013
Right Middle Temporal Gyrus	48.0 -44.0 7.0	250-350 ms	150-650 ms	3.938	0.0015
Right Superior Parietal Lobule	32.0 -60.0 54.0	450-550 ms	150-650 ms	3.243	0.0059

Table 3.3: Regions of significant enhancement in response to both pitch and formant feedback perturbations compared to unperturbed response in the beta band (13-30Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Middle Frontal Gyrus	-40.0 -2.0 54.0	50-150 ms	50-250 ms	4.673	3.59E-04
Left Insula	-39.0 -20.0 14.0	50-150 ms	50-250 ms	4.486	5.13E-04
Left Superior Temporal Gyrus	-41.0 -19.0 8.0	50-150 ms	50-250 ms	4.301	7.31E-04
Left Cerebellum, Posterior Lobe, Pyramis	-32.0 -83.0 -43.0	50-150 ms	50-450 ms	-3.650	0.0026
Left Postcentral Gyrus	-55.0 -17.0 47.0	50-150 ms	50-550 ms	4.488	5.10E-04
Left Inferior Parietal Lobule	-55.0 -36.0 45.0	150-250 ms	50-750 ms	4.594	4.17E-04
Left Precentral Gyrus	-62.0 -12.0 29.0	450-550 ms	50-750 ms	4.677	3.57E-04
Left Cerebellum, Posterior Lobe, Tuber	-39.0 -85.0 -34.0	850-950 ms	650-950 ms	-3.893	0.0016
Left Middle Occipital Gyrus	-47.0 -75.0 -10.0	850-950 ms	650-950 ms	-3.393	0.0044
Left Superior Frontal Gyrus	-24.0 50.0 29.0	850-950 ms	850-950 ms	-3.841	0.0018
Left Superior Temporal Gyrus	-62.0 -26.0 13.0	850-950 ms	850-950 ms	3.365	4.60E-03
Left Precentral Gyrus	-62.0 -3.0 31.0	850-950 ms	850-950 ms	3.092	0.008
Right Cuneus	14.0 -93.0 -1.0	50-150 ms	50-250 ms	-3.730	0.0022
Right Precentral Gyrus	48.0 -21.0 38.0	150-250 ms	50-350 ms	5.511	7.67E-05
Right Superior Temporal Gyrus	64.0 3.0 -2.0	50-150 ms	50-350 ms	5.013	1.90E-04
Right Middle Temporal Gyrus	64.0 3.0 -8.0	150-250 ms	50-350 ms	4.748	3.11E-04
Right Superior Frontal Gyrus	30.0 -11.0 68.0	150-250 ms	50-350 ms	3.828	0.0018
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	7.0 -53.0 -65.0	550-650 ms	450-950 ms	-4.213	8.69E-04
Right Fusiform Gyrus	54.0 -66.0 -18.0	550-650 ms	450-950 ms	-3.733	0.0022
Right Middle Temporal Gyrus	47.0 -60.0 6.0	550-650 ms	450-850 ms	-3.562	0.0031
Right Middle Frontal Gyrus	46.0 3.0 54.0	850-950 ms	850-950 ms	-3.060	0.0085

Table 3.4: Regions of significant enhancement in response to both pitch and formant feedback perturbations compared to unperturbed response in the low gamma band (30-55Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	-14.0 -82.0 -50.0	50-150 ms	50-250 ms	-4.711	3.34E-04
Left Precentral Gyrus	-47.0 11.0 12.0	150-250 ms	50-250 ms	4.094	1.10E-03
Left Middle Occipital Gyrus	-48.0 -75.0 -2.0	250-350 ms	150-450 ms	-4.193	9.02E-04
Left Precentral Gyrus	-39.0 -13.0 60.0	550-650 ms	450-950 ms	-3.862	0.0017
Left Inferior Temporal Gyrus	-39.0 -18.0 -33.0	550-650 ms	550-750 ms	-4.143	9.95E-04
Left Superior Parietal Lobule	-25.0 -68.0 53.0	850-950 ms	850-950 ms	-2.567	0.0224
Right Superior Frontal Gyrus	31.0 -11.0 69.0	50-150 ms	50-150 ms	5.097	1.63E-04
Right Inferior Frontal Gyrus	55.0 26.0 22.0	50-150 ms	50-150 ms	3.090	8.00E-03
Right Inferior Parietal Lobule	54.0 -53.0 45.0	150-250 ms	50-250 ms	3.500	3.50E-03
Right Cuneus	16.0 -91.0 29.0	450-550 ms	350-950 ms	-4.151	9.80E-04
Right Precuneus	8.0 -82.0 46.0	450-550 ms	450-950 ms	-4.388	6.20E-04

Table 3.5: Regions of significant enhancement in response to both pitch and formant feedback perturbations compared to unperturbed response in the high gamma band (65-150Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Cerebellum, Posterior Lobe, Declive	-46.0 -76.0 -25.0	162.5-187.5 ms	12.5-237.5 ms	6.425	1.59E-05
Left Superior Temporal Gyrus	-56.0 3.0 5.0	87.5-112.5 ms	12.5-287.5 ms	5.687	5.61E-05
Left Precentral Gyrus	-61.0 3.0 13.0	162.5-187.5 ms	12.5-287.5 ms	4.870	2.48E-04
Left Inferior Frontal Gyrus	-54.0 11.0 23.0	112.5-137.5 ms	12.5-287.5 ms	4.374	6.36E-04
Left Supramarginal Gyrus	-39.0 -49.0 30.0	112.5-137.5 ms	37.5-237.5 ms	4.860	2.53E-04
Left Middle Temporal Gyrus	-49.0 -18.0 -16.0	262.5-287.5 ms	62.5-612.5 ms	4.462	5.36E-04
Left Middle Frontal Gyrus	-40.0 58.0 -9.0	187.5-212.5 ms	87.5-287.5 ms	6.008	3.21E-05
Left Angular Gyrus	-54.0 -67.0 30.0	187.5-212.5 ms	112.5-262.5 ms	4.040	1.20E-03
Left Middle Temporal Gyrus	-52.0 -20.0 -10.0	562.5-587.5 ms	412.5-612.5 ms	4.621	3.96E-04
Left Inferior Frontal Gyrus	-56.0 11.0 36.0	512.5-537.5 ms	487.5-587.5 ms	3.558	3.20E-03
Left Middle Occipital Gyrus	-45.0 -76.0 -1.0	612.5-637.5 ms	512.5-662.5 ms	3.511	3.50E-03
Left Inferior Parietal Lobule	-47.0 -35.0 36.0	537.5-562.5 ms	537.5-637.5 ms	-4.283	7.58E-04
Left Precentral Gyrus	-40.0 -17.0 62.0	712.5-737.5 ms	662.5-937.5 ms	-3.650	2.60E-03
Right Cerebellum, Anterior Lobe, Culmen	40.0 -44.0 -34.0	12.5-37.5 ms	12.5-112.5 ms	5.265	1.19E-04
Right Fusiform Gyrus	32.0 -34.0 -24.0	37.5-62.5 ms	12.5-237.5 ms	3.260	5.70E-03
Right Precentral Gyrus	39.0 -22.0 37.0	62.5-87.5 ms	37.5-112.5 ms	4.788	2.89E-04
Right Superior Temporal Gyrus	69.0 -45.0 5.0	87.5-112.5 ms	62.5-262.5 ms	6.434	1.56E-05
Right Middle Temporal Gyrus	69.0 -43.0 3.0	87.5-112.5 ms	62.5-262.5 ms	5.835	4.33E-05
Right Cerebellum, Posterior Lobe, Declive	8.0 -86.0 -25.0	412.5-437.5 ms	387.5-487.5 ms	-5.748	5.04E-05
Right Lingual Gyrus	8.0 -92.0 -9.0	412.5-437.5 ms	387.5-487.5 ms	-3.015	0.0093
Right Superior Parietal Lobule	30.0 -67.0 45.0	637.5-662.5 ms	437.5-712.5 ms	-5.353	1.02E-04
Right Precuneus	14.0 -67.0 52.0	562.5-587.5 ms	487.5-687.5 ms	-4.110	0.0011
Right Precentral Gyrus	30.0 -12.0 61.0	587.5-612.5 ms	537.5-662.5 ms	-5.315	1.09E-04

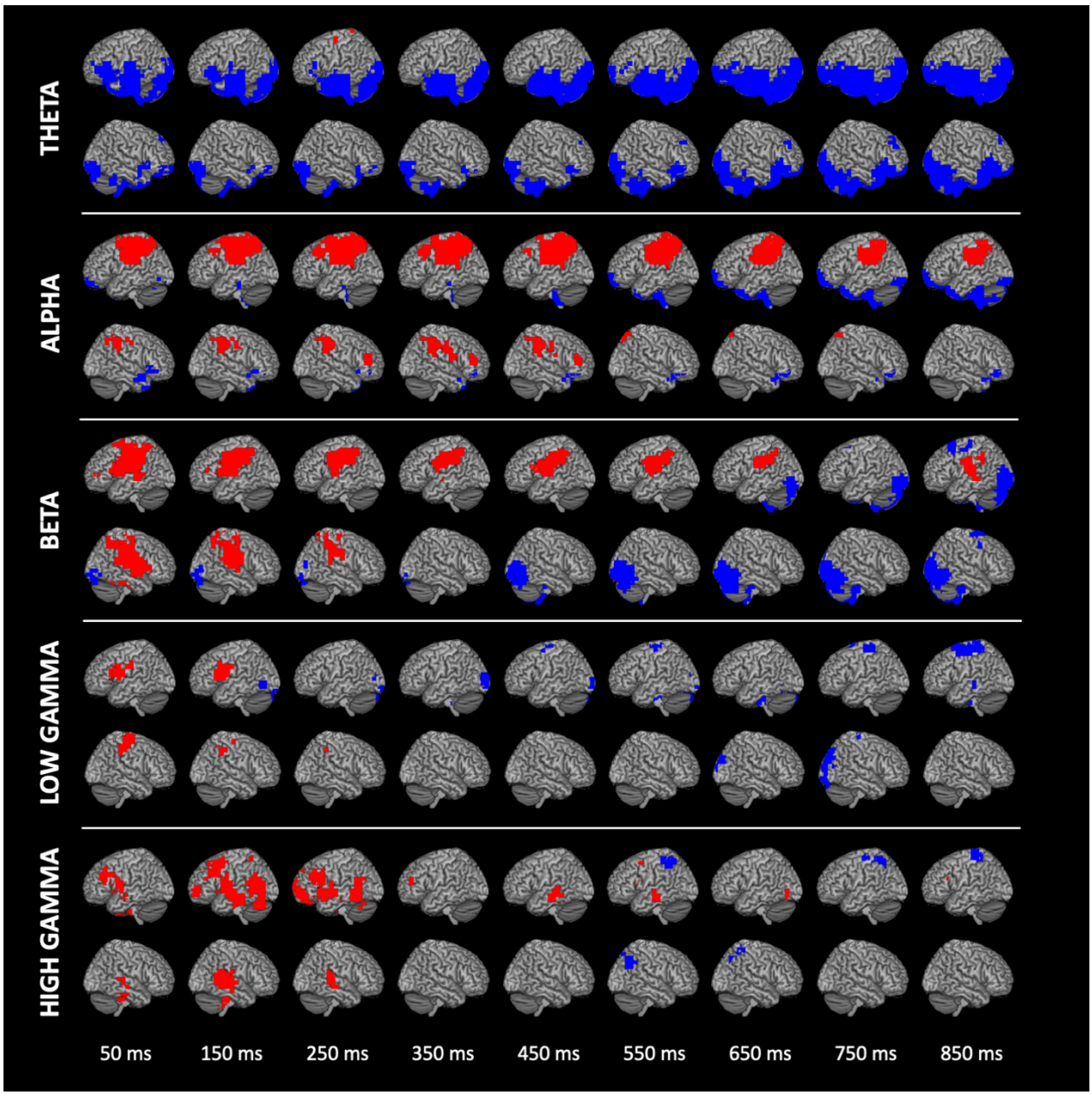

Figure 3.5: Time-frequency dynamics of neural responses shared between pitch and formant feedback perturbations. Neural responses aligned to perturbation onset that are significant in response to both pitch and formant perturbations compared to steady-state vocalizations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images represent regions where voxels survived cluster correction (30 voxels, $p<0.05$) and are significantly enhanced (red) or decreased (blue) in responses to both pitch and formant perturbations.

In the theta frequency band (4-7Hz), sustained lower activity compared to baseline are observed in response to both pitch and formant perturbations bilaterally throughout frontal, temporal,

occipital lobes and cerebellum throughout all time windows, from perturbation onset all the way up to few hundred milliseconds after perturbation offset (at 400ms). In the left hemisphere, most of the sustained lower activities began at perturbation onset peaks after perturbation offset, first in middle temporal gyrus, cerebellum, and finally at the cuneus, fusiform, inferior occipital, and inferior temporal gyrus. A decreased activity in the frontal lobe was observed post perturbation offset and peaks later in the superior frontal gyrus. In the right hemisphere, early peaks of decreased activity were observed in the medial and superior frontal gyrus, which showed sustained lower activity only in the perturbation window. The rest of regions showing decreased activity have sustained activity throughout all time windows, with peaks first in cerebellum, lingual gyrus, inferior frontal gyrus (IFG), and finally in cerebellar regions. Several enhanced activations are also observed only in the overall but not directionally overlapped responses to both pitch and formant perturbations within the perturbation window, with peaks in the middle temporal, postcentral and precentral gyrus, which are not observed in the directionally overlapped regions.

In the alpha frequency band (8-12Hz), sustained enhanced activity compared to baseline is observed in response to both pitch and formant perturbations bilaterally in several motor areas, though more pronounced and sustained in the left hemisphere. The enhanced activity in the left hemisphere peaks in the postcentral gyrus and precentral gyrus within the perturbation window and in the inferior parietal lobule after perturbation offset. In the right hemisphere, the enhanced activity all peaks within the perturbation window, first in the precentral gyrus, postcentral gyrus, and inferior parietal lobule (IPL), followed soon with peaks in middle temporal gyrus (MTG) and superior parietal lobule (SPL). Sustained decreased activity compared to baseline is also

observed bilaterally. This decreased activity peaks early post perturbation onset in the left inferior temporal, left superior frontal gyrus, and right inferior frontal gyrus. It is then followed by a peak in the left cerebellum soon after perturbation offset, and then peaks much later post perturbation offset in the left superior frontal gyrus (SFG), left MTG, right superior temporal gyrus (STG) and right middle frontal gyrus (MFG).

In the beta band (13-30Hz), enhanced activity is observed in response to both pitch and formant perturbations early bilaterally in motor and temporal regions with more sustained activity in the left hemisphere, while decreased activity is observed at different time windows in the left and right hemisphere. The enhanced activation peaks early within the perturbation time window in bilateral STG, left MFG, insula, postcentral gyrus and IPL, and in the right precentral, MTG and STG. This is soon followed by peaks in left precentral gyrus right after perturbation offset and much later in the left STG and precentral gyrus. Early decreased activity was observed transiently in the perturbation window, though only in the right hemisphere in the directionally overlapped responses to both pitch and formant perturbations, with peaks early post perturbation onset in the left cerebellum and right cuneus. This decreased activity was observed again post perturbation offset bilaterally with peaks soon after post perturbation offset in right cerebellum, fusiform and MTG and peaks much later in the left cerebellum, left middle occipital gyrus (MOG) and left SFG.

In the low gamma frequency band (30-55Hz), transient enhanced activation is observed in response to both pitch and formant perturbations bilaterally within the perturbation window and decreased activation is more sparsely observed throughout different regions and time windows.

Enhanced activation in the perturbation window peaks first in the right SFG and IFG then in the right IPL and left precentral gyrus. Peaks of decreased activation are first observed in the left cerebellum and left MOG early post perturbation onset, in the left precentral and inferior temporal gyrus (ITG) and right cuneus and precuneus post perturbation offset and finally much later in the left SPL.

In the high gamma frequency band (65-150Hz), enhanced activation is observed in response to both pitch and formant perturbations bilaterally within the perturbation window, followed by a mix of enhanced and decreased activation the left hemisphere and dominantly decrease activation in the right hemisphere post perturbation offset. Enhanced activity in the left hemisphere encompassed a large network during the perturbation window, with peaks in STG, MTG, SMG, IFG, MFG, precentral gyrus, angular gyrus and cerebellum. Enhanced activity in the right hemisphere is observed early within the perturbation window, with peaks in STG, MTG, precentral gyrus, fusiform gyrus and cerebellum. Some enhanced activity was also observed post perturbation offset in the left hemisphere with peaks in IFG, MTG and MOG. Decreased activity was observed only post perturbation offset in both hemispheres. In the left hemisphere, peaks of decreased activity are observed in the inferior parietal lobule and precentral gyrus. In the right hemisphere, peaks of decreased activity are observed in precentral gyrus, SPL, precuneus, lingual gyrus and cerebellum.

3.4.3 Lateralization of neural responses to formant vs pitch feedback perturbations

Figure 3.6 shows the contrast in neural responses to formant vs pitch feedback perturbations (with peak of activations shown in Table 3.6-3.10). Generally, neural responses to formant

feedback perturbations are more left lateralized in the lower frequency bands but more right lateralized in the higher frequency bands. Vice versa, neural responses to pitch feedback perturbations are more right lateralized in the lower frequency bands though it became sparser bilaterally in the higher frequency bands.

