Title
Articulation and Altered Auditory Feedback

Author
Bakst, Sarah G.

Publication Date
2017

Articulation and Altered Auditory Feedback

by

Sarah G. Bakst

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Keith A. Johnson, Co-chair
Professor Susan S. Lin, Co-chair
Professor John F. Houde
Professor Richard Ivry

Summer 2017

Articulation and Altered Auditory Feedback

Abstract

Articulation and Altered Auditory Feedback
by
Sarah G. Bakst
Doctor of Philosophy in Linguistics
University of California, Berkeley
Professor Keith A. Johnson, Co-chair
Professor Susan S. Lin, Co-chair

Information from auditory feedback plays an important role in speech motor control. This has been shown in particular in experiments that make real-time changes to speakers' own productions, which are fed back to them while speaking (e.g. Houde and Jordan 1998). In response, speakers change their articulations so that their own altered feedback sounds more like the target word. While such experiments have succeeded in showing the importance of the role of auditory feedback, they have focused on the *acoustic* output of the speech system, the fact that speakers did change their production in response to altered auditory feedback. The experiments presented here focus on *how* speakers change their motor plans using the novel addition of ultrasound imaging. Well-documented cases of articulatory variation in American English provided the test case to study this question. The articulatory strategies which speakers use to compensate for the altered feedback are evidence of how much knowledge they have about how their articulations map onto acoustics. These experiments consider how individual vocal tract anatomy might affect this knowledge.

Chapter 2 uses an articulatory synthesizer with three different palates to show that there is a more sensitive map between articulation and acoustics for flatter than more domed palates. This provided the motivation for Chapter 3, which tested whether these differences influenced variability in the production of normal, unperturbed speech. This ultrasound experiment found that people with flatter palates reduce their articulatory variability in comparison with people who have more domed palates. Reducing articulatory variability effectively constrains acoustic variability in /r/ production but not /s/ production.

Chapter 3 found significant differences in articulatory variability that correlated strongly with palate shape. Chapter 4 presents an experiment that uses ultrasound to capture the strategies that speakers use during adaptation to altered auditory feedback to test whether these production differences determined by palate shape carry over into how speakers make corrections to speech production on line. In one block, speakers said the target word “head,” and their F1 was gradually raised. In a second block, speakers said the target word “hood,” and their F2 was gradually raised. The ultrasound revealed how speakers search their

articulatory-acoustic space and settle on a compensation strategy. A global effect of palate shape was not seen in compensation; degree of compensation did not correlate with palate shape. If palate shape does inform articulatory habits or the detail in a speaker's knowledge of how their articulation maps onto acoustics, this knowledge does not influence speakers' compensation degree. Despite a lack of overall trend in amount of compensation, palate shape did correlate with secondary characteristics in compensation.

Finally, Chapter 5 describes a preliminary study where speakers said the target word "heard," and their F3 was gradually raised. The modeling in Chapter 2 predicted that a generally effective way to lower F3 is to raise and retract the tongue to minimize constriction size, a strategy that every speaker used.

The results from this dissertation suggest that while palate shape influences articulatory and acoustic variability in normal, unaltered speech, this behavior is not predictive of how *much* speakers will compensate when presented with altered feedback, only of certain secondary characteristics of this compensation. Speakers' overall compensation is likely impacted more by higher-level factors which may influence sensitivity to auditory and somatosensory feedback.

Contents

List of Figures

List of Tables

Acknowledgments

I truly enjoyed my Ph.D., for all its ups and downs and many moments of satisfaction. For that, and for having come out the other end of it all in one piece, I have many people to thank.

I had an impossibly kind and supportive committee who helped me navigate the last five years. Keith Johnson has been both an academic advisor and self-described cheerleader. After meetings with Keith, all things were magically improved! If I get to be an advisor one day, I will consider it a success if I can be even half as caring, encouraging, and patient as Keith.

Susan Lin made me learn Python, which was one of the absolutely most useful assignments I have ever had. Thanks also for teaching me how to use ultrasound (and for trusting me with it) and for always giving detailed and thoughtful comments on my work, and most importantly for clearly caring about my personal well-being above all else.

Many thanks to John Houde and Sri Nagarajan for welcoming me to UCSF (and to John for advice on a most excellent head-clearing bike route to get there). I also learned from watching John that it is possible to solve experiment problems by talking out every step with patience and clarity, and that my default method of sinking into self-doubt is actually not necessary or helpful.

In addition to the (figurative) brain stimulation at UCSF's Speech Neuroscience Lab, I met a wonderful group of respectful, collaboratively-minded, and kind colleagues. Thank you for patiently helping me get a handle on brain literature, and ultimately helping me find my future directions! Thanks especially to Megan Thompson, Inez Raharjo, Hardik Kothare, and Zarinah Agnew for being particularly helpful.

This dissertation (among other projects from the last few years) might literally have been impossible without Wizard Ron Sprouse. In addition to fixing every computer or equipment problem, he also was OpenSesame/ultrasound paramedic on multiple occasions (including one false alarm). I aspire to Ron levels of not losing one's cool. A very special extra thanks for getting me set up for a successful jaunt to Chennai!

Thanks to Rich Ivry for input, especially in the planning stages of this dissertation. I am also grateful to Rajesh Kumar at IIT-Madras for helping me get to IIT to do research there, and for helping me jump through all the bureaucratic hoops to make sure that I was able to get the research done.

Thanks to Sharon Inkelas for encouraging my interest in rhotics very early on, which clearly influenced this dissertation and some earlier projects, and for making the Berkeley Linguistics Department a supportive place to be.

Many thanks to the department for providing funding to go to conferences and for encouraging research with undergraduates, which not only increased research efficiency but also provided the very rewarding opportunity to mentor many wonderful undergraduates: volunteer Akshay Aitha, and through the Linguistics Research Apprenticeship Program (LRAP), Evan Gunter, Shannon Foster, Garret Kurteff, Emily van Loo, and Vivienne Finche. Vivienne deserves special recognition for measuring a total of 78 palates in addition to much Praat TextGrid correcting, which combined is a massive amount of mind-numbing work.

The reason that this PhD went as smoothly as it did was because we have Belén Flores and Paula Floro. Thank you Belén for making sure I didn't miss a deadline and for the homemade ice cream. Thank you Paula for making sure I didn't accidentally spend all my money in one place, and thank you both for your warm support.

As countless dissertations before this one have noted, a Ph.D. can be a lonely place without friends. I was lucky to have a desk early on in the Phon Lab (thanks, Keith, for making that happen!) with a bunch of accomplished, grounded, and friendly Lab Ladies (Emily Cibelli, Clara Cohen, Andréa Davis, Shinae Kang, honorarily Stephanie Shih) and some dudes. Thank you for being an enduring support system. Matt Faytak is a major reason that we have a collaborative, non-competitive atmosphere in the lab. (He also makes a yummy tart, so maybe he should've been an honorary Lab Lady as well.) Thanks to John Sylak-Glassman for being so encouraging when I was a insecure prospective student and continuing to be when I was an insecure first-year. My time at Berkeley would have been a lot less fun overall without Nico Baier, Erin Donnelly, Alice Shen, Meg Cychosz, Emily Remirez, Julia Nee, and Karee Garvin.

The reason I'm still in one piece now is Nate. Thanks for having pie delivered to my office, for adding sparkle to the mundane, for the macarons, for sneaking me off to waterfalls and far-flung swimming holes when I clearly needed a break but otherwise wouldn't give myself one, and for paying the breakfast tax. You made this final sprint to the finish line so much more enjoyable.

I count myself incredibly lucky to be the daughter of two parents who have always supported the decisions I've made and who have offered sound guidance in making the tougher ones, who never even thought to ask what I "would do with a Ph.D. in Linguistics" (the answer is obvious: I'll frame it and put it on my wall), who multiply the happiness that comes with every success and who share the weight of the disappointments. I'm also lucky to have a brother who (bonus!) helped me get participants for studies on multiple occasions, and who always makes me laugh.

Chapter 1

Introduction

1.1 Introduction

This dissertation investigates factors in speech motor control. The experiment adds lingual ultrasound imaging to the altered auditory feedback experiment paradigm, during which speakers hear themselves saying something different from what they actually produced. Their speech is altered and resynthesized in real time and sent back to them over headphones while talking. Typically, subjects will compensate in the direction opposite of a perturbation so that the altered version of themselves matches or is closer to the speech target. Such research indicates that speakers are constantly monitoring their output during speech production, and they use this auditory feedback information to adjust their motor plans.

What is less known is how speakers adjust their articulators in order to compensate, and what types of factors affect compensation. Here, two factors are investigated. The first is individual differences in vocal tract anatomy, as measured by the doming of the hard palate, and the second is the role of native language. There is also variation in whether and how much a subject will compensate, though there is less work on factors affecting a subject's compensation. The dissertation considers how these different factors interact with different types of sensory feedback a subject receives during the experiment, helping to answer questions concerning the hierarchy of factors influencing speech motor control.

1.2 Background

Altered Auditory Feedback

Houde and Jordan (1998) used real-time formant analysis and resynthesis of speech to change vowel formant frequencies during speech production. In the experiment, speakers whispered monosyllables with the vowel [ɛ] as in "head", and F1 was either lowered so that the vowel was closer to [ɪ] as in "hid," or raised so that the vowel was [æ] as in "had." This altered speech was fed back to the speakers through headphones in real time, so speakers heard themselves saying a different vowel from what they actually produced. In response, participants changed their F1 in the direction opposite the perturbation, in the first case producing a vowel closer to [æ] and in the second case [ɪ], so that the altered feedback sounded like the target word "head". Speakers overrode their usual well-practiced motor plans for producing "head" in order to compensate for the altered feedback, indicating that auditory feedback is heavily weighted during speech production.

Although it is clear that speakers do pay attention and are influenced by their auditory feedback, compensation is neither automatic nor entirely predictable in altered auditory feedback experiments, and different kinds of experimental and perceptual factors can influence how much a speaker might compensate for a given auditory perturbation. Niziolek and Guenther (2013) performed an fMRI study that manipulated how much auditory feedback was altered based on a speaker's own vowel category boundaries. For the same shift *magnitude* in acoustic space, speakers compensated more and showed a greater cortical response to shifts that pushed the output closer to a vowel category boundary than those that did not. That is, given a 50 Hz shift, if those 50 Hz pushed /ɛ/ closer to /ɪ/ (or 'head' closer to 'hid') rather than a 50 Hz shift that kept /ɛ/ well within the production space for /ɛ/, speakers will show a greater response in the former case than the latter. This indicates that speakers are not just trying to meet a goal at the acoustic level, but that they are aiming to land within a target perceptual category. If there is a perceived danger that the produced vowel will not be the target category, the response to the perturbation will increase. Thus compensation is not an automatic, low-level auditory process.

Experiments perturbing only articulation actually further show how speakers pay attention to their own acoustic feedback. Baum and McFarland (1997) fitted participants with artificial palates that had a 6mm alveolar ridge to test whether speakers could adapt to such a perturbation while producing /s/. Participants did improve with practice, but since only acoustic data was analyzed, it is unknown what types of articulatory strategies participants attempted. Thibeault et al. (2011) conducted a study using the same type of artificial palate, but they also tracked tongue trajectories using electromagnetic articulography (EMA). The aim of the study was to find out to what extent speakers could improve their acoustic output and whether their articulatory adjustments were extended to phonemes not affected by the palate. Articulatory data was collected, but not on a trial-by-trial basis, so the time-course of articulatory learning is not known. Although speakers never fully compensated for the effect of the perturbed articulatory space during either of these experiments, the fact that they made adjustments that correlated with improved acoustic output indicates that speakers changed their articulations in accordance with acoustic feedback: they could not be relying on tactile feedback alone due to the perturbation and reduced sensation due to the artificial palate. The fact that speakers' productions improved over multiple trials also suggests that the subjects were improving their articulatory strategy on successive trials and learning from previous failures, thus updating their new articulatory-acoustic map.

Auditory information is not the only feedback that speakers receive while talking. Somatosensory (tactile) feedback also provides information about where articulators are in space. Nasir and Ostry (2006) performed an experiment to isolate the role of somatosensory feedback in speech, independent of auditory feedback. A robotic device applied a lateral force on the jaw while subjects were speaking. Speakers compensated for this force, despite the fact that the perturbation did not affect speech acoustically. These studies show that speakers pay attention to their somatosensory feedback during speech production, regardless of whether there is an acoustic consequence for compensating or not. This could be interpreted as evidence for an articulatory target, but most certainly it is evidence that speakers are sensitive to what an articulation should feel like, and they can use this somatosensory information to help determine whether their articulation was on target or not.

Further, if somatosensory feedback does not match what was expected for the production of a given phone, speakers may limit how much they are willing to change their articulation, even if that means compensating less for an auditory perturbation. Larson et al. (2008) numbed the vocal fold mucosa with local anesthetic in an F0 perturbation study. Speakers' responses were greater in the numbing condition than the normal condition, suggesting that somatosensation is a limiting factor in compensation responses. In the absence of somatosensory information reporting a mismatch between the compensation response and a normal articulation, speakers relied only on their auditory feedback and compensated more strongly.

To address this apparent tradeoff between auditory and somatosensory feedback, Lametti et al. (2012) both independently and simultaneously altered both types of feedback. The somatosensory feedback was delivered by a robot that pulled the jaw forward so that formants would not be affected, but the position of the jaw would be altered. In this way, auditory and articulatory targets could be studied independently. Speakers varied in their compensation: some speakers compensated for one perturbation but not the other, and other speakers compensated for both. Speakers who compensated for somatosensory feedback showed less or no compensation for auditory feedback.

Size of perturbation could account for some differences seen in sensitivity to auditory and somatosensory feedback. Katseff et al. (2012) tested perturbation at 50, 100, 150, 200, and 250 Hz shift sizes and found that speakers' compensation did not scale linearly. Speakers compensated entirely or almost entirely for a 50 Hz shift, but only 100 Hz at the 150 Hz shift and not much more at greater shift sizes. They hypothesize that speakers may be trying to minimize both auditory and somatosensory mismatch. At smaller shift sizes, the entire shift can be compensated for without much tongue movement. This is not possible at larger shifts, so speakers must trade off between somatosensory and auditory mismatch. Given the results in Lametti et al. (2012), it is possible that at greater shifts, there might be greater speaker variation in how much of the acoustic perturbation they are willing to compensate for.

Taken together, these perturbation experiments show that the speech motor control system constantly checks and makes adjustments based on both auditory and tactile feedback. Speakers will forego practiced motor plans if they do not result in the expected acoustic output, and they will typically change how they are speaking so that their feedback matches their speech target, but the degree to which they compensate might be limited by perceived somatosensory mismatch. Further, speakers may differ in the respective weight they place on these different types of feedback.

Trading relations

Articulations are known to vary, even within an individual's own productions. One way articulation can vary is in the interaction between the articulators used to produce the phone. These are "trading relations," because one articulator (or part of an articulator) "trades off" with another. This was shown most famously for /p/ by Abbs and Gracco (1984). In this study, a robot was used to introduce an unanticipated load to the lower lip during the production of a bilabial stop. The weight of the load varied unpredictably, and it prevented a complete closure unless subjects compensated for the load. Subjects compensated for the load despite reportedly being frequently unaware of the perturbation. Importantly, compensation did not occur in the lower lip alone. The upper lip frequently compensated to meet the lower lip. This suggests that the articulatory target for a bilabial stop is not a concrete point in space that both lips must reach, but rather it is a more flexible target that is defined by a meeting of the lips with a movable geographical target. Further, the articulators were able to compensate for the load in real time. Speakers have internal knowledge of their articulators that allows them to modulate online how much "work" each articulator is doing.

The trading relations that Abbs and Gracco (1984) found call into question the nature of a speech target. It is debated whether a speech target is articulatory or auditory, and the results in Abbs and Gracco (1984) suggest that at the very least, if there is a sense in which the speech target is articulatory, it is certainly not spatially fixed, and neither is the muscular trajectory.

Namasivayam et al. (2008) explored how speakers may differ in their compensation strategies for articulatory perturbation depending on speech rate and whether or not they stutter. In this study, speakers were given a bite block and were point-tracked using an electromagnetic midsagittal articulograph (EMMA). Speakers said bisyllabic non-words containing bilabial stops, and both lips were tracked. In both populations, both lips were recruited to compensate for the articulatory perturbation, and speakers were able to compensate immediately after insertion of the bite block. These results support the idea of motor equivalence and that lip movements are part of a coordinated system.