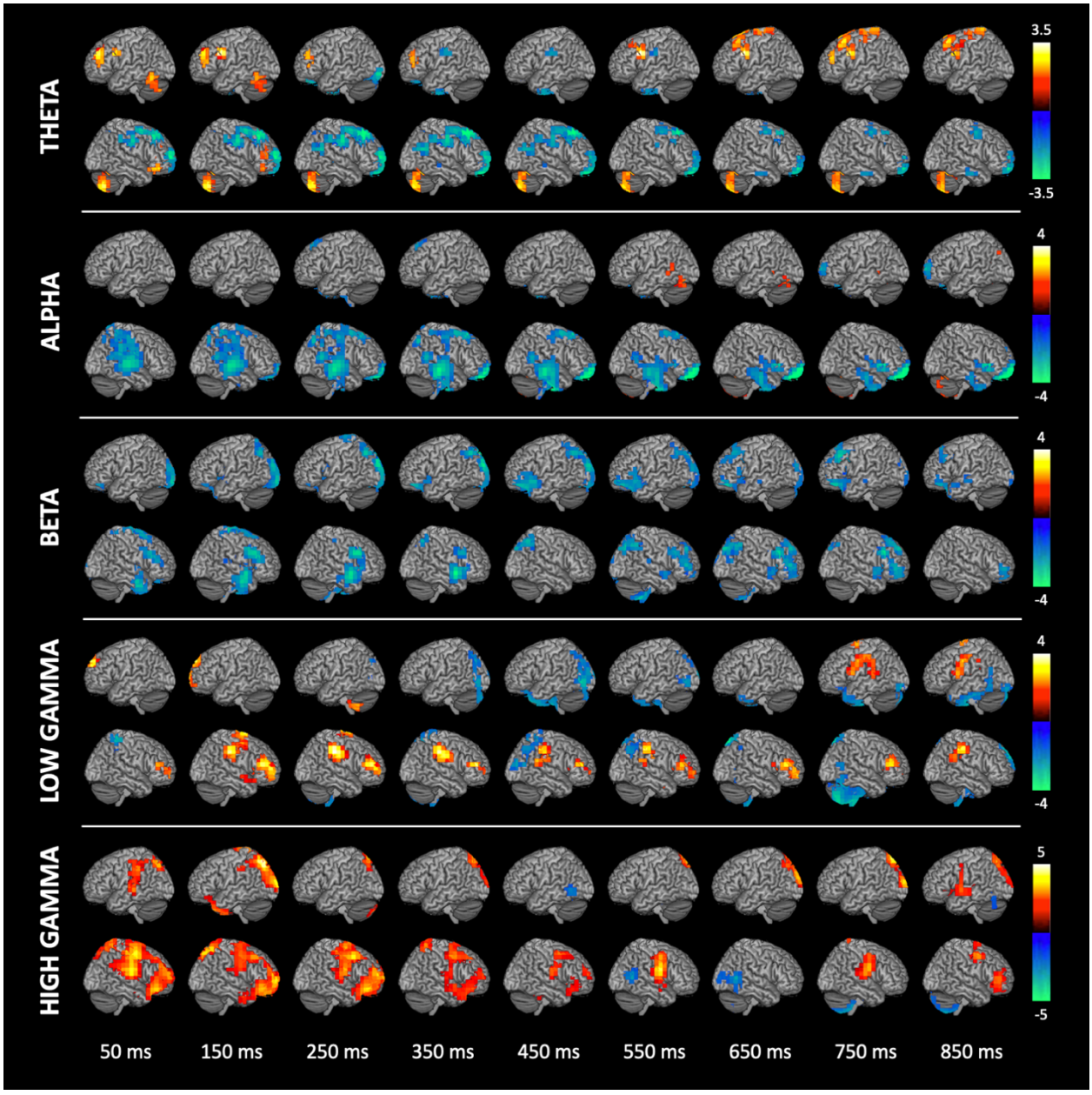

Figure 3.6. Time-frequency dynamics of neural responses specific to formant vs pitch feedback processing. Neural responses aligned to perturbation onset that are significantly

different in response to formant and pitch perturbations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images are cluster corrected, 30 voxels, $p<0.05$. Color scale represents t-value, where positive values indicate regions more enhanced in responses to formant perturbations and negative values indicate regions more enhanced in responses to pitch perturbations.

Table 3.6: Regions of significant contrast in response to formant vs pitch feedback perturbations compared to unperturbed response in the theta band (4-7Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Inferior Temporal Gyrus	-53.0 -58.0 -17.0	50-150 ms	50-250 ms	2.238	1.04E-02
Left Precentral Gyrus	-63.0 2.0 29.0	150-250 ms	50-250 ms	2.896	4.40E-03
Left Middle Frontal Gyrus	-39.0 42.0 30.0	50-150 ms	50-450 ms	2.781	9.10E-03
Left Medial Globus Pallidus	-14.0 -1.0 -2.0	350-450 ms	150-650 ms	-2.883	4.40E-03
Left Inferior Occipital Gyrus	-31.0 -93.0 -9.0	250-350 ms	250-350 ms	-3.003	9.00E-04
Left Inferior Frontal Gyrus	-61.0 12.0 31.0	550-650 ms	550-950 ms	3.776	5.00E-03
Left Superior Frontal Gyrus	-7.0 18.0 56.0	650-750 ms	650-950 ms	2.367	8.20E-03
Left Postcentral Gyrus	-24.0 -41.0 69.0	850-950 ms	650-950 ms	1.971	1.18E-02
Left Middle Frontal Gyrus	-47.0 44.0 20.0	750-850 ms	750-850 ms	2.433	2.51E-02
Right Inferior Frontal Gyrus	32.0 19.0 -18.0	50-150 ms	50-150 ms	2.133	3.35E-02
Right Postcentral Gyrus	48.0 -20.0 30.0	150-250 ms	50-650 ms	-3.305	1.10E-03
Right Superior Frontal Gyrus	24.0 26.0 60.0	350-450 ms	50-750ms	-4.206	1.00E-04
Right Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	33.0 -68.0 -50.0	150-250 ms	50-950 ms	3.434	9.00E-04
Right Superior Temporal Gyrus	58.0 -60.0 23.0	350-450 ms	250-550 ms	-2.754	2.00E-04
Right Inferior Temporal Gyrus	64.0 -12.0 -25.0	850-950 ms	650-850 ms	-2.166	2.28E-02

Table 3.7: Regions of significant contrast in response to formant vs pitch feedback perturbations compared to unperturbed response in the alpha band (8-12Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Superior Frontal Gyrus	-14.0 34.0 55.0	250-350 ms	250-450 ms	-2.434	1.07E-02
Left Medial Frontal Gyrus	-8.0 20.0 -18.0	450-550 ms	250-650 ms	-2.420	1.09E-02
Left Fusiform Gyrus	-49.0 -71.0 -18.0	550-650 ms	550-750 ms	2.144	2.39E-02
Left Middle Temporal Gyrus	-55.0 -45.0 -2.0	550-650 ms	550-850 ms	1.929	4.39E-02
Left Precuneus	-40.0 -75.0 37.0	850-950 ms	550-950 ms	1.830	3.88E-02
Left Superior Frontal Gyrus	-30.0 67.0 6.0	850-950 ms	750-950 ms	-3.372	1.80E-03
Right Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	8.0 -68.0 -56.0	550-650 ms	450-950 ms	2.477	1.05E-02
Right Middle Temporal Gyrus	69.0 -27.0 -3.0	150-250 ms	50-650 ms	-3.475	5.00E-04
Right Superior Temporal Gyrus	62.0 -53.0 13.0	250-350 ms	50-650 ms	-2.925	2.80E-03
Right Inferior Temporal Gyrus	40.0 -19.0 -34.0	450-550 ms	250-750 ms	-3.346	2.00E-04
Right Superior Frontal Gyrus	31.0 59.0 -15.0	550-650 ms	250-950 ms	-4.428	1.30E-03
Right Inferior Frontal Gyrus	55.0 17.0 -1.0	850-950 ms	550-850 ms	-3.032	5.00E-03

Table 3.8: Regions of significant contrast in response to formant vs pitch feedback perturbations compared to unperturbed response in the beta band (13-30Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Lingual Gyrus	-25.0 -97.0 -9.0	150-250 ms	50-950 ms	-3.154	3.50E-03
Left Inferior Parietal Lobule	40.0 -67.0 48.0	250-350 ms	150-650 ms	-3.088	1.09E-02
Left Cuneus	-16.0 -78.0 21.0	250-350 ms	150-950 ms	-4.839	1.00E-04
Left Postcentral Gyrus	-10.0 -35.0 70.0	250-350 ms	250-350 ms	-2.306	2.47E-02
Left Superior Temporal Gyrus	-50.0 10.0 -10.0	450-550 ms	350-850 ms	-2.820	2.60E-03
Left Middle Frontal Gyrus	-32.0 28.0 30.0	450-550 ms	650-950 ms	-3.446	2.20E-03
Right Postcentral Gyrus	23.0 -43.0 69.0	50-150 ms	50-250 ms	-2.166	1.88E-02
Right Inferior Frontal Gyrus	54.0 14.0 29.0	250-350 ms	50-450 ms	-3.494	3.00E-04
Right Inferior Parietal Lobule	49.0 -28.0 23.0	150-250 ms	150-350 ms	-2.214	4.10E-03
Right Superior Temporal Gyrus	56.0 3.0 -10.0	350-450 ms	150-450 ms	-4.056	1.00E-04
Right Precuneus	30.0 -83.0 38.0	650-750 ms	450-850 ms	-3.131	7.40E-03
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	18.0 -43.0 -56.0	550-650 ms	550-750 ms	-3.075	1.10E-03
Right Superior Frontal Gyrus	30.0 44.0 37.0	650-750 ms	550-850 ms	-2.904	1.13E-02
Right Middle Frontal Gyrus	49.0 20.0 29.0	750-850 ms	550-850 ms	-3.618	3.40E-03
Right Superior Temporal Gyrus	55.0 5.0 -10.0	750-850 ms	550-850 ms	-2.335	1.53E-02
Right Inferior Frontal Gyrus	49.0 42.0 7.0	650-750 ms	550-950 ms	-2.913	4.70E-03

Table 3.9: Regions of significant contrast in response to formant vs pitch feedback perturbations compared to unperturbed response in the low gamma band (30-55Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Superior Frontal Gyrus	-21.0 53.0 38.0	50-150 ms	50-250 ms	3.213	1.70E-03
Left Medial Frontal Gyrus	-5.0 67.0 5.0	150-250 ms	150-250 ms	2.785	1.50E-03
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-46.0 -52.0 -54.0	250-350 ms	250-350 ms	2.494	1.74E-02
Left Precuneus	-30.0 -75.0 38.0	350-450 ms	250-550 ms	-2.439	1.60E-03
Left Middle Occipital Gyrus	-35.0 -76.0 0.0	450-550 ms	350-650 ms	-3.478	8.00E-04
Left Superior Frontal Gyrus	-25.0 -2.0 70.0	750-850 ms	750-950 ms	2.579	1.83E-02
Left Inferior Occipital Gyrus	-37.0 -90.0 -15.0	750-850 ms	750-950 ms	-3.228	6.10E-03
Left Precentral Gyrus	-62.0 6.0 14.0	850-950 ms	750-950 ms	2.513	1.62E-02
Left Inferior Temporal Gyrus	-41.0 -19.0 -35.0	850-950 ms	850-950 ms	-2.373	1.41E-02
Right Postcentral Gyrus	38.0 -44.0 64.0	50-150 ms	50-250 ms	-2.871	1.88E-02
Right Middle Frontal Gyrus	46.0 54.0 7.0	150-250 ms	50-750 ms	3.222	6.70E-03
Right Inferior Frontal Gyrus	48.0 35.0 15.0	150-250 ms	50-850 ms	3.439	3.40E-03
Right Superior Temporal Gyrus	54.0 -4.0 -10.0	150-250 ms	150-250 ms	1.724	1.27E-02
Right Superior Frontal Gyrus	17.0 -2.0 70.0	150-250 ms	150-350 ms	2.896	6.20E-03
Right Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	15.0 -77.0 -65.0	250-350 ms	250-450 ms	-3.144	3.00E-03
Right Precuneus	7.0 -54.0 30.0	450-550 ms	350-950 ms	-2.997	6.40E-03
Right Middle Temporal Gyrus	47.0 -82.0 22.0	450-550 ms	450-550 ms	-2.564	8.40E-03
Right Superior Parietal Lobule	18.0 -74.0 55.0	650-750 ms	550-850 ms	-3.667	2.60E-03
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	25.0 -36.0 -67.0	750-850 ms	650-950 ms	-3.603	1.60E-03

Table 3.10: Regions of significant contrast in response to formant vs pitch feedback perturbations compared to unperturbed response in the theta band (65-150Hz)

Region	MNI peak voxel	Peak	Duration	t-value	P-value
Left Superior Temporal Gyrus	-57.0 -28.0 0.0	62.5-87.5 ms	12.5-137.5 ms	3.151	2.10E-03
Left Inferior Parietal Lobule	-64.0 -34.0 38.0	62.5-87.5 ms	12.5-162.5 ms	3.049	3.50E-03
Left Superior Parietal Lobule	-32.0 -75.0 46.0	112.5-137.5 ms	12.5-437.5 ms	4.674	2.00E-04
Left Cuneus	-16.0 -101.0 9.0	137.5-162.5 ms	87.5-237.5 ms	3.822	2.00E-04
Left Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	-14.0 -81.0 -51.0	262.5-287.5 ms	187.5-337.5 ms	3.201	1.40E-03
Left Inferior Temporal Gyrus	-50.0 -60.0 -11.0	412.5-437.5 ms	412.5-512.5 ms	-2.626	5.80E-03
Left Precuneus	-7.0 -77.0 46.0	737.5-762.5 ms	512.5-987.5 ms	3.184	4.00E-03
Left Middle Occipital Gyrus	-24.0 -99.0 14.0	712.5-737.5 ms	562.5-862.5 ms	3.780	1.00E-04
Left Middle Temporal Gyrus	-65.0 -4.0 -3.0	812.5-837.5 ms	787.5-962.5 ms	2.685	7.90E-03
Left Cerebellum, Anterior Lobe, Culmen	-17.0 -43.0 -18.0	812.5-837.5 ms	812.5-862.5 ms	-2.371	2.12E-02
Right Superior Frontal Gyrus	30.0 66.0 3.0	87.5-112.5 ms	12.5-287.5 ms	4.179	1.00E-04
Right Postcentral Gyrus	63.0 -10.0 23.0	37.5-62.5 ms	12.5-312.5 ms	3.803	3.00E-04
Right Inferior Frontal Gyrus	47.0 36.0 -9.0	87.5-112.5 ms	12.5-462.5 ms	3.663	8.00E-04
Right Precentral Gyrus	48.0 -5.0 57.0	37.5-62.5 ms	12.5-562.5 ms	3.886	8.00E-04
Right Precentral Gyrus	62.0 -2.0 22.0	537.5-562.5 ms	12.5-612.5 ms	3.952	8.00E-04
Right Superior Parietal Lobule	32.0 -61.0 56.0	112.5-137.5 ms	37.5-187.5 ms	3.286	9.00E-04
Right Middle Frontal Gyrus	48.0 12.0 46.0	262.5-287.5 ms	37.5-537.5 ms	3.834	1.00E-04
Right Middle Temporal Gyrus	62.0 -60.0 7.0	587.5-612.5 ms	537.5-712.5 ms	-3.684	1.90E-03
Right Precentral Gyrus	62.0 -4.0 23.0	737.5-762.5 ms	687.5-837.5 ms	3.211	1.50E-03
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	34.0 -58.0 -57.0	787.5-812.5 ms	687.5-987.5 ms	-3.114	1.70E-03
Right Inferior Frontal Gyrus	47.0 44.0 8.0	837.5-862.5 ms	762.5-912.5 ms	2.654	8.20E-03
Right Middle Frontal Gyrus	47.0 5.0 48.0	862.5-887.5 ms	737.5-962.5 ms	3.167	3.20E-03

In the theta band, neural responses to formant perturbations have peaks within the perturbation window in left ITG, left precentral gyrus, left MFG, right IFG. Activation in the right cerebellum

is sustained throughout all time windows with its peak early in the perturbation window. Post perturbation offset peaks of formant perturbation responses are observed mostly in the left frontal lobule including IFG, SFG, MFG and postcentral gyrus. Neural responses to pitch perturbations in the right hemisphere are sustained throughout, with peaks early in the perturbation window in the postcentral gyrus, SFG and STG and much later post perturbation offset in ITG. Some transient neural responses to pitch perturbations in the left hemisphere are also observed within the perturbation time window with peaks in the IOG.

In the alpha band, neural responses to pitch perturbations are more dominant, especially in the right hemisphere. Within the perturbation window, pitch perturbation responses showed sustained activation in the right hemisphere with peaks in the MTG and STG, as well as some faint activation in the left hemisphere with peak in the SFG. This is followed by peaks in the left MFG and right ITG and SFG right after perturbation offset, and with a final peak in the right IFG at the latest time window. Only very few regions showed more significant activation for formant perturbations, all after post perturbation offset with peaks in left fusiform gyrus, left MTG, left precuneus, and right cerebellum.

In the beta band, only neural responses to pitch perturbations are showing more significant activation than to formant perturbations, and these activations are sustained throughout all time windows. Within the perturbation window, neural responses to pitch perturbations have peaks in bilateral postcentral gyrus, left lingual gyrus, left IPL, left cuneus, left postcentral gyrus, right IFG, right IPL and right STG. Right after perturbation offset, activation peaks in the left STG

and MFG, followed later by peaks in the right hemisphere including cerebellum, precuneus, SFG, MFG, IFG and STG.

In the low gamma frequency band, neural responses to formant perturbations are more significant than to pitch in the right hemisphere especially in the earlier time window. In the right hemisphere, neural responses to formant perturbations are more significant in the frontal lobe is sustained throughout the time windows, but peaks within the perturbation time window including in the SFG, MFG, IFG, and STG. In the left hemisphere, neural responses to formant perturbations only more significantly peaks early in the SFG, MFG, and cerebellum and later peaks post perturbation offset in SFG and precentral gyrus. Neural responses that are more significant to pitch perturbations are mostly post perturbation with the exception of left precuneus, right postcentral gyrus and right cerebellum. Peaks of higher activation in pitch post perturbation can be found in left MOG, right precuneus and right MTG right after perturbation offset followed much later by peaks in left IOG, left ITG, right SPL and right cerebellum.