Similar kinds of trading relations are also found *across* speakers for the same phone in the absence of a perturbation. Johnson et al. (1993) conducted a study using the Wisconsin x-ray microbeam database. In the study that generated the data, gold pellets were glued to the lips, tongue, and gums (as a measure of jaw height) and tracked with x-ray. Johnson et al. (1993) considered the displacement of the articulators during vowel production. They found that speakers differed from one another in how much they controlled the height of their tongue by using the jaw as opposed to the tongue itself, but that speakers tended to be consistent within their own articulations.

One phone considered in this dissertation is /r/. In American English, /r/ can have drastically different articulations that exist on a continuum (Delattre and Freeman, 1968) from retroflex, with a tongue tip that points up, to bunched, where the primary constriction is in the dorsum and the tongue tip points down (Lawson et al., 2011). Individuals not only vary in their articulation of /r/, they also vary in how many /r/ variants they might produce, often depending on the phonological context. Mielke et al. (2010) found, for example, that many speakers who have retroflex /r/ in onset context might have bunched /r/ in coda context; others might have retroflex /r/ only before non-front vowels.

Guenther et al. (1999) conducted an EMMA study examining differences in articulation and acoustics for American /r/. Subjects uttered nonce words of the shape /wa(C)rav/. Subjects differed in tongue shape as well as trajectory. Further, some subjects had gradient tongue shape differences by token, where separate categories of tongue shape were not easily identifiable, while others were more bimodal in the distribution of their tongue shapes, so that while there was some between-token variation in tongue shape, the articulation fell into one of two categories, which is determined by phonological context. The authors argue that the these trading relations between different parts of the tongue are not random, but rather that they actually help constrain acoustic variability.

The two different behaviors observed draws a distinction in the two different types of trading relations described here. The trading relations observed in the more gradient /r/ group are similar to those that Abbs and Gracco (1984) describe for /p/. For this /r/ group, the articulations are more or less similar to each other, though there is some variability in, for example, the exact location of the main constriction, and the tongue body compensates for slight shifts. For /p/, there is a sense in which there is a specific and well-defined articulatory target: a complete closure at the lips. Even though the distance traveled by each lip may vary by token for /p/, the articulatory target is more or less the same. In both instances, it is possible to define an articulatory target, and in both instances, it is possible to imagine that these tradeoffs might be happening in real time.

For the bimodal articulation /r/ subjects, it is clear that the articulatory trading relations are of a different nature than those for /p/. The differences in articulation for /r/ are not only a matter of degree of a certain quality, such as curvature of the tongue or constriction location. Retroflex and bunched /r/ differ substantially in the quality of the shape of the tongue. In an extreme case, a retroflex tongue is straight or even concave, and a bunched tongue is convex. In this case, it is more plausible to say that there are two separate articulatory targets, and it is less possible to imagine real-time corrections to speech resulting in one category of /r/ or the other. Although there might be some variation within each category of /r/, the category itself is predetermined based on phonological context.

This evidence suggests that the target for /r/ might not be articulatory for all subjects, or that there must not always be a single articulatory target. Where a closure at the lips might be the target for /p/, for /r/ the target must be auditory. In fact, Guenther et al. (1999) also found that these by-subject differences in articulation resulted in greater consistency in F3 for /r/. Subjects had individualized articulatory targets for /r/ that sometimes differed by phonological context in order to maintain consistency in the output.

Physical differences/Individual differences

One anatomical factor that is known to vary across people and to have an effect on articulation is the shape of the hard palate. Lammert et al. (2013) performed a principal components analysis over mid-sagittal MRI images of the hard palate and post pharyngeal walls of 36 adults who came from a variety of ethnic and language backgrounds. They found three main modes of variation: concavity of the palate, anteriority of the apex of the dome of the palate, and sharpness of the apex.

Brunner et al. (2009) used electropalatography to study the effect of palate shape on articulatory variability in the production of front vowels. Speakers with flatter palates had less articulatory variability (i.e. were more precise) than speakers with more domed palates. Brunner et al. hypothesized that this is because, while the mapping between articulation and acoustics is non-linear, it is less non-linear for people with flatter palates. That is, all else being equal, a smaller change in articulation results in a larger change in acoustics for people with flatter palates than for people with more domed palates. The acoustics of vocal tracts with flatter palates are more sensitive to changes along the vocal tract than that of vocal tracts with more domed palates. This vocal tract sensitivity function, or the relationship between articulation and acoustics for a given vocal tract, thus varies by individual due to differences in morphology.

I extended this study to the consonants /r/ and /s/, which have very different vocal tract configurations and aerodynamic requirements from the vowels studied in Brunner et al. (2009) and also from each other. I recorded speakers' /r/ and /s/ in different syllable positions and vowel contexts with lingual ultrasound and synchronized audio and took casts of speakers' palates and upper dentition. Like in Brunner et al.'s study, speakers who had flatter palates were less variable in the configuration of their /r/ than speakers with more domed palates, and they were also less variable in the configuration of their /s/. In the case of /r/, Brunner et al. (2009)'s hypothesis still stands: speakers did not vary much in the acoustic variability of their /r/, regardless of palate shape. Speakers appear to reduce their articulatory variability to constrain acoustic variability. In the case of /s/, though, speakers' acoustic variability was inversely correlated with the flatness of their palate, meaning that despite decreased *articulatory* variability, speakers with flatter palates had greater variability in the *acoustic* variability of their /s/.

These results suggest that speakers' internal knowledge of their own vocal tract sensitivity function, or how acoustics will change as a result of a given change in articulation, does affect their motor plans. Speakers whose acoustics are more sensitive to changes in articulation will constrain their articulatory variability, even if palate flatness has the final say in how acoustically variable they will be.

As the modeling evidence in Chapter 2 shows, speakers with flatter palates have fewer discontinuities in their articulation-acoustics map. For a given change in acoustics, speakers with flatter palates will require a smaller change in articulation compared with a speaker with a more domed palate. In contrast, speakers with more domed palates are more likely to have areas of the acoustic space that have no solution in the articulatory space. This would imply that speakers with flatter palates would be better able to adapt to altered auditory feedback, because they would be more likely to find an articulatory solution for each step of the perturbation. Conversely, such speakers are known to reduce their articulatory variability. If they have a more practiced motor plan for a given target, they might have greater constraints on how far they are willing to stray articulatorily from a given motor plan.

Theories of speech motor control

These altered feedback experiments and motor equivalence studies specifically test the role of different types of feedback in speech production, and these effects have been used as evidence for different theories of speech motor control, specifically how different types of feedback and feedforward mechanisms are integrated into the system.

The Directions in Velocities of Articulators model (DIVA) (Guenther, 1994; Guenther and Perkell, 2004; Perkell et al., 1997) describes a speech production system where speech targets are "convex regions" in space comprising multiple dimensions in auditory and orosensory space. In this model, the speaker has an articulatory-acoustic map learned through babbling in infancy. During this babbling phase, the speaker learns to associate auditory and tactile feedback with a given articulation. Because the articulatory-acoustic map is many-to-one, in order to achieve a particular acoustic target, the speech system finds the articulation that involves the minimum change from the current articulatory state of the system.

Houde and Nagarajan (2011) argue that the DIVA model is insufficient because it does not account for the very fast movements that are used in speech. If feedback were the primary way speech controls were generated, error correction would be delayed in a way that is not observed in actual online speech error correction. Houde and Nagarajan propose a state feedback control (SFC) model, which has previously been used in non-speech realms to explain how CNS controls motor output by estimating the state of a limb or articulator and using this estimate to generate controls. CNS is not able to use feedback to access the current dynamical state of the speech motor system at a given point in time and generate controls fast enough to correct for any error evident in the feedback, so it predicts the state of the system and uses this prediction to make controls.

These models of speech motor control contain feedback loops that send messages to the articulator control system. The experiments in this dissertation test an interaction between feedback and the control system. There is variation in the compensation response, suggesting that the feedback loop does not affect all individuals in the same way. The experiment in Chapter 2 shows that the shape of the palate shapes the initial mapping that speakers have between articulation and acoustics, and Chapter 3 shows that this mapping plays a significant role in the forward control system. The experiments in Chapter 4 and 5 test whether and how the shape of the palate may influence the compensation response, potentially affecting how CNS is able to predict the state of the system.

Chapter 2

Experiment 1: Modeling

Previous research (Brunner et al., 2009) shows that articulatory variability is reduced for people with flatter palates. It has been hypothesized (Brunner et al., 2009) that this is because the mapping between articulation and acoustics is more linear for flatter than for more domed palates. A combination of two synthesizers were used to model how vocal tract anatomy influences the mapping of articulation onto acoustics, using American English /r/ as a test case. A retroflex-able tongue tip was added to the articulatory parameters. Two additional palate shapes and a sublingual cavity that appears during /r/ production were also added to the synthesizer. A Python script searched the articulatory-acoustic space for vocal tract configurations that resulted in a low F3 (the hallmark acoustic cue for /r/) for each palate. Palate shape influences not only the overall sensitivity of the articulatory-acoustic mapping, but also the effect of each individual articulatory parameter on F3.

2.1 Introduction

Background

This research tests a hypothesis that explains how vocal tract anatomy influences variability in speech production. One segment that is known to vary both within and across individuals is the American English rhotic /r/. This segment can be either retroflex, bunched, or in between (Delattre and Freeman, 1968), and individuals may have more than one type of /r/, typically varying by phonological context (Mielke et al., 2010). This type of variation is an example of the many-to-one mapping between articulation and acoustics: multiple qualitatively different articulatory configurations may exist that will result in similar if not identical acoustics (Atal et al., 1978).

Even within one variant (such as bunched or retroflex) there may be variability. This is because the mapping between articulation and acoustics is also quantal (Stevens and Blumstein, 1975), meaning that for some regions of the vocal tract, a given difference in articulation will have a greater effect on acoustics than the same difference in a different region. Some regions of the vocal tract have relatively stable acoustic regions. Between regions of stability might be regions where small differences in articulation result in comparatively large changes in acoustics.

The degree of nonlinearity may not be the same for all individuals. In their electropalatographical study of front vowels, Brunner et al. (2009) found that people with flatter palates exhibit less articulatory variability than people with more domed palates. They hypothesize that this is because the mapping between articulation and acoustics is more linear for flatter palates, but for domed palates, there are greater regions of the articulatory space that are acoustically stable. Assuming speakers aim to maintain a degree of acoustic consistency, speakers with flatter palates must be more precise in their articulations than speakers with more domed palates.

Questions and Hypotheses

This study seeks to answer two main questions about the role of palate shape in variability in speech production. First, the modeling broadly examines how the mapping between articulation and acoustics varies by exploring the F3 acoustic space for different articulatory configurations for the different palate shapes. Second, the modeling assesses differences in the influence of various articulatory parameters on F3 for these different palate shapes. Specifically, this study tests the hypothesis in Brunner et al. (2009) that the increased articulatory precision observed in people with flatter palates is a result of a more linear mapping between articulation and acoustics for such palates. This hypothesis predicts a greater range of F3 space for flatter palates and a stronger relationship between each articulatory parameter and F3.

2.2 Methods

The hypothesis is that the shape of vocal tract (in particular the hard palate) plays an important role in the quality and variability of articulation in production. The work specifically considers whether, how, and to what extent articulatory configurations might differ in producing a low F3 for different palate shapes.

The original intent of the Maeda synthesizer (Maeda, 1990) was to model French vowels and includes a single palate based off a real speaker. The articulatory parameters are principal components based off of X-ray data from this speaker. There is an apex position parameter that controls the proximity of the apex to the palate, but this principal component also affects the tongue root. Because French vowels do not typically include retroflex tongue configurations, we created a new tip-curling parameter, which controls the orientation of the tongue tip only. We also created two new palates, one flatter and one much more domed than the default. We tested each of these palates with a spectrum of tongue shapes. Figure 2.1 shows the implementation of all three palate shapes and the tongue tip parameter.

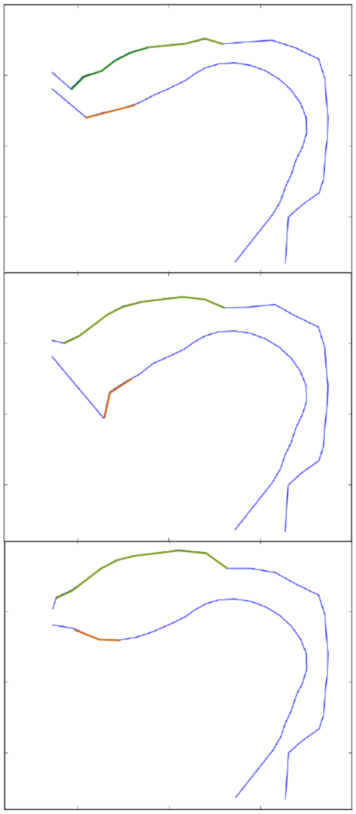


Figure 2.1: The three different palate shapes (from top: flat, default, domed) and examples of tongue tip configurations: neutral, extreme tip-down, and extreme tip-up.

The active articulators are driven by the user of the synthesizer. There are a number of parameters, such as the shape and position of the tongue dorsum or the protrusion and aperture of the lips, that the user adjusts to make different phones. The user manipulates the shape of the vocal tract by indicating a setting for each articulator. This setting is a multiplier for the principal component that represents the articulatory parameter. The four active articulator parameters considered here are the dorsum position and shape, the protrusion of the lips, and the orientation of the tip (whether it is pointing up or down), which represents an important difference in retroflex versus bunched articulations.

The Maeda synthesizer models the vocal tract as a series of cross-sectional areas. To calculate the area from the width in the sagittal plane we have to assume something about the shape of the tract at that point (if the vocal tract is a cylinder then the cross-sectional area is $A(x) = \pi r^2$, if the vocal tract is a square the cross-sectional area is $A(x) = d^2$, etc.). Within the oral cavity (as opposed to at the lips or in the pharynx) the Maeda synthesizer assumes that cross-sectional area is a function of palate doming, so that the cross-sectional area $A(x)$ at a point with a given width in the sagittal plane x is calculated based on the formula in (2.1), where and were determined from real production data from a single speaker and hard-coded into the original model. These α values are a special ratio of the width and

	Domed	Default	Flat
α	1.3	(1.7)	2.7
Minimum F3	1789	1371	1578
Maximum F3	2789	2713	3428
Min F2	788	704	684
Max F2	2252	2227	2035

Table 2.1: F2 and F3 ranges for each palate shape, considering all articulations.

depth of the palate and actually correspond to the same metric of domedness as in Brunner et al. (2009).

$$A(x) = \alpha x^{\beta} \tag{2.1}$$

The α values for the sections corresponding to the hard palate were set to empirically-derived values for the domed and flat palates based on Bakst and Lin (2015) in order to reflect realistic differences in cross-sectional area.

We used the Maeda model in conjunction with the Manzara tube model[1]. The Manzara model is a series of tubes of varying widths that are joined together. We added a short tube as a side-branch (after Espy-Wilson et al. (2000)) to model the sublingual cavity that emerges in /r/ production.

The model was run with each of the three palates. We wrote a program which cycled over the range of settings for each of the four articulatory parameters that would result in a possible articulatory configuration for /r/. The script produced cross-sectional area values, which were used as inputs for the Manzara tube synthesizer (Manzara, 1993). The script also called a program to perform acoustic analysis (Ueda et al., 2007) over the synthesized output and rejected tokens that were silent or not speech-like based on their RMS amplitude (amplitude < 1200). The program also recorded F2 and F3 measurements from the midpoint of the sound file.

2.3 Results

A summary of the results is in Table 2.1. The flattest palate has the widest range of F3 values, suggesting that articulatory-acoustic mapping may indeed be more sensitive for a flatter palate than a more domed palate, given that the same range of articulation was used for all palates.

Figure 2.2 shows the spread of F3 values for each palate. The generated sound files were sorted by F3 value; the spread (not value) along the x-axis corresponds to the number of sound files generated at a given F3 value. The closer to zero the slope is, the greater

---

[1] We used two synthesizers: the Maeda synthesizer is more faithful to articulation, and the Manzara synthesizer produces files of better sound quality.