In the high gamma frequency band, neural responses to formant perturbations are much more significant overall, involving a wide range of network especially in the right hemisphere. Within the perturbation windows, only neural responses to formant perturbations are most significant than to pitch bilaterally, with peaks in the left hemisphere (STG, IPL, SPL, cuneus and cerebellum) and right hemisphere (SFG, MFG, IFG, postcentral gyrus, precentral gyrus, and SPL). Post perturbation offset, neural responses to formant perturbations are more significant in left precuneus, left MOG, left MTG, right precentral gyrus, right IFG and right MFG. Neural

responses that are more significant to pitch perturbations are after perturbation offset, with peaks in left ITG, left cerebellum, right MTG and right cerebellum.

3.4.4 Neural-behavioral correlations in responses to pitch perturbations

Figure 3.7 shows regions that are correlation to the peak compensation responses to pitch perturbations (with peak of correlations shown in Table 3.11-3.15). In the theta band, neural activation in the left hemisphere mostly have negative correlations to behavioral responses, while in the right hemisphere they are mostly positively correlated. Negative correlations are sustained in the left hemisphere shortly after perturbation onset throughout few hundred milliseconds after perturbation offset in the middle frontal gyrus. In the right hemisphere a transient negative correlation peaks shortly after perturbation onset in SFG and MFG. Peaks of positive correlations early within the perturbation time window are mostly in the right hemisphere (MTG, cerebellum, precentral gyrus and fusiform gyrus) with one peak in the left cerebellum as well. Sustained positive correlation is also observed in the left cuneus which peaks after perturbation offset. In the right hemisphere, significant positive correlations are observed post perturbation offset in the temporal lobe with peaks in MTG and STG.

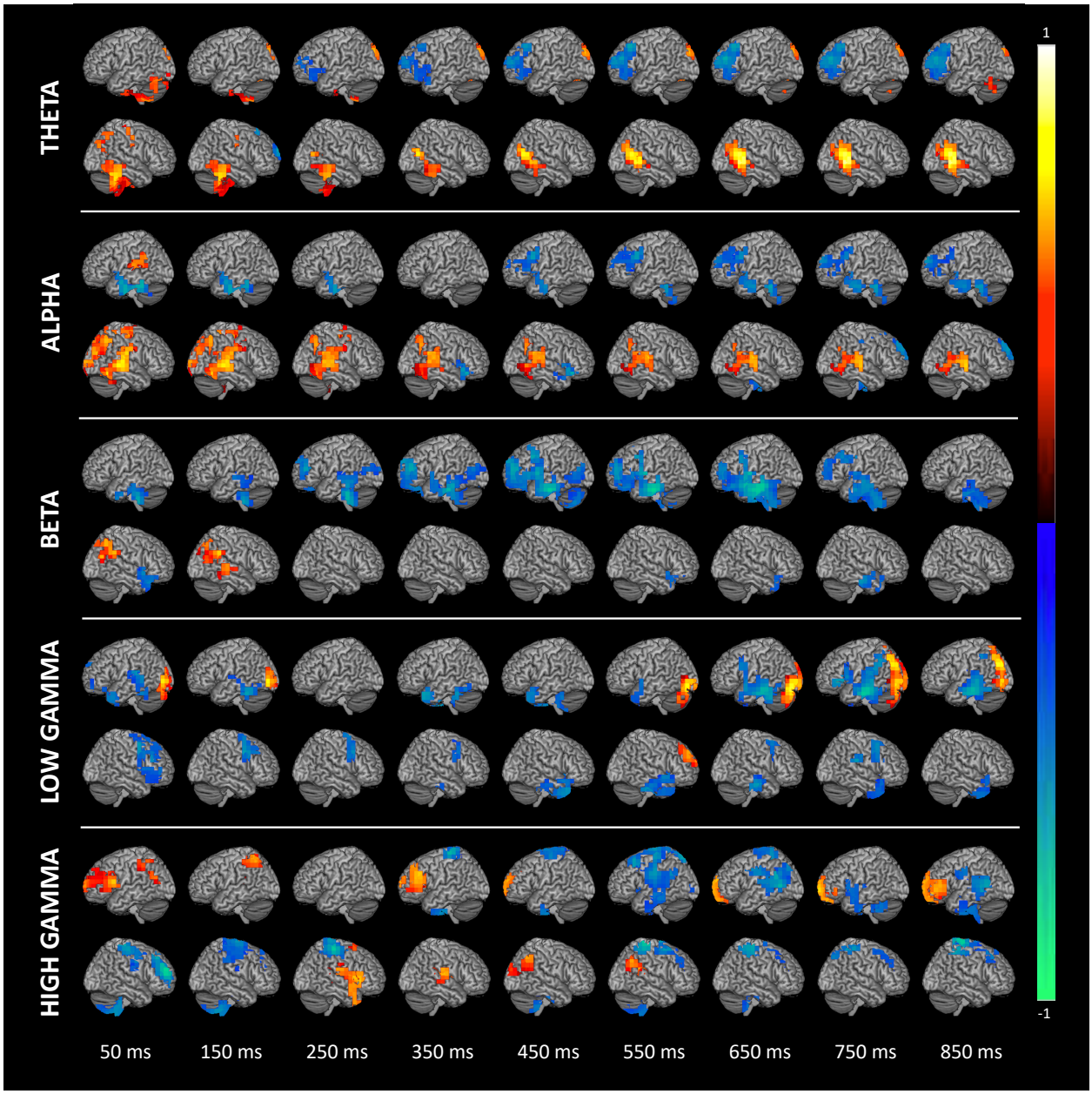

Figure 3.7: Time-frequency dynamics of correlations between neural activity and pitch compensation responses. Significant correlations between neural and behavioral (peak compensation, Figure 3.1c) responses aligned to perturbation onset in response to pitch perturbations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images are cluster corrected, 40 voxels, $p<0.05$. Color scale represents R-value.

Table 3.11: Regions of significant correlations between neural and behavioral responses to pitch feedback perturbations in the theta band (4-7Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-15.0 -51.0 -50.0	50-150 ms	50-350 ms	0.616	0.0033
Left Cuneus	-15.0 -82.0 30.0	550-650 ms	50-950 ms	0.692	0.016
Left Middle Frontal Gyrus	-38.0 27.0 44.0	550-650 ms	350-950 ms	-0.654	0.0019
Right Middle Temporal Gyrus	40.0 -60.0 15.0	50-150 ms	50-150 ms	0.603	7.20E-03
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	53.0 -51.0 -42.0	50-150 ms	50-250 ms	0.593	9.00E-03
Right Precentral Gyrus	39.0 -8.0 36.0	50-150 ms	50-250 ms	0.576	0.0263
Right Fusiform Gyrus	53.0 -42.0 -25.0	50-150 ms	50-550 ms	0.698	9.80E-03
Right Superior Frontal Gyrus	23.0 27.0 60.0	150-250 ms	150-250 ms	-0.601	0.0223
Right Medial Frontal Gyrus	8.0 67.0 13.0	150-250 ms	150-250 ms	-0.561	4.40E-03
Right Middle Temporal Gyrus	48.0 -50.0 -0.0	750-850 ms	350-950 ms	0.900	7.00E-04
Right Superior Temporal Gyrus	48.0 -50.0 15.0	850-950 ms	450-950 ms	0.839	0.0043

Table 3.12: Regions of significant correlations between neural and behavioral responses to pitch feedback perturbations in the alpha band (8-12Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Superior Temporal Gyrus	-41.0 -35.0 14.0	50-150 ms	50-150 ms	0.573	0.0114
Left Fusiform Gyrus	-54.0 -11.0 -31.0	50-150 ms	50-350 ms	-0.662	0.0122
Left Inferior Temporal Gyrus	-62.0 -11.0 -24.0	450-550 ms	450-550 ms	-0.537	0.0276
Left Inferior Frontal Gyrus	-49.0 10.0 30.0	450-550 ms	450-950 ms	-0.608	0.0075
Left Cerebellum, Anterior Lobe, Culmen	-46.0 -45.0 -33.0	650-750 ms	550-950 ms	-0.589	0.0194
Left Fusiform Gyrus	-40.0 -43.0 -24.0	650-750 ms	650-950 ms	-0.648	0.0351
Left Superior Frontal Gyrus	-23.0 59.0 22.0	850-950 ms	650-950 ms	-0.575	0.0031
Right Cuneus	21.0 -97.0 7.0	50-150 ms	50-250 ms	0.628	0.0231
Right Superior Temporal Gyrus	39.0 -26.0 6.0	50-150 ms	50-350 ms	0.760	4.40E-03
Right Postcentral Gyrus	36.0 -28.0 40.0	50-150 ms	50-350 ms	0.722	0.0048
Right Middle Occipital Gyrus	53.0 -64.0 -16.0	50-150 ms	50-450 ms	0.542	0.0392
Right Inferior Parietal Lobule	46.0 -34.0 24.0	350-450 ms	50-750 ms	0.670	0.003
Right Insula	47.0 -28.0 16.0	50-150 ms	50-950 ms	0.721	2.50E-03
Right Superior Temporal Gyrus	45.0 21.0 -25.0	350-450 ms	350-450 ms	-0.399	0.043
Right Inferior Frontal Gyrus	47.0 31.0 -14.0	450-550 ms	350-550 ms	-0.589	0.023
Right Superior Temporal Gyrus	54.0 -19.0 -1.0	850-950 ms	650-950 ms	0.687	0.0208
Right Superior Frontal Gyrus	16.0 59.0 29.0	850-950 ms	750-950 ms	-0.586	0.0065

Table 3.13: Regions of significant correlations between neural and behavioral responses to pitch feedback perturbations in the beta band (13-30Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Middle Temporal Gyrus	-47.0 4.0 -40.0	50-150 ms	50-150 ms	-0.506	3.70E-02
Left Parahippocampal Gyrus	-22.0 -27.0 -9.0	50-150 ms	50-350 ms	-0.588	5.00E-04
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-32.0 -34.0 -40.0	250-350 ms	50-450 ms	-0.704	1.00E-04
Left Middle Frontal Gyrus	-32.0 46.0 12.0	350-450 ms	250-750 ms	-0.653	0.0036
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-30.0 -53.0 -49.0	750-850 ms	250-950 ms	-0.539	0.0073
Left Superior Temporal Gyrus	-55.0 3.0 5.0	450-550 ms	350-750 ms	-0.662	0.0037
Left Inferior Temporal Gyrus	-62.0 -21.0 -26.0	550-650 ms	350-850 ms	-0.841	0.0029
Left Middle Frontal Gyrus	-40.0 13.0 37.0	450-550 ms	450-850 ms	-0.655	0.0024
Right Superior Temporal Gyrus	32.0 18.0 -27.0	50-150 ms	50-150 ms	-0.523	4.20E-03
Right Middle Temporal Gyrus	53.0 -67.0 23.0	50-150 ms	50-250 ms	0.638	4.70E-03
Right Inferior Parietal Lobule	39.0 -41.0 40.0	150-250 ms	150-250 ms	0.753	0.0192
Right Superior Temporal Gyrus	54.0 -27.0 5.0	150-250 ms	150-250 ms	0.584	0.0302
Right Superior Temporal Gyrus	53.0 19.0 -24.0	550-650 ms	550-950 ms	-0.429	0.0299
Right Inferior Temporal Gyrus	54.0 -20.0 -24.0	750-850 ms	750-850 ms	-0.584	0.0029

Table 3.14: Regions of significant correlations between neural and behavioral responses to pitch feedback perturbations in the low gamma band (30-55Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Transverse Temporal Gyrus	-41.0 -27.0 13.0	50-150 ms	50-150 ms	-0.587	0.0048
Left Superior Temporal Gyrus	-31.0 10.0 -25.0	50-150 ms	50-150 ms	-0.539	0.0146
Left Medial Frontal Gyrus	-8.0 50.0 22.0	50-150 ms	50-150 ms	-0.528	0.0127
Left Superior Frontal Gyrus	-32.0 52.0 -17.0	50-150 ms	50-150 ms	-0.380	0.0304
Left Middle Occipital Gyrus	-39.0 -84.0 -1.0	150-250 ms	50-250 ms	0.742	1.26E-02
Left Fusiform Gyrus	-48.0 -43.0 -24.0	150-250 ms	50-250 ms	-0.621	0.0083
Left Inferior Temporal Gyrus	-48.0 -10.0 -40.0	350-450 ms	350-450 ms	-0.548	2.68E-02
Left Superior Temporal Gyrus	-38.0 19.0 -25.0	350-450 ms	350-550 ms	-0.626	0.0181
Left Fusiform Gyrus	-41.0 -45.0 -24.0	350-450 ms	350-550 ms	-0.577	0.0054
Left Cerebellum, Anterior Lobe, Culmen	-47.0 -45.0 -35.0	350-450 ms	350-550 ms	-0.537	0.0041
Left Cerebellum, Posterior Lobe, Declive	-45.0 -82.0 -28.0	650-750 ms	550-850 ms	0.736	0.0025
Left Inferior Occipital Gyrus	-38.0 -83.0 -13.0	650-750 ms	550-950 ms	0.800	0.0022
Left Inferior Frontal Gyrus	-48.0 11.0 31.0	650-750 ms	650-750 ms	-0.487	0.0277
Left Cerebellum, Anterior Lobe, Culmen	-39.0 -44.0 -28.0	650-750 ms	650-950 ms	-0.612	0.0022
Left Inferior Temporal Gyrus	-62.0 -19.0 -26.0	750-850 ms	650-950 ms	-0.789	0.0035
Left Inferior Frontal Gyrus	-54.0 28.0 -2.0	750-850 ms	750-850 ms	-0.634	0.0388
Left Supramarginal Gyrus	-40.0 -43.0 39.0	750-850 ms	750-950 ms	-0.622	0.0105
Left Superior Occipital Gyrus	-40.0 -83.0 30.0	850-950 ms	750-950 ms	0.748	0.0063
Right Middle Frontal Gyrus	46.0 13.0 39.0	250-350 ms	50-450 ms	-0.523	2.60E-03
Right Superior Temporal Gyrus	40.0 12.0 -40.0	450-550 ms	450-650 ms	-0.536	3.00E-04
Right Fusiform Gyrus	39.0 -25.0 -31.0	550-650 ms	450-750 ms	-0.601	0.0156
Right Inferior Temporal Gyrus	63.0 -20.0 -24.0	650-750 ms	450-850 ms	-0.581	0.0139
Right Superior Frontal Gyrus	22.0 52.0 38.0	550-650 ms	550-650 ms	0.613	0.0061
Right Middle Frontal Gyrus	45.0 13.0 52.0	750-850 ms	650-850 ms	-0.555	0.0154
Right Insula	47.0 -27.0 21.0	750-850 ms	750-850 ms	-0.616	0.016
Right Middle Temporal Gyrus	39.0 5.0 -42.0	850-950 ms	750-950 ms	-0.438	0.0037

Table 3.15: Regions of significant correlations between neural and behavioral responses to pitch feedback perturbations in the high gamma band (65-150Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Precentral Gyrus	-49.0 11.0 15.0	12.5-37.5 ms	12.5-87.5 ms	0.689	2.11E-02
Left Superior Frontal Gyrus	-13.0 66.0 5.0	62.5-87.5 ms	12.5-87.5 ms	0.594	0.0146
Left Middle Temporal Gyrus	-39.0 -68.0 23.0	37.5-62.5 ms	37.5-62.5 ms	0.538	0.0493
Left Inferior Parietal Lobule	-43.0 -53.0 54.0	112.5-137.5 ms	37.5-212.5 ms	0.603	1.80E-03
Left Middle Frontal Gyrus	-32.0 27.0 32.0	337.5-362.5 ms	312.5-412.5 ms	0.676	0.0172
Left Precentral Gyrus	-31.0 -28.0 70.0	312.5-337.5 ms	312.5-737.5 ms	-0.736	0.0091
Left Superior Frontal Gyrus	-23.0 60.0 20.0	337.5-362.5 ms	337.5-512.5 ms	0.609	0.0361
Left Inferior Temporal Gyrus	-40.0 -3.0 -42.0	462.5-487.5 ms	337.5-562.5 ms	-0.514	0.0016
Left Middle Frontal Gyrus	-32.0 20.0 53.0	537.5-562.5 ms	487.5-587.5 ms	-0.617	0.0055
Left Postcentral Gyrus	-62.0 -20.0 37.0	562.5-587.5 ms	487.5-587.5 ms	-0.622	0.0045
Left Supramarginal Gyrus	-63.0 -53.0 23.0	662.5-687.5 ms	512.5-762.5 ms	-0.699	0.0011
Left Middle Temporal Gyrus	-47.0 -60.0 6.0	637.5-662.5 ms	562.5-712.5 ms	-0.703	0.0037
Left Postcentral Gyrus	-21.0 -43.0 71.0	662.5-687.5 ms	637.5-737.5 ms	-0.612	0.0043
Left Cerebellum, Anterior Lobe, Culmen	-31.0 -53.0 -34.0	737.5-762.5 ms	687.5-762.5 ms	-0.449	0.0088
Left Inferior Temporal Gyrus	-47.0 -10.0 -39.0	712.5-737.5 ms	687.5-912.5 ms	-0.544	0.0044
Left Middle Temporal Gyrus	-50.0 -3.0 -17.0	737.5-762.5 ms	712.5-787.5 ms	-0.374	0.0307
Left Medial Frontal Gyrus	-16.0 68.0 -1.0	737.5-762.5 ms	737.5-887.5 ms	0.699	0.0124
Left Inferior Frontal Gyrus	-47.0 29.0 6.0	812.5-837.5 ms	787.5-962.5 ms	0.671	0.0205
Left Middle Temporal Gyrus	-64.0 -53.0 5.0	812.5-837.5 ms	812.5-912.5 ms	-0.724	0.0081
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	24.0 -44.0 -54.0	62.5-87.5 ms	12.5-212.5 ms	-0.600	4.70E-03
Right Postcentral Gyrus	45.0 -27.0 60.0	237.5-262.5 ms	12.5-312.5 ms	-0.793	0.0027
Right Middle Frontal Gyrus	25.0 -12.0 55.0	62.5-87.5 ms	37.5-237.5 ms	-0.800	1.80E-03
Right Superior Frontal Gyrus	21.0 -10.0 69.0	62.5-87.5 ms	62.5-212.5 ms	-0.670	0.0176
Right Superior Temporal Gyrus	48.0 12.0 -24.0	212.5-237.5 ms	162.5-337.5 ms	0.626	0.018
Right Precentral Gyrus	63.0 12.0 8.0	287.5-312.5 ms	162.5-337.5 ms	0.682	0.0103