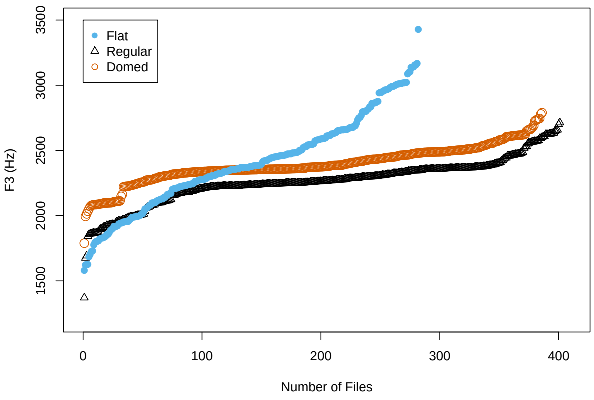


Figure 2.2: The F3 range for all three palates in increasing order of F3. The steady region for the flat palate is smallest and it is largest for the most domed palate.

the region of acoustic stability, and the less sensitive the mapping between articulation and acoustics. There is a greater range of values for the flat palate, less so for the default palate, and the smallest range for the domed palate. The overall acoustic flexibility is similar for domed and default palates; for much of the graph, the slope of the line is shallow. This indicates a large region of acoustic stability, where many articulatory configurations can result in similar if not identical acoustics. In contrast, the flattest palate has the steepest slope in this acoustic region, indicating the least acoustic stability for this palate.

While the F3 values reported in Figure 2.2 all come from articulations that might have hypothetically produced an /r/, some of these values are far too high to correspond with a phone that could be perceived as /r/. If we restrict our view to only those files produced with F3 values under 2300Hz, which is a reasonable cutoff for an /r/, there is less stability for the domed and default palates, but still more than for the flattest palate.

	Tip curl	Backing	Bunching	Lip
flat	0.15*	-0.78**	-0.56**	-0.06 (n.s.)
regular	-0.05 (n.s.)	-0.7**	-0.33**	0.15*
domed	-0.31**	-0.45**	0.36**	0.15*

Table 2.2: Correlations between articulators and F3 for each palate. $^{*}p < 0.05$, $^{**}p < 0.001$

Table 2.2 shows the correlation of each parameter with F3 for each palate. For all palates, the position of the tongue dorsum had a greater lowering effect on F3 than any other articulator, while lip rounding had minimal if any effect. The shape of the dorsum (bunching) had a surprising effect: for flat and regular palates, bunching of the tongue lowered F3, but for the domed palate, bunching actually raised F3. Conversely, raising the tongue tip slightly raised F3 for the flat palate, had almost no effect for the regular

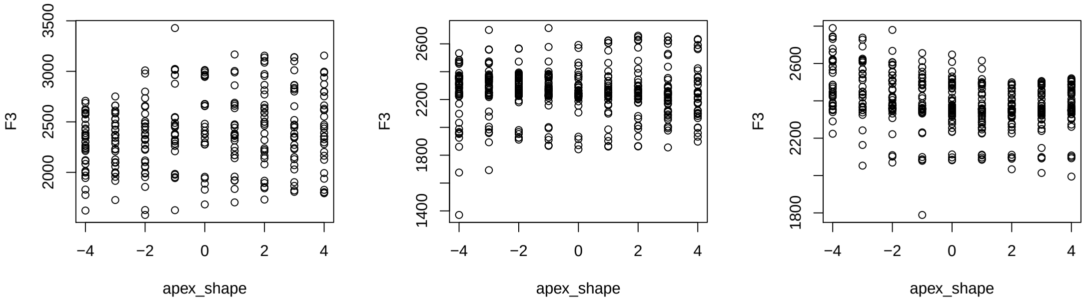


Figure 2.3: Spread in F3 for the flat, regular, and domed palates, respectively. Scale is the same although the range differs.

palate, and significantly lowered F3 for the domed palate. This result suggests that a flat palate would favor a bunched /r/, and a domed palate might favor a retroflex /r/. This relationship is weak, though: for all three palates, there was a wide range in F3 values for different settings of the tongue tip. Figure 2.3 shows that, given the other three articulatory factors tested here, a wide range of F3 values is possible for any setting of tip orientation.

For the flat palate, dorsum position and dorsum shape both had a great effect on F3. This supports the hypothesis that for a flatter palate, changes in articulation will generally have a greater effect on acoustics than for more domed palates. This holds true for all articulatory factors except retroflexion, where there is a much stronger relationship with acoustics for the most domed palate than for the other palates.

Figure 2.3 shows the articulatory settings for the five lowest F3 settings for each palate. For the first three parameters, settings range from 0 to 4, and for retroflexion, from -4 (tip down) to 4 (tip up). For all three palates, a low F3 was achieved primarily by retracting the tongue as much as possible, and secondarily by bunching the tongue. Lip protrusion appears to have no discernible effect in this table, even though lip protrusion is significantly correlated with F3 for regular and domed palates. Only the flat palate has a consistent pattern for the tongue tip; all five articulations for this palate have the tip either at a neutral or downward orientation.

2.4 Discussion

The modeling here suggests that the vocal tract sensitivity function is related to the flatness of the palate: the flatter the palate, the more acoustics are affected by a change in articulation. This was shown both in the range of F3 values produced with a flat palate in comparison with the default and more domed palates, and also in the large region of acoustic stability that was present for the more domed palates but not the flat palate.

The hypothesis proposed by Brunner et al. (2009) is that people with flat palates must

F3	lip protrusion	dorsum backing	dorsum bunching	retro-flexion
Flat				
1578.7	1	4	3	-4
1622.2	2	4	2	-2
1625.8	3	4	2	-1
1627.4	4	4	2	0
1683.0	4	4	3	-2
Default				
1371.0	4	4	3	-4
1675.6	3	4	3	-4
1692.8	4	4	3	-3
1842.6	2	4	2	0
1856.0	4	4	2	3
Domed				
1789.1	3	3	4	-1
1994.5	1	4	3	4
2013.9	1	4	3	3
2033.7	0	4	3	2
2053.1	0	4	3	-3

Table 2.3: Parameter settings yielding the five lowest F3 values for each palate.

reduce their articulatory variability to maintain acoustic consistency because their vocal tracts have smaller regions of acoustic stability. The hypothesis specifically applies to people with flat palates, and does not make predictions for the articulatory precision of people with domed palates. In the modeling here, the differences in the results from the three palates do not form a gradient. Rather, the domed and default palates have very similar results, with a large region of acoustic stability and similar slopes, but the flat palate has no regions of acoustic stability at all.

Palate shape not only influences the overall acoustic stability and flexibility of a vocal tract but also the effect of individual articulators on acoustics. Each of the articulators manipulated here had a different effect on F3. Most surprisingly, some factors (bunching of the tongue and orientation of the tongue tip) had opposite influences on F3 for the flat and domed palate shapes. This difference in effect of individual articulators provides a glimpse of an answer to the long-standing question of why some speakers have a retroflex /r/ and others have a bunched /r/. The shape of the palate is likely not the sole determining factor of a speaker's articulation, but it is certainly possible that the vocal tract is influential indirectly through this relationship between individual articulators and acoustics.

2.5 Conclusions

The models here test the hypothesis that the reason that people with flatter palates are articulatorily more precise is that the articulatory-acoustic mapping is most sensitive for such vocal tract shapes. The modeling shows a greater acoustic range overall for flatter palates. Changes in articulation are more closely correlated with acoustics for flatter palates than for more domed palates. This shown in how incremental changes in articulatory parameter settings have a greater effect on F3 for the flattest palate and the least on the most domed palate. Further, articulators seem to have different influences on acoustics in relation to each other for different palate shapes. This is seen in how closely linked each articulator is with acoustics; for example, while the position of the tongue dorsum has the strongest lowering effect on F3 for all three palates, the shape of the dorsum and orientation of the tongue tip have opposite effects for flatter and more domed palates.

The work supports the hypothesis in Brunner et al. (2009) that the articulatory-acoustics relationship is less quantal for flatter palates, and that this may be the reason that people with flatter palates are more articulatorily precise. The results also begin to answer how different palate shapes could influence articulatory variants for phones like /r/, which can have drastically different articulations.

Finally, the results here have implications for the organization of sound systems and may provide an explanation for the instigation of sound change. It provides evidence for the hypothesis in Stevens (1972) that the phonemes of a language are attracted to regions of acoustic stability. In a hypothetical community with a high ratio of speakers with flatter palates (and therefore less acoustic stability), we might find higher rates of sound change.

Chapter 3

Experiment 2: The role of palate shape on articulatory and acoustic variability in non-perturbed speech

This ultrasound and acoustics study of American English /r/ and /s/ investigates whether variability in production is related to individual differences in vocal tract morphology, namely the shape of the palate.[1] There was reduced articulatory variability for flatter palates for both /r/ and /s/. There was a relationship between acoustics and palate shape for /s/ only, where flatter palates had increased acoustic variability, but for /r/, there was no relationship between palate shape and acoustic variability.

3.1 Background

Individuals vary in palate size and shape (Vorperian et al., 2005). Brunner et al. (2009) have proposed that there is a relationship between palate shape and articulatory variability. In their electropalatographical study of front vowels, they found that individuals with flatter, less domed palates exhibit less articulatory variability. There is a more quantal relationship between articulation and acoustics for domed palates (Stevens, 1972), which predicts that a given amount of tongue displacement should result in greater acoustic change for flatter palates than for domed palates. Given this relationship, we would expect that a flat palate shape requires greater articulatory precision to remain under a certain threshold of acoustic variability. The experiments presented here test whether the Brunner et al. (2009) finding for front vowels extends to the two consonants /r/ and /s/.

Stevens (1972) wrote that the quantal mapping between articulation and acoustics gives rise to the *invariability* of speech. This hypothesis relies on the idea that regions of the

---

[1] American English /r/ is known to have a continuum of variants, from retroflex, where the tongue tip points up, to bunched, where the tongue tip points down. No relationship was found between palate shape and whether a participant's /r/ is bunched or retroflex, nor was one necessarily expected.

vocal tract vary in their acoustic stability. Stevens hypothesized that language "seeks out" such regions with greater acoustic stability, and from these regions "assembles an inventory of phonetic elements that are used to form the code for communication by language." This hypothesis links the shape of the vocal tract, and the resulting sensitivity of the mapping between articulation and acoustics, with the phonology of a language.

If we assume that the vocal tract does not differ across individuals, then these regions of acoustic stability also should not differ across individuals. This would predict that all languages draw from the same set of phones. The experiment presented here continues a line of investigation into the role of the shape of the vocal tract in speech production. If the size or location of these regions of acoustic stability *do* differ across individuals, it opens the possibility that communities of speakers with similarly-shaped vocal tracts may have a phonetic inventory that is skewed towards regions of stability for those palate shapes. This could potentially explain how physical differences between speakers could influence the phonological organization of different languages and language families.

3.2 Experimental Methods

Participants

Twenty-eight native speakers of Californian English, mainly students at the University of California, Berkeley, took part in this study (7 male, 21 female). Participants were compensated $15 for one hour of their time. Four (two female) were excluded from the final analysis due to experimental errors or ultrasound imaging that failed to capture a large portion of the tongue blade.

Stimuli

Stimuli for the experiment were English CV(C) words that contained /r/ or /s/ in onset or coda. The vowel environment was {a, i, o}. All words were presented in the carrier phrase "I'm a _____." Other words of the same shape containing /ʃ/ or /l/ as the segment of interest were collected but are not analyzed. There was also another set of words that served as fillers for this experiment that contained nasals and final coronal stops. All target stimuli and fillers are shown in Table 3.1, but only the /s/ and /r/ rows contain words analyzed in this paper.

Procedure

Ultrasound images were recorded using an Ultrasonix SonixTablet using a C9-5/10 microconvex transducer, operating at 107 fps. Audio was recorded with an AKG 535 EB microphone at a sampling rate of 48kHz and digitized with a Steinberg UR22 USB audio interface. A second audio channel of synchronization pulses (one per ultrasound frame) was sent from

Table 3.1: Stimuli

In this paper						
s	sob	sew	sea	boss	dose	piece
	/sab/	/soʊ/	/si/	/bas/	/doʊs/	/pis/
r	rah	Rome	ream	bar	bore	beer
	/ra/	/roʊm/	/rim/	/bar/	/bor/	/bir/
Not analyzed here						
ʃ	shaw	show	she	posh	gauche	quiche
l	lob	lobe	lee	ball	bowl	meal
Fillers						
wV(N)T	wad	Watt	wand	want	wan	
cV(N)T	cad	cat	canned	can’t	can	
bV(N)T	bed	bet	bend	bent	Ben	

the ultrasound to the UR22 and used to time-align the ultrasound frames with the speech audio signal.

A stabilization helmet from Articulate Instruments Ltd was used to hold the ultrasound probe in place under the chin in the midsagittal plane. A teleprompter displayed the test items, and the set of stimuli was randomized and repeated eight times. Participants also swallowed a small amount of water to generate a palate trace.

Each participant also provided a dental cast after the ultrasound session. Participants held a dental tray filled with alginate against the palate and upper dentition. Dental stone was poured into the resulting impression to form a permanent model of the participant’s palate.

Analysis

A script using the Penn Phonetics Lab Forced Aligner used the synchronized audio and the corresponding stimulus file (Yuan and Liberman, 2008) to create Praat (Boersma and Weenink, 2009) TextGrids for each acquisition. The script combed the resulting TextGrids for /r/ and /s/ and extracted the corresponding ultrasound frames using the synchronization file.

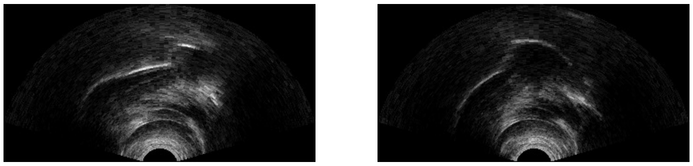
For each utterance, the frame corresponding to the acoustic midpoint of the segment was determined, and subsequent analyses were performed on those frames. The acoustic midpoint was chosen for consistency across tokens.

Measuring articulatory variability

EdgeTrak (Li et al., 2005) was used to trace the tongue contours in the selected ultrasound frames for four subjects (102, 103, 106, and 109). These traces were then visualized using the SSANOVA package (Gu, 2002; Davidson, 2006) in R and were used to make initial observations visually in order to corroborate and validate the subsequent principal components analysis (PCA)[2], which was then used to analyze the rest of the data. PCA was performed by-subject and by-phoneme.

Data from the first principal component (PC1), which by definition accounts for the majority of the variation in the data, was generally used. The exception was for /r/: all subjects who had a retroflex /r/ in some contexts also had a bunched /r/ in others. Therefore all subjects had a bunched /r/ in some context. A comparison between retroflex and bunched /r/ variants would likely result in an artificially large value of articulatory variability for those subjects who had both variants. Therefore, in order to maintain comparability across subjects, only variability in bunched /r/ is investigated in this study. For the few subjects who did have retroflex /r/ in some contexts, PC1 always separated out bunched from retroflex /r/, and PC2 showed differences within bunched /r/. For these subjects, further analysis was carried out using PC2.

Representative midpoint frames showing the difference in retroflex and bunched /r/ for subject 103 are shown in Figure 3.1a and 3.1b as examples. A retroflex /r/ is usually marked by an abrupt discontinuity in the tongue contour where the tongue tip is flipped up, whereas a bunched /r/ generally shows a smooth hill-like contour.



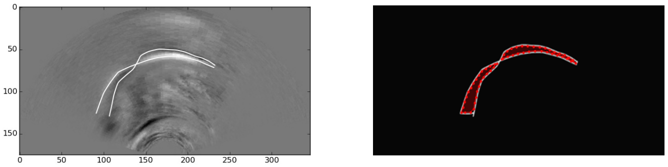
(a) 103 retroflex /r/ (b) 103 bunched /r/

Figure 3.1: Representative frames from 103's retroflex /r/ and bunched /r/. The bright white line in the middle of the fan is the air above the tongue, which for all intents and purposes here can be interpreted as the top surface of the tongue. Posterior is to the left and anterior is to the right.

Because PCA was performed over ultrasound images, the principal components themselves can also be represented as images, where the brightness of each pixel indicates how much that pixel loads on a PC, and lightness or darkness indicates positive or negative value for that PC. For example, a very white pixel indicates a strong positive loading for that PC, black a strong negative loading, and gray a loading closer to zero. In each image are two

---

[2] PCA is a dimensionality reduction analysis technique for data such as numbers or images that finds components along which variance in the data can be accounted for.



(a) Tongue contours traced in Pixelmator

(b) Area computed in GraphClick

Figure 3.2: Contours corresponding to the highest positive and negative loadings of PC1 were traced, and then the area bound by the two curves was computed.

tongue contours that correspond to strong positive and strong negative loading on that PC. The image representing the loadings for PC1 were loaded into the graphics program Pixelmator, and the two tongue contours were traced using an external tablet. The perimeter of the traces was automatically detected using the program GraphClick (Inc., 2010), and the area between the curves was calculated in pixels. The area was divided by the length of the tongue that was visible in order to account for individual differences in imageability. The resulting value was used as a metric to quantify individual variability in tongue shape. An example of a trace and the resulting area calculation is shown in Figures 3.2a and 3.2b.