Right Transverse Temporal Gyrus	47.0 -21.0 13.0	337.5-362.5 ms	212.5-387.5 ms	0.654	0.0116
Right Superior Frontal Gyrus	23.0 13.0 61.0	262.5-287.5 ms	237.5-287.5 ms	0.509	0.0359
Right Insula	46.0 -18.0 20.0	287.5-312.5 ms	287.5-312.5 ms	0.717	0.0032
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	30.0 -36.0 -54.0	437.5-462.5 ms	362.5-587.5 ms	-0.486	0.0184
Right Fusiform Gyrus	56.0 -19.0 -32.0	412.5-437.5 ms	387.5-437.5 ms	-0.408	0.0306
Right Cuneus	29.0 -89.0 22.0	437.5-462.5 ms	437.5-487.5 ms	0.457	0.0131
Right Supramarginal Gyrus	54.0 -51.0 38.0	512.5-537.5 ms	437.5-562.5 ms	0.751	0.0061
Right Superior Temporal Gyrus	40.0 -35.0 14.0	562.5-587.5 ms	437.5-637.5 ms	0.629	0.0076
Right Middle Frontal Gyrus	38.0 3.0 60.0	562.5-587.5 ms	487.5-662.5 ms	-0.623	0.0029
Right Postcentral Gyrus	31.0 -34.0 69.0	537.5-562.5 ms	512.5-662.5 ms	-0.763	0.0066
Right Superior Frontal Gyrus	23.0 -5.0 71.0	537.5-562.5 ms	537.5-662.5 ms	-0.442	0.0272
Right Middle Frontal Gyrus	37.0 28.0 45.0	862.5-887.5 ms	662.5-912.5 ms	-0.539	0.0152
Right Precentral Gyrus	15.0 -27.0 71.0	837.5-862.5 ms	787.5-937.5 ms	-0.802	0.0012

In the alpha band, similar to in the theta band, neural activation in the left hemisphere mostly have negative correlations to the behavioral responses to pitch perturbations, while in the right hemisphere they are mostly positively correlated. The negative correlations in the left hemisphere have an early peak in the fusiform gyrus but are mostly observed post perturbation offset, with peaks in ITG, IFG, SFG, fusiform gyrus and cerebellum. Several negative correlations in the right hemisphere are observed near and after perturbation offset, with peaks in the STG, IFG and SFG. Positive correlation in the left hemisphere is only observed transiently after perturbation onset in the STG. In the right hemisphere, the positive correlations peaks early post perturbation onset in the cuneus, STG, postcentral gyrus, MOG, IPL and insula and are sustained throughout the rest of the time windows after perturbation offset with another peak in STG.

In the beta band, neural activation to pitch perturbations mostly have negative correlations to the behavioral responses in the left hemisphere, with only transient positive correlations in the right hemisphere. In the left hemisphere, only negative correlations are observed throughout all the time windows, with some peaks in the perturbation window (MTG, parahippocampal gyrus, cerebellum, and MFG) and peaks post perturbation offset (STG, ITG, MFG, and cerebellum). In the right hemisphere, positive correlations are only observed early post perturbation onset, with peaks in the MTG, STG, and IPL. Negative correlations in the right hemisphere showed transiently early post perturbation onset in STG, and again later post perturbation offset in STG and ITG.

In the low gamma frequency band, positive correlations are mostly found in the left hemisphere while negative correlations are observed bilaterally in various regions. Positive correlations in the left hemisphere peaks early post perturbation onset in the MOG, and later post perturbation offset in IOG, SOG and cerebellum. Positive correlation in the right hemisphere is only observed post perturbation offset in SFG. Negative correlations early post perturbation onset is observed in the left temporal and frontal lobes (transverse temporal gyrus (TTG), STG, MFG, SFG, and fusiform gyrus) and in the right MFG, and later closer but before perturbation offset in left ITG, STG and fusiform gyrus. Post perturbation offset, negative correlations are observed more strongly in the left hemisphere, with peaks in the IFG, ITG, SMG and cerebellum. In the right hemisphere post perturbation offset negative correlations are observed in MFG, MTG and insula.

In the high gamma frequency band, both negative and positive correlations are spread bilaterally. Positive correlations in the left hemisphere are mostly observed in the frontal regions, where peaks are observed early post perturbation onset in precentral gyrus, SFG, and MFG and much later post perturbation offset in MFG and IFG, as well as in MTG and IPL early post perturbation onset. Positive correlations in the right hemisphere are observed within the perturbation window in TTG, STG, precentral gyrus, SFG, and insula as well as shortly after perturbation offset in cuneus, SMG and STG. Negative correlations in the left hemisphere are mostly observed late in the perturbation window with wide negatively correlated regions observed post perturbation offset. In the left hemisphere, peak negative correlation in the late perturbation window is observed in the precentral gyrus, whereas peaks post perturbation offset are observed in MFG, postcentral gyrus, SMG, ITG, MTG and cerebellum. In the right hemisphere, negative correlations within the perturbation window are observed in MFG, SFG, postcentral gyrus, MTG and cerebellum, while negative correlations post perturbation offset are observed in frontal lobe (SFG, MFG, precentral gyrus, postcentral gyrus), fusiform gyrus and cerebellum.

3.4.5 Neural-behavioral correlations in responses to formant perturbations

Figure 3.8 shows regions that are correlation to the peak compensation responses to formant perturbations (with peak of correlations shown in Table 3.16-3.20). In the theta band, positive correlations are mostly observed bilaterally with some negative correlations in the left hemisphere post perturbation offset. In the left hemisphere, only positive correlations are observed, which are sustained throughout most of the time windows. Peaks of positive correlation within the perturbation time window are observed in inferior and middle occipital gyrus as well as cerebellum, followed by peaks in IFG, SFG, precuneus and ITG post

perturbation offset. In the right hemisphere, strong positive correlations are observed early in the perturbation window, with peaks in the MFG, SFG, IPL, MOG, angular gyrus and MTG, followed by a later peak in IFG post perturbation offset. Negative correlations are only observed post perturbation offset in MFG, ITG and cerebellum.

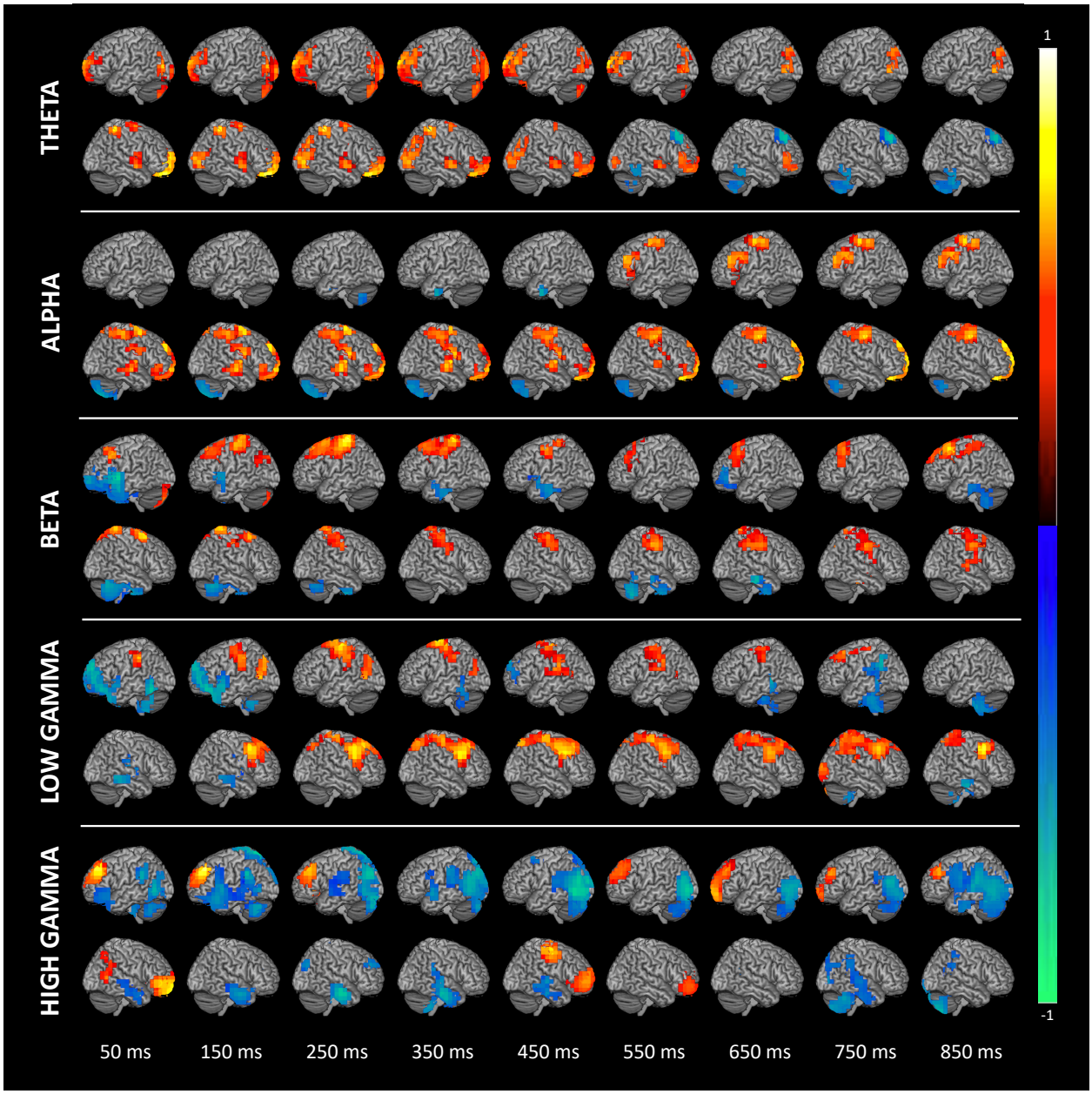

Figure 3.8: Time-frequency dynamics of correlations between neural activity and formant compensation responses. Significant correlations between neural and behavioral (peak compensation, Figure 3.1d) responses aligned to perturbation onset in response to formant perturbations across different frequency bands: theta (4-7Hz), alpha (8-12Hz), beta (13-30Hz), low gamma (30-55Hz) and high gamma (65-150Hz notched at 120Hz). Images are cluster corrected, 40 voxels, $p<0.05$. Color scale represents R-value.

Table 3.16: Regions of significant correlations between neural and behavioral responses to formant feedback perturbations in the theta band (4-7Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Inferior Occipital Gyrus	-45.0 -84.0 -8.0	50-150 ms	50-450 ms	0.536	6.70E-03
Left Superior Frontal Gyrus	-7.0 66.0 13.0	550-650 ms	50-750 ms	0.754	0.0035
Left Middle Occipital Gyrus	-51.0 -81.0 7.0	50-150 ms	50-950 ms	0.65	0.0057
Left Cerebellum, Posterior Lobe, Tuber	-47.0 -82.0 -33.0	350-450 ms	150-650 ms	0.592	1.00E-04
Left Inferior Frontal Gyrus	-41.0 43.0 12.0	450-550 ms	150-650 ms	0.586	0.0252
Left Precuneus	-38.0 -75.0 36.0	750-850 ms	450-950 ms	0.573	1.00E-03
Left Inferior Temporal Gyrus	-57.0 -68.0 -2.0	850-950 ms	450-950 ms	0.645	1.00E-04
Right Medial Frontal Gyrus	15.0 59.0 8.0	50-150 ms	50-450 ms	0.744	2.00E-04
Right Inferior Parietal Lobule	39.0 -44.0 60.0	250-350 ms	50-550 ms	0.793	1.00E-04
Right Middle Temporal Gyrus	56.0 -3.0 -9.0	350-450 ms	50-650 ms	0.602	0.0044
Right Superior Frontal Gyrus	15.0 51.0 -12.0	50-150 ms	50-950 ms	0.771	0.0024
Right Middle Occipital Gyrus	33.0 -92.0 5.0	250-350 ms	150-650 ms	0.663	0.0025
Right Angular Gyrus	46.0 -80.0 29.0	250-350 ms	250-550 ms	0.652	0.0222
Right Inferior Frontal Gyrus	54.0 35.0 4.0	650-750 ms	350-750 ms	0.548	0.018
Right Inferior Temporal Gyrus	62.0 -51.0 -17.0	550-650 ms	550-750 ms	-0.563	0.0464
Right Middle Frontal Gyrus	25.0 35.0 45.0	750-850 ms	550-950 ms	-0.778	0.0018
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	40.0 -60.0 -58.0	750-850 ms	650-950 ms	-0.546	0.0182

Table 3.17: Regions of significant correlations between neural and behavioral responses to formant feedback perturbations in the alpha band (8-12Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-39.0 -60.0 -42.0	250-350 ms	250-350 ms	-0.523	0.0155
Left Inferior Temporal Gyrus	-54.0 -3.0 -35.0	450-550 ms	350-550 ms	-0.664	1.43E-02
Left Postcentral Gyrus	-46.0 -27.0 60.0	650-750 ms	550-950 ms	0.651	0.0097
Left Middle Frontal Gyrus	-41.0 12.0 30.0	750-850 ms	550-950 ms	0.638	0.0084
Left Precentral Gyrus	-31.0 -11.0 61.0	850-950 ms	550-950 ms	0.724	0.0163
Right Cerebellum, Posterior Lobe, Inferior Semi-Lunar Lobule	32.0 -78.0 -66.0	50-150 ms	50-250 ms	-0.732	0.0101
Right Superior Frontal Gyrus	18.0 44.0 30.0	150-250 ms	50-350 ms	0.831	6.00E-04
Right Precuneus	15.0 -52.0 55.0	150-250 ms	50-350 ms	0.512	2.14E-02
Right Middle Frontal Gyrus	25.0 3.0 65.0	150-250 ms	50-450 ms	0.713	0.008
Right Precentral Gyrus	49.0 -4.0 7.0	350-450 ms	50-650 ms	0.721	0.0011
Right Superior Temporal Gyrus	49.0 -12.0 -2.0	250-350 ms	50-750 ms	0.778	1.00E-04
Right Insula	40.0 -2.0 16.0	250-350 ms	50-750 ms	0.733	0.0035
Right Superior Frontal Gyrus	23.0 -3.0 71.0	250-350 ms	50-950 ms	0.754	0.0029
Right Medial Frontal Gyrus	8.0 61.0 -9.0	250-350 ms	50-950 ms	0.738	1.00E-03
Right Rectal Gyrus	7.0 35.0 -24.0	650-750 ms	50-950 ms	0.773	0.0029
Right Superior Frontal Gyrus	14.0 58.0 36.0	850-950 ms	50-950 ms	0.800	8.00E-04
Right Medial Frontal Gyrus	7.0 68.0 12.0	850-950 ms	50-950 ms	0.767	0.0091

Table 3.18: Regions of significant correlations between neural and behavioral responses to formant feedback perturbations in the beta band (13-30Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Middle Frontal Gyrus	-49.0 12.0 45.0	50-150 ms	50-150 ms	0.593	0.0112
Left Superior Frontal Gyrus	-31.0 57.0 -10.0	50-150 ms	50-150 ms	-0.508	6.60E-03
Left Superior Temporal Gyrus	-54.0 5.0 -2.0	50-150 ms	50-350 ms	-0.819	2.00E-04
Left Postcentral Gyrus	-32.0 -35.0 70.0	250-350 ms	150-550 ms	0.791	7.00E-04
Left Superior Frontal Gyrus	-7.0 30.0 48.0	250-350 ms	150-450 ms	0.668	0.0021
Left Precentral Gyrus	-49.0 -11.0 53.0	450-550 ms	250-550 ms	0.600	2.00E-04
Left Middle Temporal Gyrus	-61.0 3.0 -9.0	450-550 ms	350-550 ms	-0.598	4.60E-03
Left Inferior Temporal Gyrus	-54.0 -12.0 -24.0	450-550 ms	350-550 ms	-0.541	0.0294
Left Middle Frontal Gyrus	-45.0 50.0 13.0	650-750 ms	650-750 ms	-0.565	2.91E-02
Left Middle Frontal Gyrus	-26.0 21.0 52.0	750-850 ms	650-950 ms	0.746	5.00E-04
Left Inferior Temporal Gyrus	-54.0 -19.0 -24.0	850-950 ms	850-950 ms	-0.477	0.0288
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-55.0 -53.0 -43.0	850-950 ms	850-950 ms	-0.476	1.56E-02
Right Middle Frontal Gyrus	22.0 10.0 68.0	50-150 ms	50-250 ms	0.776	0.0039
Right Cerebellum, Posterior Lobe, Tuber	41.0 -60.0 -32.0	50-150 ms	50-350 ms	-0.647	1.59E-02
Right Postcentral Gyrus	32.0 -35.0 70.0	150-250 ms	50-450 ms	0.579	4.50E-03
Right Precentral Gyrus	40.0 -20.0 48.0	550-650 ms	150-750 ms	0.658	0.0074
Right Inferior Temporal Gyrus	38.0 -5.0 -41.0	250-350 ms	250-350 ms	-0.506	0.0326
Right Inferior Temporal Gyrus	47.0 -2.0 -40.0	550-650 ms	250-650 ms	-0.517	0.0178
Right Precentral Gyrus	56.0 -10.0 45.0	850-950 ms	250-950 ms	0.577	0.0037
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	54.0 -60.0 -49.0	550-650 ms	450-750 ms	-0.684	0.0101
Right Middle Temporal Gyrus	65.0 -18.0 -15.0	650-750 ms	550-750 ms	-0.474	0.0363
Right Postcentral Gyrus	22.0 -51.0 69.0	750-850 ms	650-950 ms	0.495	0.0089
Right Superior Temporal Gyrus	70.0 -20.0 5.0	850-950 ms	850-950 ms	0.521	0.0256

Table 3.19: Regions of significant correlations between neural and behavioral responses to formant feedback perturbations in the low gamma band (30-55Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Temporal Gyrus	-63.0 -60.0 -10.0	50-150 ms	50-150 ms	-0.717	8.00E-04
Left Superior Frontal Gyrus	-16.0 51.0 27.0	50-150 ms	50-250 ms	-0.807	0.0038
Left Middle Frontal Gyrus	-24.0 35.0 -2.0	50-150 ms	50-250 ms	-0.754	0.0092
Left Cerebellum, Posterior Lobe, Cerebellar Tonsil	-46.0 -52.0 -57.0	50-150 ms	50-250 ms	-0.600	0.0215
Left Inferior Frontal Gyrus	-40.0 28.0 -17.0	150-250 ms	50-250 ms	-0.735	1.20E-02
Left Inferior Parietal Lobule	-61.0 -36.0 44.0	50-150 ms	50-550 ms	0.675	0.0029
Left Middle Temporal Gyrus	-41.0 -68.0 5.0	150-250 ms	150-350 ms	0.671	0.0089
Left Precuneus	-39.0 -76.0 38.0	150-250 ms	150-450 ms	0.607	0.0124
Left Precentral Gyrus	-40.0 -20.0 63.0	250-350 ms	150-750 ms	0.702	2.00E-04
Left Superior Frontal Gyrus	-8.0 -4.0 71.0	350-450 ms	250-550 ms	0.683	1.10E-03
Left Inferior Temporal Gyrus	-62.0 -52.0 -17.0	350-450 ms	350-450 ms	-0.509	0.0047
Left Inferior Parietal Lobule	-58.0 -33.0 28.0	450-550 ms	450-650 ms	0.589	0.0139
Left Supramarginal Gyrus	-39.0 -53.0 30.0	750-850 ms	750-850 ms	0.715	0.0218
Left Middle Frontal Gyrus	-38.0 19.0 53.0	750-850 ms	750-850 ms	0.585	0.003
Left Inferior Temporal Gyrus	-62.0 -35.0 -24.0	750-850 ms	750-850 ms	-0.516	0.0257
Right Middle Temporal Gyrus	64.0 -42.0 -9.0	50-150 ms	50-250 ms	-0.631	0.0287
Right Middle Frontal Gyrus	42.0 10.0 30.0	350-450 ms	150-950 ms	0.734	1.50E-03
Right Superior Parietal Lobule	24.0 -68.0 54.0	450-550 ms	250-550 ms	0.642	1.56E-02
Right Precentral Gyrus	31.0 -28.0 70.0	450-550 ms	250-950 ms	0.661	0.0029
Right Middle Occipital Gyrus	31.0 -98.0 12.0	750-850 ms	750-850 ms	0.632	0.0119
Right Cerebellum, Posterior Lobe, Uvula	8.0 -88.0 -32.0	750-850 ms	750-850 ms	0.519	0.0294
Right Cerebellum, Posterior Lobe, Cerebellar Tonsil	45.0 -46.0 -55.0	750-850 ms	750-950 ms	-0.559	0.0379
Right Middle Frontal Gyrus	55.0 11.0 38.0	850-950 ms	750-950 ms	0.753	0.0044
Right Inferior Temporal Gyrus	62.0 -20.0 -24.0	850-950 ms	850-950 ms	-0.662	0.0119

Table 3.20: Regions of significant correlations between neural and behavioral responses to formant feedback perturbations in the high gamma band (65-150Hz)

Region	MNI peak voxel	Peak	Duration	R-value	P-value
Left Middle Occipital Gyrus	-41.0 -66.0 0.0	37.5-62.5 ms	12.5-112.5 ms	-0.701	0.0139
Left Inferior Parietal Lobule	-64.0 -43.0 31.0	112.5-137.5 ms	12.5-187.5 ms	-0.746	4.00E-04
Left Cerebellum	-55.0 -68.0 -41.0	137.5-162.5 ms	12.5-237.5 ms	-0.663	7.00E-04
Left Superior Frontal Gyrus	-31.0 44.0 37.0	12.5-37.5 ms	12.5-312.5 ms	0.827	4.00E-04
Left Superior Temporal Gyrus	-46.0 11.0 -17.0	112.5-137.5 ms	62.5-162.5 ms	-0.677	0.0028
Left Postcentral Gyrus	-7.0 -58.0 71.0	162.5-187.5 ms	87.5-512.5 ms	-0.807	2.50E-03
Left Precuneus	-7.0 -68.0 22.0	162.5-187.5 ms	112.5-212.5 ms	-0.680	0.0177
Left Medial Frontal Gyrus	-7.0 -29.0 71.0	187.5-212.5 ms	112.5-262.5 ms	-0.744	0.01
Left Cerebellum	-45.0 -82.0 -32.0	212.5-237.5 ms	137.5-562.5 ms	-0.774	0.0016
Left Angular Gyrus	-38.0 -80.0 29.0	262.5-287.5 ms	187.5-437.5 ms	-0.781	0.0125
Left Middle Occipital Gyrus	-39.0 -67.0 -9.0	462.5-487.5 ms	187.5-937.5 ms	-0.873	1.60E-03
Left Middle Temporal Gyrus	-47.0 -77.0 13.0	462.5-487.5 ms	212.5-962.5 ms	-0.832	0.0028
Left Superior Temporal Gyrus	-61.0 3.0 -2.0	312.5-337.5 ms	312.5-362.5 ms	-0.625	0.0056
Left Middle Frontal Gyrus	-38.0 35.0 30.0	887.5-912.5 ms	512.5-962.5 ms	0.670	0.0061
Left Superior Frontal Gyrus	-16.0 66.0 -9.0	662.5-687.5 ms	562.5-762.5 ms	0.682	0.0042
Left Superior Temporal Gyrus	-41.0 -28.0 5.0	837.5-862.5 ms	812.5-962.5 ms	-0.672	0.0011
Right Middle Temporal Gyrus	61.0 -58.0 5.0	12.5-37.5 ms	12.5-87.5 ms	0.590	8.90E-03
Right Medial Frontal Gyrus	13.0 66.0 -16.0	12.5-37.5 ms	12.5-112.5 ms	0.801	1.00E-04
Right Inferior Temporal Gyrus	56.0 -13.0 -26.0	237.5-262.5 ms	87.5-437.5 ms	-0.787	1.80E-03
Right Superior Frontal Gyrus	24.0 45.0 31.0	237.5-262.5 ms	187.5-262.5 ms	-0.584	0.0104
Right Angular Gyrus	38.0 -76.0 31.0	237.5-262.5 ms	237.5-262.5 ms	-0.496	0.0489
Right Cerebellum	32.0 -52.0 -48.0	362.5-387.5 ms	312.5-387.5 ms	-0.636	0.0329
Right Insula	55.0 -35.0 21.0	337.5-362.5 ms	337.5-387.5 ms	-0.538	0.0199
Right Superior Frontal Gyrus	23.0 52.0 5.0	412.5-437.5 ms	362.5-537.5 ms	0.627	0.0017
Right Precentral Gyrus	46.0 -11.0 61.0	412.5-437.5 ms	387.5-512.5 ms	0.751	0.0037
Right Superior Temporal Gyrus	64.0 -19.0 -2.0	712.5-737.5 ms	662.5-762.5 ms	-0.484	0.0061
Right Fusiform Gyrus	56.0 -20.0 -32.0	737.5-762.5 ms	662.5-762.5 ms	-0.453	0.0199

Right Supramarginal Gyrus	43.0 -49.0 28.0	812.5-837.5 ms	712.5-912.5 ms	-0.610	0.0062
Right Middle Occipital Gyrus	47.0 -84.0 -2.0	737.5-762.5 ms	737.5-762.5 ms	-0.579	0.0054
Right Cerebellum	39.0 -75.0 -55.0	812.5-837.5 ms	737.5-887.5 ms	-0.739	0.0033

In the alpha band, sustained positive correlations are observed in the right hemisphere, with some positive correlations post perturbation offset in the left hemisphere and sustained negative correlation in the right cerebellum. In the left hemisphere, positive correlations are only observed post perturbation offset in postcentral, precentral and middle frontal gyrus, while negative correlations are observed in the late perturbation window in cerebellum and ITG. In the right hemisphere, negative correlation is only observed in the cerebellum which peaks early post perturbation onset, while positive correlations are observed mostly in frontal lobe. Peaks of positive correlations in the right hemisphere within the perturbation window are observed in SFG, MFG, precentral gyrus, STG, precuneus, and insula, while peaks post perturbation offset are observed in SFG, MFG and rectal gyrus.

In the beta band, positive correlations are observed in the frontal and motor regions, while negative correlations are observed in frontal, temporal and cerebellar regions. In the left hemisphere, positive correlations are stronger within the perturbation window, with peaks in MFG, SFG, and postcentral gyrus and later peaks post perturbation offset in precentral gyrus and MFG. Peak negative correlation is observed early post perturbation onset in SFG and STG, while weaker peaks of negative correlations are observed later in MTG, ITG, MFG, and cerebellum. In the right hemisphere, positive correlations are sustained throughout the time windows, with early peaks post perturbation onset in MFG and postcentral gyrus, followed by later peaks post perturbation offset in precentral gyrus, postcentral gyrus and STG. Transient negative

correlations in the right hemisphere are observed shortly after perturbation onset and perturbation offset in the cerebellum, ITG and MTG.

In the low gamma band, similar to beta band, positive correlations are mostly observed in frontal and motor regions while negative correlations are mostly observed in the frontal, temporal and cerebellar regions. In the left hemisphere, an early peak positive correlation is observed in IPL, MTG and precuneus, then peaks in the late perturbation window in precentral gyrus and SFG which are sustained for a few hundred milliseconds post perturbation offset. Strong negative correlations in the left hemisphere are observed early post perturbation onset in temporal gyrus, SFG, MFG, IFG and cerebellum, late within the perturbation window in ITG and IPL, then late post perturbation offset in MSG, MFG and ITG. In the right hemisphere, negative correlations are observed transiently post perturbation onset in MTG and much later post perturbation offset in ITG and cerebellum. Positive correlations in the right hemisphere, however, began shortly after perturbation onset and are sustained throughout the rest of the time windows, with peaks in MFG, precentral gyrus, SPL, MOG and cerebellum.

In the high gamma, positive correlations are mostly observed in the frontal regions while negative correlations are observed throughout a wider region. Strong positive correlations are observed early post perturbation onset in left SFG, right MTG and right MFG, as well as post perturbation offset first in right SFG and precentral gyrus and then in left MFG and SFG. Negative correlations in the left hemisphere are observed within the perturbation window in MOG, IPL, cerebellum, STG, postcentral gyrus, precuneus, MFG, and angular gyrus, followed by peak negative correlations post perturbation offset in MOG, MTG, and STG. Negative

correlations in the right hemisphere are sustained within the perturbation time window in ITG, SFG, angular gyrus, insula and cerebellum, and are also observed much later post perturbation offset in STG, fusiform gyrus, SMG, MOG and cerebellum.

3.5 Discussion

In this study, we explored the cortical mechanisms involved in the control of pitch and formants by studying both behavioral and neural responses to pitch and formant perturbations using MEG. Previous studies which have investigated responses to formant feedback perturbations utilized fMRI (Niziolek & Guenther, 2013; Tourville et al., 2008), which lack the temporal resolution that MEG offers. Moreover, since it is possible that multiple frequency bands correlate with BOLD fMRI activity (Hall et al., 2014), we decided to explore this feedback perturbation processing across all frequency bands. With MEG, we were able to observe a large network of regions across all frequency bands that are involved in feedback processing for pitch and/or formants, with some regions more specific for either pitch or formant feedback processing. Though we listed all activations observed in Tables 3.1-3.20, here we only focus on regions that are observed across multiple time windows and more than two frequency bands. We also discuss how these regions compared to findings from previous studies by highlighting similarities as well as regions that have not been previously observed, most likely due to its transient activity that is not observable with fMRI. Lastly, we explored the correlations between the neural activation to the peak compensation responses to pitch and formant perturbations, respectively. Given that participant's compensation responses to the feedback perturbations on average were feature-specific, we can obtain the correlation between the feature-specific compensation responses to the feature-specific neural activity observed using MEG. We performed a data-

driven correlation and reported all observations to be comprehensive, however here we will only discuss regions showing correlation that also showed activation, since it would be difficult to interpret otherwise.

3.5.1 Neural regions involved in pitch and formant feedback processing

A wide network of activity throughout all frequency bands is involved for pitch and formant pitch processing, with some regions showing sustained activity throughout all time windows and others showing processing-specific activity. Notable regions that have been previously been observed in previous studies include STG, MTG, MFG, IFG, SMC, MC, IPL and cerebellum (Kort et al., 2016; Niziolek & Guenther, 2013; Parkinson et al., 2012; Ranasinghe et al., 2019; Tourville et al., 2008).

Activity in the STG has been bilaterally observed in both formant and pitch perturbation fMRI studies (Parkinson et al., 2012; Tourville et al., 2008). In this study, its suppression is observed anteriorly throughout all time windows in the right hemisphere (theta band), while its enhanced activity is observed posteriorly in the first 350ms bilaterally (beta and high gamma bands) with peaks in the first 50ms time window. Our findings indicated that this region is involved in both pitch and formant feedback processing, which is in line with what was found in previous fMRI studies. Furthermore, as can be observed from the contrast map, the right STG seemed to be more involved for pitch processing in theta, alpha, and beta (bilateral STG).

Activity in the MTG has been bilaterally observed in pitch perturbation MEG and fMRI studies (Kort et al., 2016; Parkinson et al., 2012). In this study, suppressed activity in MTG is observed

in the left hemisphere throughout all the time windows (theta and alpha bands) and in the right hemisphere for 300ms after perturbation offset (beta band). Meanwhile, enhanced activity of the MTG is sustained for almost half a second after perturbation onset in right hemisphere (theta, alpha, and beta bands) and bilaterally for few hundred milliseconds as well after perturbation onset though more sustained in the left hemisphere (high gamma band). The early enhanced activity in the MTG throughout the frequency bands, which is in line with what's observed in a previous MEG pitch perturbation study (Kort et al., 2016), indicate its involvement in the early error detection process. Furthermore, it can also be observed that its right-hemispheric activity in the lower frequency bands, especially the theta band, is more dominant for pitch, as this enhanced activity is not observed in the directionally overlapped map of activation. This right-lateralization for pitch processing is further observed in the contrast map, where pitch shows more enhanced activity in the right hemisphere during the perturbation window (alpha band) and post perturbation offset (low and high gamma bands), and formant shows more enhanced activity in the left hemisphere during late post perturbation offset time windows (alpha and high gamma).

Activity in the MFG has been bilaterally observed in both pitch and formant MEG and fMRI studies (Kort et al., 2016; Niziolek & Guenther, 2013; Ranasinghe et al., 2019). In this study, its early activity is observed in the left hemisphere as suppression in the theta band and as enhancement in beta and high gamma bands. In the right hemisphere, its suppression is observed throughout all time windows in the alpha band. The early enhanced activity compared to normal speaking after perturbation onset in the high gamma is in line with previous studies of pitch perturbations in MEG, though it was found in either bilaterally (Ranasinghe et al., 2019) or in the right hemisphere, not just in the left hemisphere. When we observed the feature-specific map of

high gamma activity more closely, we can see that enhanced right MFG is found in response to formant perturbations in the early time windows, similar to what was found in formant perturbation study with fMRI, though yet again we did not observe enhanced activity in the right MFG is found in response to pitch perturbations. This can be further confirmed from the contrast map, where early and sustained enhanced activity in the right MFG for formant processing compared to pitch can be observed in both low gamma and high gamma bands. Overall, the right MFG seems to be involved in both pitch and formant processing, though according to our findings it has a preference for formant processing.

Another frontal region observed in other studies as well as this study is the IFG (Kort et al., 2016; Niziolek & Guenther, 2013; Tourville et al., 2008). In this study, we observed suppression of IFG in the right hemisphere throughout all time windows in the theta band and throughout the perturbation time window in the alpha band. We also observed enhanced activity briefly early after perturbation onset in the right hemisphere (low gamma band) and more sustainedly after perturbation onset and offset in the left hemisphere (high gamma band). Enhanced activity in the IFG has been observed in fMRI formant perturbations studies bilaterally and in pitch perturbation in the right hemisphere (high gamma band). By examining the contrast map, we can observe that the right IFG has more sustained enhanced activity after perturbation onset for formant processing in the low gamma and high gamma bands but more for pitch processing in the alpha and beta bands. Therefore, the right IFG seems to be involved in both pitch and formant processing across different frequency bands, though in the high gamma it is likely to have a preference for formant processing.

The somatosensory cortex is, as expected, involved in both pitch and formant feedback processing as observed in this study, though it has only been previously observed in the left hemisphere in a pitch perturbation study (high gamma band) (Kort et al., 2016). In this study, we observed enhanced activity bilaterally (right theta, bilateral alpha, left beta bands) for both pitch and formant processing, and from the contrast map it seems that the right somatosensory cortex is much more enhanced for pitch processing in the lower frequency bands (theta, alpha, beta and low gamma) but more enhanced activity for formant processing in the high gamma band.

The motor cortex is also as expected involved in both pitch and formant feedback processing, as previously found in other studies as well (Kort et al., 2016; Tourville et al., 2008). In this study, the motor cortex is surprisingly enhanced quite early after perturbation onset bilaterally throughout all frequency bands (right theta, bilateral alpha, left low gamma, and bilateral high gamma), with decreased activation in some frequency bands (left beta, left low gamma, bilateral high gamma) post perturbation offset. Previous studies have shown motor cortex's involvement bilaterally in the high gamma for pitch and in the right hemisphere for formants. From the high gamma band in the contrast map, we further observed that the right motor cortex is more involved for formant processing.

The IPL, though previously only been shown to be involved in the processing of pitch (Kort et al., 2016), is observed in this study to be involved in both pitch and formant processing. Sustained enhanced activity compared to normal speaking is observed bilaterally in alpha and beta band, with peaks of enhanced activity taking place later in the alpha band than in the beta band. Activity in the IPL including the supramarginal gyrus has been previously observed early

post perturbation onset in the high gamma band for pitch processing, but we did not observe such activity in this study. Rather, larger enhanced activity in the left IPL is observed for formants compared to pitch processing in the high gamma band, whereas larger decreased activity in bilateral IPL is observed for pitch compared to formants in the beta band.