Measuring acoustic variability

The acoustic measures for /r/ and /s/ were F3 and spectral peak, respectively, taken at the midpoint of the segment to match the point at which the ultrasound frame was chosen. A Praat script extracted F3, and a Python script calculated the spectral peak. The standard deviations of these measures were calculated over all of the utterances for each subject. These standard deviations were used as a measure of acoustic variability.

Measuring palate doming

Palates were measured following the method described in Brunner et al. (2009). They describe a value α that is correlated with the curvature of the palate and can be used to compare palate shapes. This value is calculated as below by approximating a slice of the palate in the coronal plane as a parabola (shown in Figure 3.3).

$$\alpha = \frac{4}{3\sqrt{|a|}}$$

In the equation above, a is the coefficient of the x^2 term of the equation describing the parabola, where half the width of the palate is the x-intercept of the parabola, and the depth of the palate is the distance between the x-axis and the vertex of the parabola. The α term is thus a kind of ratio of width to depth. The value of α is inversely correlated with domedness, so values closer to 3 are flatter, while values closer to 1 are more domed.

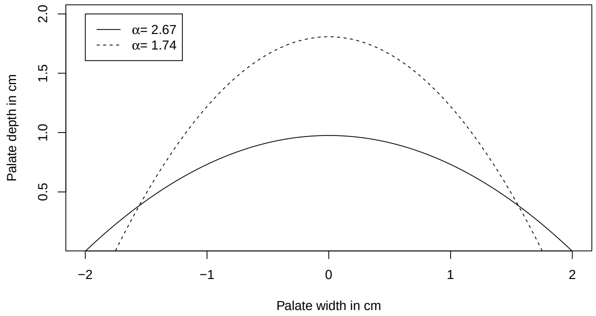


Figure 3.3: Approximation of the palate as a parabola for flat and domed palate shapes taken from participants in this study. For the derivation of α, see Brunner et al. (2009, p. 3948)

Results

	Articulation-palate	**Acoustics-palate**	**Articulation-acoustics**
/r/	$\mathbf{r = -0.43, p = 0.03}$*	$r = -0.18, p = 0.39$	$r = -0.12, p = 0.57$
/s/	$\mathbf{r = -0.41, p < 0.05}$*	$\mathbf{r = 0.50, p = 0.01}$*	$r = -0.20, p = 0.33$

Table 3.2: Summary of the significant relationships found in the data. Significance indicated by “*”.

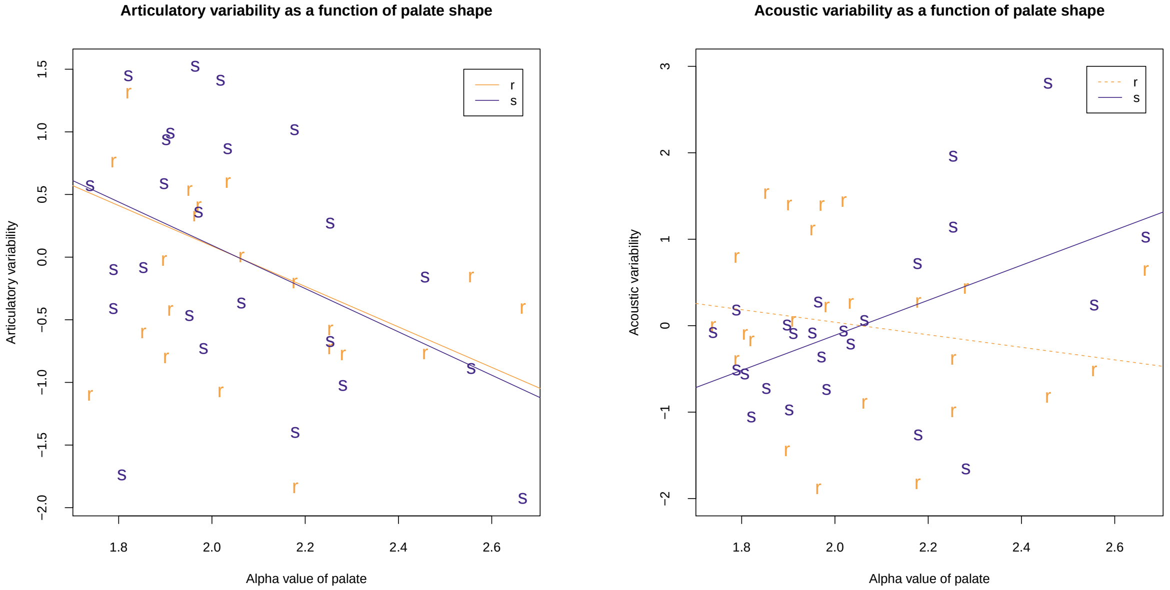
Qualitative results

Domedness of the palate was not a significant predictor of retroflexion or bunching ($p = 0.09$). Of the 26 subjects included in the qualitative analysis (two excluded for experimental errors that precluded further quantitative analysis), seven had a retroflex /r/ in some context. The mean alpha values for speakers with retroflex and bunched /r/ were similar (2.09 and 2.07, respectively). Domedness also not a predictor of /s/ articulation ($p = 0.23$). The split between apical and laminal variants was more even; 11 speakers had apical articulations. Mean alpha values for apical and laminal /s/ were also similar (2.09 and 2.06) respectively. Therefore, the data in this study do not support the idea that palate shape is a significant predictor of articulatory variant.

Quantitative results

Figure 3.4 shows articulatory variability plotted against palate shape in (a) and acoustic variability plotted against palate shape in (b). The relationship between articulatory variability and palate shape was significant for both /r/ ($r = -0.43, p = 0.03$) and /s/

($r = -0.41, p < 0.05$). The relationship between acoustics and palate shape, though, is significant only for /s/ ($r = 0.50, p = 0.01$). No relationship was found between acoustics and palate shape for /r/ ($r = -0.18, p = 0.39$). There was also no relationship between articulatory and acoustic variability for either /r/ or /s/. A summary of the results of the relationships between palate shape and variability in articulation and acoustics for /r/ and /s/ are summarized in Table 3.2.



(a) /r/: $r = -0.43, p = 0.03$
/s/: $r = -0.41, p < 0.05$

(b) /r/: $r = -0.18, p = 0.39$
/s/: $r = 0.50, p = 0.01$

Figure 3.4: Figure 3.4a shows articulatory variability graphed against palate shape for /r/ and /s/, and 3.4b shows the same for acoustic variability. All measures have been z-scored to allow for comparison across segment and variability type. Solid lines indicate significant correlations.

The observations above reflect group dynamics, but there is substantial individual variation. A hierarchical clustering analysis was performed on the data to identify categories of speakers. To help understand the interaction of the different variables in forming these clusters, multidimensional scaling (MDS) analysis was performed. The four factors (articulatory and acoustic standard deviation data for both /r/ and /s/) were reduced to two. The palate data was intentionally not included in the MDS analysis in order to test how well it correlated with the aggregate production data.

Dimension 1 of the MDS analysis was strongly negatively correlated with variability in spectral peak for /s/ ($r = -0.87, p < 0.001$) but very positively with variability in F3 for /r/ ($r = 0.73, p < 0.001$). Dimension 2 was strongly correlated with /s/ articulatory variability ($r = 0.87, p < 0.001$), but there was virtually no relationship with articulatory variability in /r/ ($r = 0.23, p = 0.27$), possibly due to the proportion of participants with domed palates and low articulatory variability. Although palate shape was not included in the MDS analysis, both dimensions significantly correlated with α value (Dimension 1: $r = -0.43, p = 0.04$, Dimension 2: $r = -0.49, p = 0.01$).