We observed a sustained decreased activation in the cerebellum across the lower frequency bands (bilateral theta, left alpha, bilateral beta and left low gamma), as well as a brief enhanced activity bilaterally in the high gamma band. In previous studies, activity in the cerebellum has only been observed in the right hemisphere through fMRI formant perturbation studies (Tourville et al., 2008). What is observed in this study, on the other hand, is larger enhanced activity in the left cerebellum soon after perturbation onset in the theta, low gamma and high gamma bands for formants compared to pitch, with more enhanced activity in the right cerebellum for pitch compared to formants in much later time windows post perturbation offset (beta, low gamma and high gamma bands).

We also observed several other regions that are involved in both pitch and formant processing that has not been observed in previous studies. In the frontal lobe, the SFG is involved in both pitch and formant processing: we observed decreased activity in the left hemisphere (alpha band) that is sustained throughout and in the right hemisphere (theta band) early post perturbation onset, as well as enhanced activity in the right hemisphere (beta and low gamma bands) transiently after perturbation onset.

In the temporal lobe, we observed the involvement of the left ITG and fusiform gyrus for both pitch and formant processing. Specifically, we observed sustained decreased activity in the left ITG (theta and alpha bands) and fusiform gyrus (theta band) throughout all time windows, similar to that observed in the left MTG. The right fusiform gyrus also followed the decreased activity pattern of the right MTG beta band after perturbation offset. Furthermore, interestingly, from the contrast map we observed that the right ITG in the late perturbation window (alpha band) and late after perturbation offset (theta band) is more enhanced in pitch processing than in formants.

In the parietal lobe, we observed involvement of SPL. Enhanced activity in the right SPL (alpha band) is observed soon after perturbation onset and sustained until perturbation offset, and, specifically in the precuneus, decreased activity in the right SPL (low and high gamma bands) after perturbation offset. From the contrast activation, the right SPL seems to be more enhanced for pitch processing after perturbation offset (beta and low gamma bands) while SPL seems to be more enhanced for formant processing in the high gamma band during the perturbation window bilaterally and after perturbation offset in the left hemisphere.

In the occipital lobe, we observed the involvement of the cuneus, MOG, IOG and lingual gyrus for both pitch and formant processing. A sustained decreased activation in the left IOG, left cuneus and right lingual gyrus is observed throughout all time windows in the theta band. The left MOG showed decreased activation transiently after perturbation onset (low gamma band) and after perturbation offset (beta band) as well as enhanced activation after perturbation offset (high gamma band). The right cuneus showed decreased activation transiently after perturbation

onset (beta band) and sustained after perturbation offset (low gamma band). From the contrast activation, we see more enhanced involvement for pitch processing in the left MOG (low gamma band, transiently during perturbation offset) left cuneus (beta band, all time windows), left lingual gyrus (alpha band, all time windows), and more enhanced involvement for formant processing in after perturbation offset in MOG (high gamma band) and cuneus (alpha band).

3.5.2 Correlations between neural activation and peak compensation responses to pitch perturbations

Previous MEG studies have investigated correlations between neural activation in the high gamma band and peak compensation responses to pitch perturbations (Kort et al., 2016; Ranasinghe et al., 2019), and several regions observed in those studies include the bilateral prefrontal cortex, bilateral MTG and left MOG. We observed correlations in some of these regions across multiple frequency bands, as well as other regions not previously observed.

In the prefrontal cortex high gamma band, positive correlation has been observed in the right hemisphere by one study (Kort et al., 2016) and negative correlation in the left hemisphere by another study (Ranasinghe et al., 2019), both 200ms after pitch perturbation onset. In this study, we instead observed the opposite with an earlier timescale in the high gamma band where the left prefrontal cortex shows positive correlation and right prefrontal cortex shows negative correlation to peak pitch compensation responses right after perturbation onset. The left prefrontal cortex (high gamma band) also showed positive correlation again in the later perturbation window. Across other frequency bands, we observed negative correlation with the

bilateral prefrontal cortex, mostly in the late perturbation window extending until hundred milliseconds after perturbation offset.

In the MTG high gamma band, positive correlation has been observed posteriorly in the right hemisphere by one study (Ranasinghe et al., 2019) and negative correlation has been observed anteriorly in the left hemisphere by another study (Kort et al., 2016), both 200ms after pitch perturbation onset. Our findings are yet again not in line with those, as not much correlation is observed in the MTG in the high gamma band, though we observed positive correlation in the nearby STG region more anteriorly (compared to Ranasinghe et al., 2019) in the right hemisphere 250ms after perturbation onset. In the left hemisphere, mostly negative correlations were observed in the temporal lobe across multiple frequency bands (alpha, beta, low gamma, high gamma. In the right hemisphere, however, many positive correlations are observed in the temporal lobe, right after perturbation onset (alpha band, posteriorly and sustained throughout all time windows) and some in the later perturbation window (high gamma) or after perturbation offset (theta and beta bands).

In the MOG, positive correlation to pitch compensation responses was observed in the left hemisphere. In this study we did not observe this positive correlation in the high gamma band, but instead in the low gamma band (left MOG) and alpha band (right MOG), both right after perturbation onset and sustaining for a few hundred milliseconds. We also observed a positive correlation in a nearby region, left cuneus, that is sustained throughout all time windows in the theta band.

In our study, we observed several new regions showing correlations to peak pitch compensation responses. Among those are the somatosensory and motor cortex, supramarginal gyrus, IPL, and cerebellum. Positive correlation is observed transiently after perturbation onset in the right somatosensory (alpha band) and right motor (theta band) cortices, while negative correlations in right SMC and left MC (at a later time scale) were observed in the high gamma band. The IPL shows positive correlations in the alpha and high gamma bands, specifically transiently right after perturbation onset in the left hemisphere high gamma band and more sustainedly in the alpha band and transiently beta band in the right hemisphere. The supramarginal gyrus is negatively correlated in the left hemisphere but positively correlated in the right hemisphere after perturbation offset in the high gamma band, with some transiently negative correlation in the left hemisphere much later in the low gamma band. In the cerebellum, we observed positive correlations soon after perturbation onset bilaterally in the theta band, and negative correlations in the same time scale in the right hemisphere high gamma band, throughout all time windows in the left hemisphere beta band, and after perturbation offset in the left hemisphere across alpha, low gamma and high gamma bands.

3.5.3 Correlations between neural activation and peak compensation responses to formant perturbations

Only one previous fMRI study has investigated the neural-behavioral correlation to formant perturbations (Niziolek & Guenther, 2013), which found positive correlation primarily in the bilateral STG (which covered the entire gyrus through MTG and possibly ITG as well) and IFG with several sparsely located correlations in other regions. In our study, we observed positive correlations in the right STG in the alpha band, right MTG and left ITG in the theta band, and in

the bilateral IFG in the theta ban, all of which are sustained through many time windows. We also observed negative correlations in the left temporal lobe (alpha, beta, low gamma and high gamma bands) and in the left IFG in the low gamma band during the perturbation window. In this study, we also observed correlations in the prefrontal region besides IFG. We observed positive correlations in the right SFG and MFG across all frequency bands that are sustained through multiple time windows, as well as some negative correlations in the left prefrontal cortex transiently after perturbation onset that is followed immediately by positive correlation (beta, low gamma and high gamma bands).

In this study, we saw bilateral positive correlations across all frequency bands (except theta) in the motor cortex, which mostly started soon after perturbation onset and sustained for few time windows. We also saw bilateral positive correlations in the somatosensory cortex in the beta band and negative correlation in the left somatosensory cortex in the high gamma band during the perturbation window. Other regions showing correlations that we observed in this study are IPL and cerebellum. In the IPL, we saw positive correlations in the right theta band and left low gamma band throughout the perturbation, and transient negative correlation in the left high gamma band after perturbation onset. In the cerebellum, we mostly observed negative correlations bilaterally throughout al frequency bands, with left low gamma and high gamma and right alpha and beta band showing early negative correlation after perturbation onset, though we also observed faint positive correlation in the left theta band during the perturbation window.

3.6 Summary

In this study, we revealed regions involved in the feedback processing of both pitch and formants, including those that have not been previously observed. We also revealed right lateralization for pitch feedback processing in the lower frequency bands but right lateralization for formants feedback processing in the higher frequency bands. Lastly, our neurobehavioral correlation findings for pitch seemed to oppose findings from previous studies, specifically in the high gamma band, though we observed many more neurobehaviorally-correlated regions for both pitch and formants feedback processing than previously observed.

3.7 References

Barnes, G. R., Hillebrand, A., Fawcett, I. P., & Singh, K. D. (2004). Realistic spatial sampling for MEG beamformer images. *Human Brain Mapping*, *23*(2), 120–127. https://doi.org/https://doi.org/10.1002/hbm.20047

Bauer, J. J., Hain, T. C., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, *119*(4), 2363–2371. https://doi.org/10.1121/1.2173513

Bearelly, S., & Cheung, S. W. (2017). Sensory Topography of Oral Structures. *JAMA Otolaryngology–Head & Neck Surgery*, *143*(1), 73–80. https://doi.org/10.1001/jamaoto.2016.2772

Burke, B. D. (1969). Reduced auditory feedback and stuttering. *Behaviour Research and Therapy*, *7*(3), 303–308. https://doi.org/https://doi.org/10.1016/0005-7967(69)90011-4

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE*, *7*(7), 1–13. https://doi.org/10.1371/journal.pone.0041830

Casserly, E. D., Rowley, M. C., Marino, F. R., & Pollack, E. A. (2016). Acoustic analysis of speech produced with degradation of acoustic and somatosensory feedback. *Proceedings of Meetings on Acoustics*, *29*(1), 60016. https://doi.org/10.1121/2.0000651

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7). https://doi.org/10.1073/pnas.1216827110

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

Dalal, S. S., Guggisberg, A. G., Edwards, E., Sekihara, K., Findlay, A. M., Canolty, R. T., Berger, M. S., Knight, R. T., Barbaro, N. M., Kirsch, H. E., Nagarajan, S. S. (2008). Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. *NeuroImage*, *40*(4), 1686–1700. https://doi.org/10.1016/j.neuroimage.2008.01.023

Dalal, S. S., Zumer, J. M., Guggisberg, A. G., Trumpis, M., Wong, D. D. E., Sekihara, K., & Nagarajan, S. S. (2011). MEG/EEG Source Reconstruction, Statistical Evaluation, and Visualization with NUTMEG. *Computational Intelligence and Neuroscience*, *2011*, 758973. https://doi.org/10.1155/2011/758973

Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679. https://doi.org/10.1152/jn.00638.2010

Guenther, F. H. (2016). *Neural Control of Speech*. The MIT Press.

Hall, E. L., Robson, S. E., Morris, P. G., & Brookes, M. J. (2014). The relationship between MEG and fMRI. *NeuroImage*, *102*, 80–91. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.11.005

Hari, R., Levänen, S., & Raij, T. (2000). Timing of human cortical functions during cognition: role of MEG. *Trends in Cognitive Sciences*, *4*(12), 455–462.

https://doi.org/https://doi.org/10.1016/S1364-6613(00)01549-7

Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain Research*. https://doi.org/10.1016/j.brainres.2009.04.033

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*, *6*(3), 131–137. https://doi.org/10.1121/1.1931747

Ho, A. M. H., Chung, D. C., To, E. W. H., & Karmakar, M. K. (2004). Total airway obstruction during local anesthesia in a non-sedated patient with a compromised airway. *Canadian Journal of Anesthesia*, *51*(8), 838. https://doi.org/10.1007/BF03018461

Houde, J. F., & Jordan, M. I. (1998). *Sensorimotor adaptation in speech production. Science, 279(5354), 1213-1216..pdf. 1*.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*(October), 1–14. https://doi.org/10.3389/fnhum.2011.00082

Jacks, A., & Haley, K. L. (2015). Auditory Masking Effects on Speech Fluency in Apraxia of Speech and Aphasia: Comparison to Altered Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1670–1686. https://doi.org/10.1044/2015_JSLHR-S-14-0277

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *J. Acoust. Soc. Am.*, *108*(3 Pt 1), 1246–1251. https://doi.org/10.1121/1.1288414

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, *113*(1),

532–543. https://doi.org/10.1121/1.1529670

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308. https://doi.org/10.1177/0023830911417802

Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., Ballard, K. J., Chang, S., Chao, S., Murrray, E. S. H., Scott, T. L., Guenther, F. H. (2020). A Simple 3-Parameter Model for Examining Adaptation in Speech and Voice Production . *Frontiers in Psychology* , Vol. 10, p. 2995. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.02995

Keough, D., Hawco, C. S., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience*, *14*(1), 25. https://doi.org/10.1186/1471-2202-14-25

Kim, K. S., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research : JSLHR*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_jslhr-19-00419

Kort, N. S., Cuesta, P., Houde, J. F., & Nagarajan, S. S. (2016). Bihemispheric network dynamics coordinating vocal feedback control. *Human Brain Mapping*, *37*(4), 1474–1485. https://doi.org/10.1002/hbm.23114

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The*

Journal of the Acoustical Society of America, *89*(2), 859–866. https://doi.org/10.1121/1.1894647

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research*, *187*(4), 613–621. https://doi.org/10.1007/s00221-008-1330-z

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, *110*(6), 2845–2848. https://doi.org/10.1121/1.1417527

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (1999). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America*, *107*(1), 559–564. https://doi.org/10.1121/1.428323

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862–2872. https://doi.org/10.1121/1.2715657

Maas, E., Mailend, M.-L., & Guenther, F. H. (2015). Feedforward and Feedback Control in Apraxia of Speech: Effects of Noise Masking on Vowel Production. *Journal of Speech, Language, and Hearing Research*, *58*(2), 185–200. https://doi.org/10.1044/2014_JSLHR-S-13-0300

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068. https://doi.org/10.1121/1.3278606

Mitsuya, T., MacDonald, E. N., Munhall, K. G., & Purcell, D. W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of*

America, *138*(1), 413–424. https://doi.org/10.1121/1.4923154

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, *11*(10), 1217–1222. https://doi.org/10.1038/nn.2193

Niemi, M., Laaksonen, J. P., Ojala, S., Aaltonen, O., & Happonen, R. P. (2006). Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/. *International Journal of Oral and Maxillofacial Surgery*, *35*(10), 920–923. https://doi.org/10.1016/j.ijom.2006.06.002

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *Journal of Neuroscience*, *33*(29). https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology*, *15*(9), e1007321.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(January 2006), 2288–2297. https://doi.org/10.1121/1.2173514

Ranasinghe, K. G., Kothare, H., Kort, N., Hinkley, L. B., Beagle, A. J., Mizuiri, D., Honma, S.

M., Lee, R., Miller, B. L., Gorno-Tempini, M. L., Vossel, K. A., Houde, J. F., Nagarajan, S. S. (2019). Neural correlates of abnormal auditory feedback processing during speech production in Alzheimer's disease. *Scientific Reports*, *9*(1), 5686. https://doi.org/10.1038/s41598-019-41794-x

Reilly, K. J., & Dougherty, K. E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4812763

Scheerer, N. E., & Jones, J. A. (2018). The Role of Auditory Feedback at Vocalization Onset and Mid-Utterance. *Frontiers in Psychology*, *9*, 2019. https://doi.org/10.3389/fpsyg.2018.02019

Scott, C. M., & Ringel, R. L. (1971). Articulation Without Oral Sensory Control. *Journal of Speech and Hearing Research*, *14*(4), 804–818. https://doi.org/10.1044/jshr.1404.804

Singh, S. P. (2014). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*(Suppl 1), S107–S112. https://doi.org/10.4103/0972-2327.128676

Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of Auditory and Somatosensory Feedback to Vocal Motor Control. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2039–2053. https://doi.org/10.1044/2020_JSLHR-19-00296

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/J.NEUROIMAGE.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869. https://doi.org/10.1038/nature01710

Chapter 4: Effects of oral cavity numbing on speech responses to perturbed auditory feedback

4.1 Abstract

Feedback information, including auditory and somatosensory feedback, is crucial for the control of speech. Previous feedback perturbation studies have shown evidence for the individual role of auditory and somatosensory feedback in speech. However, it has been frequently observed that compensation responses to auditory feedback perturbations never truly reached the full magnitude of the applied perturbations, which has been postulated to be caused by competing somatosensory feedback information. Evidence supporting this idea was from a pitch perturbation study with numbing of the vocal folds, which found that compensation responses increased when the somatosensory feedback information was reduced. This study investigated whether similar effect can be observed in the compensation responses to formant feedback perturbations with numbing of the oral cavity. Formant feedback perturbations in both negative and positive directions were applied for a whole utterance before and after swishing with treatment solution, which was either a placebo non-numbing solution or lidocaine numbing solution. Results showed that oral cavity numbing has an effect on the magnitude of compensation responses to formant perturbations, though said effect depended on the applied perturbation direction. Further investigation is required to explore this direction-dependent interaction between the auditory and somatosensory feedback information for the control of formants.

4.2 Introduction

It is well understood that human's ability to produce clear speech requires precise timing and coordination of many independent articulators. The control of these speech articulators, as it turns out, heavily depends on the available feedback information. More specifically, humans rely on what they hear (auditory feedback) and what they feel (somatosensory feedback) in order to correct for errors in their speech. Numerous studies have shown evidence for significant alterations in speech behaviors when auditory or somatosensory feedback is experimentally masked/removed (Burke, 1969; Jacks & Haley, 2015; Jones & Munhall, 2003; Maas et al., 2015; Niemi et al., 2006; Scott & Ringel, 1971; Smith et al., 2020) or absent in post-lingually deaf population (Lane & Webster, 1991), underlining the role of both auditory and somatosensory feedback information in the control of speech.

Further evidence for the role of feedback in speech is shown through studies with feedback perturbations, which results in a compensation response. Speech production models (Guenther, 2016; Houde & Nagarajan, 2011; Kearney et al., 2020; Parrell et al., 2019)theorized that this ability to compensate for speech errors made possible by a feedback control mechanism, which 1) compares the feedback information to an internal representation of expected feedback and 2) sends necessary motor correction commands to the speech articulators when a mismatch is detected. Numerous studies have shown evidence for compensation responses to auditory feature feedback perturbation of loudness (Bauer et al., 2006; Heinks-Maldonado & Houde, 2005; Larson et al., 2007), pitch (Burnett et al., 1998; Hawco & Jones, 2009; Larson et al., 2001, 1999; Scheerer & Jones, 2018) and formants (Cai et al., 2012; Purcell & Munhall, 2006b; Reilly & Dougherty, 2013), as well as to somatosensory feedback perturbation (Nasir & Ostry, 2008;

Tremblay et al., 2003). However, in the auditory feedback perturbation studies, incomplete compensation responses, i.e., when compensation response magnitude is not equal to 100% of the perturbation magnitude, are often observed, especially to pitch and formant perturbations. A possible explanation for this observation would be the presence of the unperturbed somatosensory feedback information that reports no change in articulations that is competing with the auditory feedback information that reports an unexpected change in articulation.

Evidence of competing information between auditory and somatosensory feedback can be observed from studies that applied auditory perturbations while simultaneously perturbing somatosensory feedback either with external forces using robotic arms (Lametti et al., 2012) or by reducing it with a local anesthesia (Casserly et al., 2016; Larson et al., 2008). When somatosensory feedback is consistently perturbed using external forces while also receiving consistent auditory feedback perturbations, subjects showed preference for either somatosensory or auditory feedback (Lametti et al., 2012). More straightforward evidence for somatosensory feedback competing with auditory feedback information in incomplete compensation responses is found when somatosensory feedback from the vocal folds is reduced using local anesthesia, where subjects demonstrated larger responses to unexpected within-utterance pitch feedback perturbations (Larson et al., 2008). This truly demonstrates that somatosensory feedback was a hindrance to achieving full compensation response to pitch perturbations. However, whether somatosensory feedback would also restrain full compensation response to formant perturbations is still unclear.

The current study investigated the effect of reducing somatosensory feedback information on compensation responses to formant perturbations by applying a non-invasive local anesthesia

and assessing speech responses to formant perturbations during the production of the vowel /ɛ/. We first examined the effectiveness of topical oral lidocaine in reducing the somatosensory sensitivity of the oral cavity using a somatosensory force mapping method outlined by Bearelly and Cheung (2017). We decided to pursue the use of topical oral lidocaine (i.e., lidocaine solution delivered via contact with oral surface area instead of injection) for ease of use as well as to reduce possible effect of lidocaine on the oral motor function, given that 4% lidocaine has been shown to significantly affect fine motor function in the larynx (Ho, Chung, To, & Karmakar, 2004). Though a topical administration of local anesthesia has been previously demonstrated in a pitch feedback perturbation during reduced somatosensory feedback study, said local anesthesia was used to anesthetize the vocal folds, the articulator organ mainly responsible for the production of pitch. Moreover, a different local anesthetic was used in that study. Thus, it was necessary to confirm the effectiveness of topical lidocaine in reducing the somatosensory feedback of the oral cavity, specifically the tongue, which is one of the main articulators responsible for the control of formants.

After we showed the effectiveness of topical lidocaine in reducing the somatosensory feedback of the oral cavity, we then examined speech responses to formant feedback perturbations during reduced somatosensory feedback. The perturbations applied were whole-trial F1 perturbations, which have been shown to induce reliable within-trial compensation responses. We applied these perturbations at varying directions (either increases or decreases in F1), similar to prior formant feedback perturbation studies, to increase the unpredictability of the perturbations and to investigate the effect of perturbation direction. We hypothesized that responses to formant

feedback perturbations would be larger when somatosensory feedback is reduced with lidocaine, similar to what was found in pitch.

4.3 Methods

4.3.1 Somatosensory sensitivity reduction study

4.3.1.1 Participants

To evaluate the effectiveness of lidocaine in reducing the somatosensory sensitivity of the oral cavity, a somatosensory sensitivity reduction study was performed prior to the main feedback perturbation study. Healthy participants were recruited (n=21) through pamphlets and announcements at the UC San Francisco campus and online platform. Participants' ages ranged from 19 to 68 years (mean ± standard deviation of 29.9 ± 11.74 years). Participants had no allergy to lidocaine, no deficits in learning, motor, or speech and language abilities, and gave written informed consent to participate. There were no exclusion criteria based on the participants' native language for this somatosensory numbing study. The study was approved by the UCSF Institutional Review Board for human research.

4.3.1.2 Treatment solution

Participants were assigned to either placebo (n=13) or lidocaine (n=16) treatment group and were made aware of which treatment they were receiving, though some participants (n=7) were assigned to both groups (participated in the different treatments on different days). Participants in the lidocaine treatment group were asked to swish vigorously (but not gargle) with a 10ml lidocaine solution for 1 minute to make sure the solution coats the entire oral cavity. The lidocaine solution consisted of 5ml of 4% topical lidocaine solution (Roxanne Laboratories,

obtained from the UCSF Pharmacy) and 5ml of flavored water (made of 1 drop of Crystal Light Strawberry Lemonade flavoring in 1 cup of water). This specific combination was chosen for several reasons. First, a conservative amount of lidocaine was selected to avoid any potential long-term side effect on the participants. Given that the maximum conservative allowable lidocaine solution is 4mg per 10kg body weight, the maximum conservative allowable lidocaine for the average 80kg adult would be 320mg; this amounts to 8ml of 4% (40mg/ml) topical lidocaine solution. Therefore, the 5ml volume for the 4% topical lidocaine solution was chosen to take a more conservative approach as well as for easier measurement. Second, since the lidocaine solution was administered via swishing, it would be easier if the solution was larger in volume. Therefore, the additional 5ml flavored water was added to increase the chance for the lidocaine solution to come in contact with the entire surface area of the oral cavity. Third, flavored water was chosen to dilute the solution to lessen the bitterness of the lidocaine solution, which hopefully reduce the unpleasantness experienced by the participants in the lidocaine treatment group compared to the control treatment group. Participants in the placebo treatment group were similarly asked to swish with a 10ml placebo solution consisting of 5ml lemon rind juice (to mimic the bitterness of lidocaine) and 5ml flavored water.

4.3.1.3 Oral somatosensory mapping

Participants in both treatment groups underwent an oral somatosensory mapping experiment developed by Bearelly and Cheung (2017) to evaluate the oral somatosensory sensitivity before and after swishing with the treatment solution. The oral somatosensory mapping experiment utilized the Cheung-Bearelly monofilaments to measure the sensory threshold of the posterior tongue (Bearelly & Cheung, 2017). The monofilaments suture size ranged from 9-0 (smallest

caliber) to 2-0 (largest caliber). The somatosensory mapping would start by tapping (defined as buckling of the monofilament by around 30% of original length) the participant's (with their eyes closed) posterior tongue with the largest caliber monofilament (2-0), followed by successively smaller caliber monofilaments until participant could not detect the monofilament (staircase method) (Bearelly & Cheung, 2017). Cross-validation using neighboring-sized monofilaments was then performed to confirm the participant's final somatosensory threshold. For example, if a participant could not detect a 5-0 monofilament after the staircase method, four separate taps would be performed for 6-0, 5-0 and 4-0 monofilaments (in that order, largest to smallest caliber), and if 3 out of 4 affirmative responses were given for the 5-0 monofilament, 5-0 was recorded as their somatosensory threshold.

Participants underwent the oral somatosensory mapping experiment once before swishing with their treatment solution, once right after swishing, and every 5 minutes after until 30 minutes post-swish. The start time of the somatosensory mapping post-swish was recorded (instead of the order) since the somatosensory mapping process took a variable amount of time in each participant, depending on their sensitivity level and clarity of their responses; cross-validations were performed multiple times if participant failed to give 3 out of 4 affirmative responses. However, in the end, the majority of the participants had a somatosensory threshold mapped for the following time windows: pre-swish, 0-5 minutes after swishing, and 10-15 minutes after swishing. Therefore, we chose to focus our analysis on these time windows.

4.3.1.4 Statistical analysis

Participant's somatosensory thresholds, recorded as the size of monofilament cross-validated to be their somatosensory sensitivity threshold, were converted to buckling force of the monofilament, using the equation outlined in Bearelly and Cheung (2017):

$$force\ (grams) = 10^{-0.4765x+1.6765}$$

where x is the suture size. This force represents the lowest amount force that the participant was able to detect, i.e., their somatosensory force threshold. A linear mixed effects (LME) model was run in SAS 9.4 (SAS Institute Inc., Cary, NC) using the proc mixed procedure to evaluate the main effect of treatment solution (placebo vs lidocaine) and time windows (pre-swish vs 0-5 minutes post-swish vs 10-15 post-swish) on the somatosensory force thresholds with participant as a random factor. The LME model was then used for a series of contrast analyses to compare the somatosensory force thresholds between time windows in each treatment group. The contrast analyses performed in each placebo and lidocaine treatment group were: 1) pre-swish vs 0-5 minutes post-swish, 2) pre-swish vs 10-15 minutes post-swish, and 3) 0-5 minutes post-swish vs 10-15 minutes post-swish.

4.3.2 Feedback Perturbations Study

4.3.2.1 Participants

Healthy participants were recruited for the feedback perturbation study (n=26, 10 females) through pamphlets and announcements at the UC San Francisco campus and online platform, separately from the previous somatosensory sensitivity reduction study. Participants' ages ranged from 19 to 68 years (mean ± standard deviation of 30 ± 12.88 years). All participants were native English speakers with no allergy to lidocaine and no deficits in learning, motor or speech and

language abilities. Participants gave written informed consent to participate. The study was approved by the UCSF Institutional Review Board for human research.

4.3.2.2 Apparatus

The feedback perturbation experiments were performed in a quiet room equipped with sound booth. While inside the sound booth, participants sat in front of a laptop (Thinkpad W530, Lenovo Group Limited) while wearing Beyerdynamic DT 770 Pro 50 Ohm headphones and a head-mounted AKG Pro Audio C520 condenser microphone. Participant's speech from the microphone was fed into a Focusrite Scarlett 2i2 USB Recording Audio Interface and processed and recorded using MATLAB (which also displayed the word prompts) paired with a custom real-time digital signal processing program called the Feedback Utility for Speech Production (FUSP) (further details of FUSP can be found in Katseff et al., 2012). FUSP repeatedly analyzed 3ms frames of speech input from the microphone into separate pitch and formant representations that were, at times, altered (depending on the experiment) and used to synthesize the next 3ms of speech output to the participant's headphones. The speech data was recorded at a rate of 11025Hz and this feedback processing, along with hardware delays, introduced an imperceptible ~21ms delay in the auditory feedback, as measured following the methods outlined by Kim, et al. (Kim, Wang, & Max, 2020).

4.3.2.3 Experimental design and procedures

Participants were randomly assigned to either placebo or lidocaine treatment group and performed the feedback perturbation experiment before and after swishing with the treatment solution. The solution and method of swishing was as described above in the somatosensory

numbing study. In addition, their lack of an allergic response to lidocaine was confirmed by testing for allergy reaction with the 4% lidocaine solution (not the diluted lidocaine treatment solution) on their skin prior to the start of the experiment. There were 12 and 14 participants in the placebo and lidocaine treatment group, respectively. The number of participants were not completely balanced for this study because the recruitment of these participants depended on another concurrent formant adaptation study, which is outside the scope of this paper.

Each feedback perturbation experiment consisted of 3 types of trials: +200Hz F1 whole utterance perturbations, -200Hz F1 whole utterance perturbations and unperturbed trials. There were 30 trials for each trial type, randomly distributed across a total of 90 trials. During the experiment, participants were instructed to say the word 'head' (/hɛd/), extending the vowel portion of the utterance for as long as the prompt word was displayed on the screen (approximately 2 seconds). The feedback perturbation experiments were administered in two different sessions: once before swishing with treatment solution (pre-swish), and once right after swishing with treatment solution (post-swish).

4.3.2.4 Data processing and statistical analysis

All acoustic speech data was analyzed using Wave Viewer, a custom-built MATLAB-based speech analysis software (https://github.com/SpeechNeuroscienceLab/Wave-Viewer). In each trial, formants were tracked using linear predictive coding (LPC). The tracking for the first formant was further refined by manual screening, as needed, to exclude bad trials (e.g. trials with no speech response, interruption in speech production/recording, and poor formant tracking) and to occasionally fix the voice onset and offset time markings automatically detected by FUSP. On

average, less than 6% of the trials were excluded from analysis across all subjects. The raw F1 formant track trial data was extracted and separated to the 2 experimental sessions: before and after swishing with treatment solution (pre-swish and post-swish, respectively).

Time-course analyses of the responses to the whole utterance formant perturbations before and after swishing with treatment solution were performed. To obtain an F1 response time-course that could highlight each participant's F1 response changes elicited by the perturbations, we performed three linear normalization steps that eliminated within- and across-trial variance. First, each perturbed trial was normalized by subtracting the participant's F1 unperturbed response trend (average F1 response in the 30 unperturbed trials in the corresponding experimental session, i.e., pre-swish or post-swish, normalized at voice onset such that F1(t=0) = 0Hz). This was done to reduce the within-trial variations in F1 responses of each participant. Second, each perturbed trial was then normalized to the voice onset data, such that F1(t=0) = 0Hz. This was done to reduce across-trials variations in F1 responses. Third, each perturbed trial was smoothed by averaging the F1 responses within non-overlapping 25ms windows. This was done to reduce the formant tracking variations across frames. The average and standard error of the F1 response time-courses were then calculated across participants within each treatment group (placebo vs lidocaine) and experimental session (pre-swish vs post-swish).

Using the normalized F1 response time-course data points, we ran a linear mixed effects model in SAS to evaluate the main effect of treatment group (placebo vs lidocaine), experimental session (pre-swish vs post-swish) and perturbation direction (+200Hz vs -200Hz) on the within-trial compensation responses with participant as a random factor. The LME model was then used

for a series of contrast analyses to compare the F1 responses between sessions in each treatment group and perturbation direction. The contrast analyses performed compared pre-swish vs post-swish responses in: 1) placebo group with +200Hz perturbations, 2) placebo group with -200Hz perturbations, 3) lidocaine group with +200Hz perturbations, and 4) lidocaine group with -200Hz perturbations.

4.4 Results

4.4.1 Participants showed reduced somatosensory sensitivity after administration of lidocaine solution

Participants in the lidocaine group exhibited increased somatosensory force threshold over time (Figure 4.1), whereas participants in the placebo group exhibited similar if not decreased somatosensory force threshold. A linear mixed effects model of the somatosensory force thresholds showed significant main effect of the treatment group (placebo vs lidocaine: $F(83)=5.98$, $p=0.0166$) but not of time windows (pre-swish vs 0-5 minutes post-swish vs 10-15 post-swish: ($F(83)=0.62$, $p=0.5394$) with no significant interaction between treatment group and time window ($F(83)=2.04$, $p=0.1366$). The contrast analysis revealed a significant difference only between the force somatosensory threshold in the pre-swish vs 10-15 minutes post-swish session in the lidocaine group ($F(83)=4.27$, $p=0.0419$); the rest of the contrast analysis was not significant (pre-swish vs 0-5 minutes post-swish in lidocaine group: $F(83)=3.26$, $p=0.0747$; 0-5 minutes vs 10-15 minutes post-swish in lidocaine group: $F(83)=0.08$, $p=0.7723$; pre-swish vs 0-5 minutes post-swish in placebo group: $F(83)=0.09$, $p=0.7624$; pre-swish vs 10-15 minutes post-swish in placebo group: $F(83)=0.05$, $p=0.48$; 0-5 minutes vs 10-15 minutes post-swish in placebo group: $F(83)=0.16$, $p=0.6857$). Overall, participants' somatosensory sensitivity was significantly

reduced 10 minutes after administration of the lidocaine solution, though the sensitivity reduction was already observed even right after swishing.

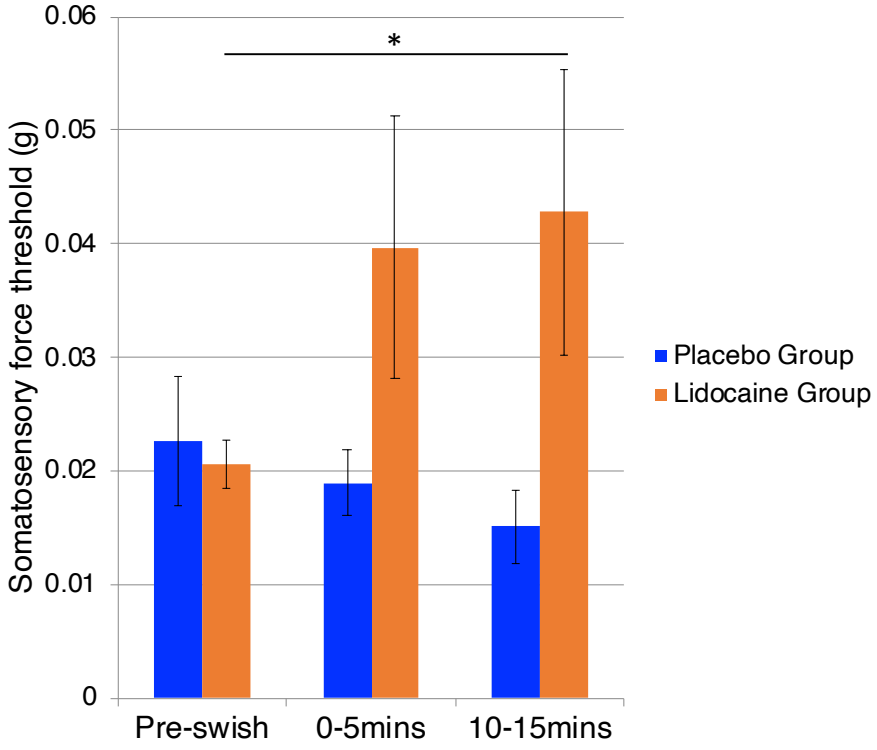

Figure 4.1: Somatosensory force threshold changes before and after swishing with treatment solution. Mean (height of bar) and SEM (bar whiskers) of the somatosensory force thresholds of participant's tongue sensitivity in each time window (pre-swish: before swishing, 0-5 minutes after swishing, and 10-15 minutes after swishing) for participants in the placebo group (blue) and lidocaine group (orange). Only force thresholds between pre-swish and 10-15 minutes after swishing time windows in the lidocaine group were significantly different from each other ($F(83)=4.27$, $p=0.0419$).

4.4.2 Online compensation responses to formant feedback perturbations were affected by swishing solution depending perturbation direction

Participants showed changes in their online compensation responses to unpredictable whole utterance +200Hz F1 perturbations but not to -200Hz F1 perturbations (Figure 4.2). A linear mixed effects model of the normalized timecourse compensation responses showed significant main effects for swishing (before vs after swishing: $F(4128)= 7.85$, $p=0.01$), treatment group

(placebo vs lidocaine, F(4128)= 6.27, p=0.01), and perturbation direction (positive vs negative: F(4128)= 7.56, p=0.01). There are significant interactions between swishing and perturbation direction (F(4128)= 47.03, p<0.0001) and between swishing, treatment group and perturbation direction (F(4128)=135.42, p<0.0001) but not between swishing and treatment group (F(4128)=0.27, p=0.60) and between treatment group and perturbation direction (F(4128)=0.07, p=0.79). A post-hoc analysis of the model reveals a significant effect of swishing in the lidocaine treatment group responses to both perturbation directions (-200Hz: F(4128)=71.22, p<0.0001, +200Hz: F(4128)=116.88, p<0.0001) and in the placebo treatment group responses to -200Hz perturbation (F(4128)=15.25, p<0.0001) but not to +200Hz perturbation (F(4128)=0.49, p=0.48). Overall, swishing with either placebo or lidocaine solution significantly affected online compensation responses -200Hz F1 perturbations, but only lidocaine solution significantly affected the responses to +200Hz F1 perturbations.

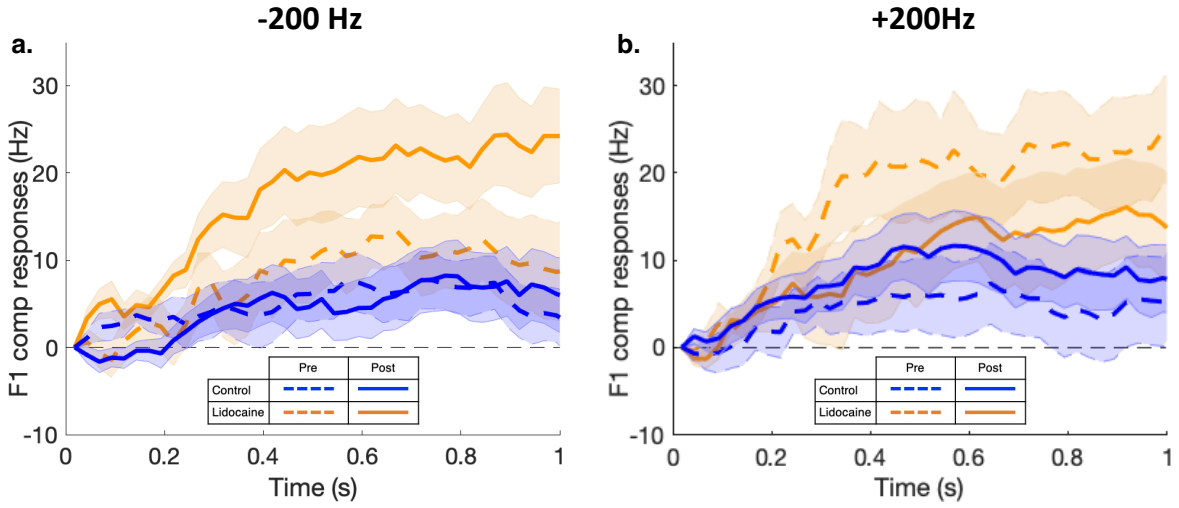

Figure 4.2: Online formant response to whole-utterance formant perturbations before and after swishing with solution. Normalized F1 responses to (**a**) -200Hz and (**b**) +200Hz F1 whole-utterance perturbations averaged across participants within each treatment group, placebo (blue) and lidocaine (orange), before (pre, dashed) and after (post, solid) participants swished with treatment solution. Mean responses (lines) and SEM (shaded colored region) are shown.

4.5 Discussion

In this study, we showed that the reduction of somatosensory feedback information on the oral cavity using lidocaine has a significant effect on the within-trial compensation responses to formant perturbations. Specifically, compensation responses to -200Hz F1 perturbations were significantly larger after swishing with the numbing lidocaine solution. This is in line with what was found in compensation responses to positive pitch perturbations (Larson et al., 2008). However, responses to +200Hz F1 perturbations were affected by swishing alone, with compensation responses significantly increased after swishing with placebo solution and significantly decreased after swishing with numbing lidocaine solution.

Though this was initially unexpected, this result actually reflects the significance of perturbation direction in responses to feedback perturbations. Direction dependent responses to perturbations has been previously observed in a sensorimotor adaptation study (Kothare, et al. in press), and one possible explanation proposed for such direction dependence is based on the need to balance competing auditory and somatosensory feedback errors in speech production. The idea that somatosensory feedback information is competing with auditory feedback information more on some vowels than others has also been proposed by Mitsuya (Mitsuya, MacDonald, Munhall, & Purcell, 2015), where a smaller adaptation response to consistently altered auditory feedback using the word ‘hid’ compared to the word ‘head’ was observed. This was postulated to be driven by the richer somatosensory feedback information in the production of word ‘hid’ compared to ‘head’.

In terms of what articulatory changes are necessary to compensate for the +200Hz F1 perturbations in this study, participants would need to decrease their produced F1, which involves raising their tongue position closer towards the palatal region, moving closer to the production of the word 'hid'. This would result in an increased in somatosensory information from the oral cavity, which would compete with the altered auditory feedback information. Once somatosensory feedback is reduced or even removed, there is no more feedback information competition and thus the auditory feedback is able to dominate and drive the compensation response, increasing the compensation response to altered auditory feedback even more. On the other hand, to compensate for the -200Hz F1 perturbations, participants would need to increase their F1 production, which involves lowering their tongue away from the palatal region, closer to the production of the word 'had'. This movement would not induce much somatosensory feedback information from the oral cavity, which means auditory feedback information was not competing much with the somatosensory feedback information to begin with. Thus, it makes sense that no increase in compensation responses were observed.

What is puzzling, however, is the fact that compensation responses to -200Hz F1 perturbations were actually affected in both placebo and lidocaine group. For the placebo group, the compensation responses significantly increased after swishing with the placebo solution. However, the increase seems visually minor, and it is possibly driven by participants heighten awareness of their feedback information overall, given that there would not be as much competing somatosensory information for the movements involved for compensation responses to -200Hz F1 perturbations (compared to +200Hz) to begin with. For the lidocaine group, the compensation responses significantly decreased after swishing with the lidocaine solution. In

terms of direction, this response is actually in the same direction as the compensation response to +200Hz F1 perturbation, which is a decrease in F1 by raising of the tongue. We can only speculate that the numbing sensation induced by lidocaine may have led to a desire to 'look for' as much somatosensory feedback information as possible. Furthermore, a previous study has shown variability in somatosensory vs auditory feedback information dominance among participants (Feng, Gracco, & Max, 2011), which might have impacted the result we observed here. Further research is required to investigate this unexpected change in response to -200Hz F1 perturbations after swishing with lidocaine.

One major limitation of our study is the possible variability in the effect of lidocaine in reducing the somatosensory information of our participants. Though our initial numbing study shows that lidocaine can significantly reduce somatosensory sensitivity, the effect did need to take some time to obtain significant numbing effect. The measurement for the somatosensory sensitivity was only on a posterior location of the tongue due to time constraint, and this area was shown to be less sensitive than other regions (Bearelly & Cheung, 2017). Moreover, we were unable to obtain somatosensory sensitivity information in the set of subjects who performed the auditory feedback perturbations study due to time constraint. It is entirely possible that these groups of participants had widely varying degrees of numbing when performing their speech tasks. Anecdotally, participants did express varying degree of numb sensation, and some participant's normal speech production (observed during break) was degraded more than others. It is indeed difficult to equalize the numbing degree across all participants given many factors that can drive the effect of lidocaine in reducing the somatosensory sensitivity.

4.6 References

Barnes, G. R., Hillebrand, A., Fawcett, I. P., & Singh, K. D. (2004). Realistic spatial sampling for MEG beamformer images. *Human Brain Mapping*, *23*(2), 120–127. https://doi.org/https://doi.org/10.1002/hbm.20047

Bauer, J. J., Hain, T. C., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, *119*(4), 2363–2371. https://doi.org/10.1121/1.2173513

Bearelly, S., & Cheung, S. W. (2017). Sensory Topography of Oral Structures. *JAMA Otolaryngology–Head & Neck Surgery*, *143*(1), 73–80. https://doi.org/10.1001/jamaoto.2016.2772

Burke, B. D. (1969). Reduced auditory feedback and stuttering. *Behaviour Research and Therapy*, *7*(3), 303–308. https://doi.org/https://doi.org/10.1016/0005-7967(69)90011-4

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & Perkell, J. S. (2012). Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation. *PLoS ONE*, *7*(7), 1–13. https://doi.org/10.1371/journal.pone.0041830

Casserly, E. D., Rowley, M. C., Marino, F. R., & Pollack, E. A. (2016). Acoustic analysis of speech produced with degradation of acoustic and somatosensory feedback. *Proceedings of Meetings on Acoustics*, *29*(1), 60016. https://doi.org/10.1121/2.0000651

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7). https://doi.org/10.1073/pnas.1216827110

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America*, *121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

Dalal, S. S., Guggisberg, A. G., Edwards, E., Sekihara, K., Findlay, A. M., Canolty, R. T., Berger, M. S., Knight, R. T., Barbaro, N. M., Kirsch, H. E., Nagarajan, S. S. (2008). Five-dimensional neuroimaging: localization of the time-frequency dynamics of cortical activity. *NeuroImage*, *40*(4), 1686–1700. https://doi.org/10.1016/j.neuroimage.2008.01.023

Dalal, S. S., Zumer, J. M., Guggisberg, A. G., Trumpis, M., Wong, D. D. E., Sekihara, K., & Nagarajan, S. S. (2011). MEG/EEG Source Reconstruction, Statistical Evaluation, and Visualization with NUTMEG. *Computational Intelligence and Neuroscience*, *2011*, 758973. https://doi.org/10.1155/2011/758973

Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of Neurophysiology*, *106*(2), 667–679. https://doi.org/10.1152/jn.00638.2010

Guenther, F. H. (2016). *Neural Control of Speech*. The MIT Press.

Hall, E. L., Robson, S. E., Morris, P. G., & Brookes, M. J. (2014). The relationship between MEG and fMRI. *NeuroImage*, *102*, 80–91. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.11.005

Hari, R., Levänen, S., & Raij, T. (2000). Timing of human cortical functions during cognition: role of MEG. *Trends in Cognitive Sciences*, *4*(12), 455–462.

https://doi.org/https://doi.org/10.1016/S1364-6613(00)01549-7

Hawco, C. S., & Jones, J. A. (2009). Control of vocalization at utterance onset and mid-utterance: Different mechanisms for different goals. *Brain Research*. https://doi.org/10.1016/j.brainres.2009.04.033

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*, *6*(3), 131–137. https://doi.org/10.1121/1.1931747

Ho, A. M. H., Chung, D. C., To, E. W. H., & Karmakar, M. K. (2004). Total airway obstruction during local anesthesia in a non-sedated patient with a compromised airway. *Canadian Journal of Anesthesia*, *51*(8), 838. https://doi.org/10.1007/BF03018461

Houde, J. F., & Jordan, M. I. (1998). *Sensorimotor adaptation in speech production. Science, 279(5354), 1213-1216..pdf. 1*.

Houde, J. F., & Nagarajan, S. S. (2011). Speech Production as State Feedback Control. *Frontiers in Human Neuroscience*, *5*(October), 1–14. https://doi.org/10.3389/fnhum.2011.00082

Jacks, A., & Haley, K. L. (2015). Auditory Masking Effects on Speech Fluency in Apraxia of Speech and Aphasia: Comparison to Altered Auditory Feedback. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1670–1686. https://doi.org/10.1044/2015_JSLHR-S-14-0277

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *J. Acoust. Soc. Am.*, *108*(3 Pt 1), 1246–1251. https://doi.org/10.1121/1.1288414

Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: The role of auditory feedback. *The Journal of the Acoustical Society of America*, *113*(1),

532–543. https://doi.org/10.1121/1.1529670

Katseff, S., Houde, J., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308. https://doi.org/10.1177/0023830911417802

Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., … Guenther, F. H. (2020). A Simple 3-Parameter Model for Examining Adaptation in Speech and Voice Production . *Frontiers in Psychology* , Vol. 10, p. 2995. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2019.02995

Keough, D., Hawco, C. S., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience*, *14*(1), 25. https://doi.org/10.1186/1471-2202-14-25

Kim, K. S., Wang, H., & Max, L. (2020). It's About Time: Minimizing Hardware and Software Latencies in Speech Research With Real-Time Auditory Feedback. *Journal of Speech, Language, and Hearing Research : JSLHR*, *63*(8), 2522–2534. https://doi.org/10.1044/2020_jslhr-19-00419

Kort, N. S., Cuesta, P., Houde, J. F., & Nagarajan, S. S. (2016). Bihemispheric network dynamics coordinating vocal feedback control. *Human Brain Mapping*, *37*(4), 1474–1485. https://doi.org/10.1002/hbm.23114

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory Preference in Speech Production Revealed by Simultaneous Alteration of Auditory and Somatosensory Feedback. *Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *The Journal of the Acoustical Society of America*, *89*(2), 859–866.

https://doi.org/10.1121/1.1894647

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Experimental Brain Research*, *187*(4), 613–621. https://doi.org/10.1007/s00221-008-1330-z

Larson, C. R., Burnett, T. A., Bauer, J. J., Kiran, S., & Hain, T. C. (2001). Comparison of voice F0 responses to pitch-shift onset and offset conditions. *The Journal of the Acoustical Society of America*, *110*(6), 2845–2848. https://doi.org/10.1121/1.1417527

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (1999). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America*, *107*(1), 559–564. https://doi.org/10.1121/1.428323

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862–2872. https://doi.org/10.1121/1.2715657

Maas, E., Mailend, M.-L., & Guenther, F. H. (2015). Feedforward and Feedback Control in Apraxia of Speech: Effects of Noise Masking on Vowel Production. *Journal of Speech, Language, and Hearing Research*, *58*(2), 185–200. https://doi.org/10.1044/2014_JSLHR-S-13-0300

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *The Journal of the Acoustical Society of America*, *127*(2), 1059–1068. https://doi.org/10.1121/1.3278606

Mitsuya, T., MacDonald, E. N., Munhall, K. G., & Purcell, D. W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of America*, *138*(1), 413–424. https://doi.org/10.1121/1.4923154

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, *11*(10), 1217–1222. https://doi.org/10.1038/nn.2193

Niemi, M., Laaksonen, J. P., Ojala, S., Aaltonen, O., & Happonen, R. P. (2006). Effects of transitory lingual nerve impairment on speech: an acoustic study of sibilant sound /s/. *International Journal of Oral and Maxillofacial Surgery*, *35*(10), 920–923. https://doi.org/10.1016/j.ijom.2006.06.002

Niziolek, C. A., & Guenther, F. H. (2013). Vowel Category Boundaries Enhance Cortical and Behavioral Responses to Speech Feedback Alterations. *Journal of Neuroscience*, *33*(29). https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314–322. https://doi.org/10.1016/j.neuroimage.2012.02.068

Parrell, B., Ramanarayanan, V., Nagarajan, S., & Houde, J. (2019). The FACTS model of speech motor control: Fusing state estimation and task-based control. *PLOS Computational Biology*, *15*(9), e1007321.

Purcell, D. W., & Munhall, K. G. (2006a). Adaptive control of vowel formant frequency: evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Purcell, D. W., & Munhall, K. G. (2006b). Compensation following real-time manipulation of formants in isolated vowels. *The Journal of the Acoustical Society of America*, *119*(January 2006), 2288–2297. https://doi.org/10.1121/1.2173514

Ranasinghe, K. G., Kothare, H., Kort, N., Hinkley, L. B., Beagle, A. J., Mizuiri, D., … Nagarajan, S. S. (2019). Neural correlates of abnormal auditory feedback processing during

speech production in Alzheimer's disease. *Scientific Reports*, *9*(1), 5686. https://doi.org/10.1038/s41598-019-41794-x

Reilly, K. J., & Dougherty, K. E. (2013). The role of vowel perceptual cues in compensatory responses to perturbations of speech auditory feedback. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.4812763

Scheerer, N. E., & Jones, J. A. (2018). The Role of Auditory Feedback at Vocalization Onset and Mid-Utterance. *Frontiers in Psychology*, *9*, 2019. https://doi.org/10.3389/fpsyg.2018.02019

Scott, C. M., & Ringel, R. L. (1971). Articulation Without Oral Sensory Control. *Journal of Speech and Hearing Research*, *14*(4), 804–818. https://doi.org/10.1044/jshr.1404.804

Singh, S. P. (2014). Magnetoencephalography: Basic principles. *Annals of Indian Academy of Neurology*, *17*(Suppl 1), S107–S112. https://doi.org/10.4103/0972-2327.128676

Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of Auditory and Somatosensory Feedback to Vocal Motor Control. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2039–2053. https://doi.org/10.1044/2020_JSLHR-19-00296

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429–1443. https://doi.org/10.1016/J.NEUROIMAGE.2007.09.054

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869. https://doi.org/10.1038/nature01710

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
539B35C660BF451...
Author Signature

12/16/2020
Date