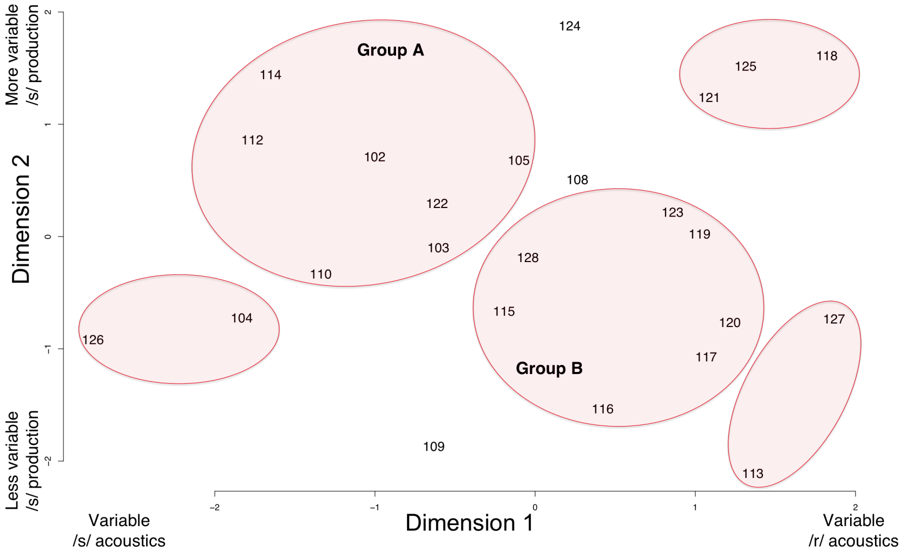


Figure 3.5: Participants grouped by cluster analysis performed on the data, plotted here on the resulting dimensions from MDS analysis. Dimension 1 is highly negatively correlated with variability in spectral peak for /s/ ($r = -0.87, p < 0.001$) but very positively with variability in F3 for /r/ ($r = 0.73, p < 0.001$). Dimension 2 is strongly correlated with /s/ articulatory variability ($r = 0.87, p < 0.001$).

Several groups emerge (shown in Figure 3.5). The two larger groups (Group A and Group B in the figure) reflect the main ways in which people differ. Group A had low acoustic variability for /r/, though the source of this consistency was different for different members of the group: some had low articulatory variability, and others had very domed palates. The members of Group B were marked by their high articulatory variability in /r/ and low /s/ acoustic variability, and all had average to domed palates. The lower left group containing 126 and 104 had fairly flat palates and correspondingly high acoustic variability for /s/ despite low articulatory variability. The group containing 118, 121, and 125 had fairly domed palates and high articulatory variability for /s/, but their acoustic variability was average. In contrast, the group in the lower right (113 and 127) had very low articulatory variability despite not having particularly flat palates. In a sense, this group's articulation was more consistent than such a palate shape would predict to reach a particular level of acoustic consistency.

3.3 Discussion

Brunner et al. (2009) hypothesized that the motivation for individual differences in articulatory variability is that all speakers aim to minimize acoustic variability. Individuals may differ in the *value* of the F3 of their /r/ or the spectral peak of their /s/, but the standard deviations in these values are probably going to be relatively consistent from one individual to another. If there is a more linear mapping between articulation and acoustics for vocal tracts with flat palates than domed palates, then vocal tracts with flatter palates will have smaller regions of acoustic stability. Speakers with such vocal tracts must reduce their articulatory variability if they wish to stay within (or not deviate far from) such a region of acoustic stability. Therefore, we expect that while variability in articulation may change in accordance with palate shape, there might not be a relationship between articulatory and acoustic variability. This explains the main trends that we see in the data: there is a relationship between articulatory variability and palate shape, but no relationship between articulatory and acoustic variability.

Although there was significantly less articulatory variability for flat palates for both /r/ and /s/, the pattern differs between the two segments. For /r/, flatter palates are articulatorily highly consistent, but within the domed group (where α is under 2.0) there is less of a relationship between articulatory and acoustic variability, and there are more participants like 105, who have reduced articulatory variability despite a more domed palate. In other words, while a flat palate predicts articulatory consistency, a domed palate does *not* necessarily predict increased variability in the articulation of /r/. For /s/, the relationship is more linear, so that even within the domed group, there is a strong relationship between palate domedness and acoustic variability. The trend is therefore the same for both segments, but the extent to which the palate shape has an effect on acoustics for all speakers depends on the segment. While palate shape might be one interfering force that results in individual differences in articulatory variability, there are other factors that have different influences over the two segments. For some participants with domed palates (e.g. 121) who have high articulatory variability for /r/, it is possible that there is not a way to reduce articulation to the same degree for /s/ while also maintaining the constriction required for turbulent airflow. This possibility is corroborated by the fact that there is no correlation between an individual's articulatory variability in /r/ and /s/. While palate shape certainly influences articulatory and acoustic variability, on an individual level there are many other possible factors yet to be investigated that might also play important roles.

Despite the decrease in articulatory variability, there is greater acoustic variability in /s/ for participants with flatter palates. The difference in the significance of the relationships between palate shape, articulatory variability, and acoustic variability for /r/ and /s/ suggests that consistency in the acoustic goals are not equally attainable across individuals and phones. From the MDS analysis, it seems that there are different degrees of attainability for this consistency. While people with flatter palates may increase their articulatory precision in order to maintain acoustic consistency, their palate shape is a limiting factor in how much acoustic consistency they can achieve.

Individuals are also likely to differ in how much variation they are willing to tolerate in their own acoustics and articulation. Although the main motivating factor seems to be maintaining acoustic consistency, it is possible that some speakers weight their somatosensory feedback more strongly than others (Lametti et al. 2012). This might explain both why some some speakers are articulatorily very consistent despite having domed palates (e.g. participant 105), and it also might explain the small differences in acoustic consistency in /r/. Speakers may also differ in their sensitivity to certain acoustic cues, or in the knowledge of their own articulatory-acoustic mapping.

3.4 Conclusion

Palate shape does play a significant role in individual differences in articulatory and acoustic variability, but the weight of this role depends on the segment. For /r/, there was relative consistency within-individual in acoustics, but this did not extend to /s/. Acoustic variability did not correlate with articulatory variability for either segment, which is to be expected if speakers' articulatory variability is related to consistency in acoustics. For /s/, there was a strong relationship between both types of variability and palate shape, but for /r/ there was a strong relationship between palate shape and articulatory but not acoustic variability. There may be a difference in attainability of acoustic goals. For /r/, speakers seem to be adjusting their articulations in order to maintain a relatively consistent F3. The spectral peak of /s/ is much less consistent than the F3 for /r/. It is possible that the acoustics for /s/ are more strongly influenced by the shape of the palate than /r/, so acoustic consistency is more easily attained for /r/ than /s/.

The results presented here provide part of the answer to how morphological differences between individuals may influence speech production. The domedness of the hard palate is one such anatomical factor, but many other factors, such as the size of the oral cavity or the size and range of motion of the tongue, are likely to exert their own influences. Further research needs to take more of these factors into consideration in order to understand how differences in vocal tract morphology affect production.

Chapter 4

Experiment 3: The role of palate shape on compensation in altered auditory feedback

4.1 Introduction

The experiment in Chapter 3 found an effect of the hard palate on articulatory and acoustic variability in normal speech. The goal of the experiment in this chapter is to determine whether this effect extends to how speakers correct their motor plans online by considering how (articulatorily) and how much (acoustically) speakers compensate in an adaptation to altered auditory feedback task. In this task, subjects hear themselves while they are speaking, but their vowels are resynthesized with different formants from what they actually produced. A typical response is compensation for this feedback, where speakers begin to alter their own formants in a way that opposes the perturbation: if F1 is raised in the experiment, participants will lower their F1 so that after alteration, their feedback matches the target word. Participants in such tasks are known to vary both in *whether* they will compensate for any of the altered feedback and also *how much* they will vary. Further, it is unknown what articulatory strategies speakers use to compensate.

Hypothesis

The previous chapters make predictions for how palate shape interacts with speech production. The modeling chapter showed that flatter palates have a vocal tract sensitivity function whose acoustics are more sensitive to differences in articulation than more domed palates. In other words, for a given magnitude of change in articulation, there will be a greater acoustic change in a vocal tract with a flatter palate than a more domed palate.

The experiment following this modeling prediction determined that this vocal tract sensitivity difference has an effect on articulatory and acoustic variability in normal, unperturbed speech. The results indicate that for some types of speech segments, including vowels (Brunner et al., 2009) and /r/ (Chapter 3), speakers with flatter palates will constrain their articulatory variability in order to maintain acoustic consistency. This follows directly from the difference in vocal tract sensitivity observed for different degrees of palate curvature. If speakers with flatter palates have a more sensitive articulatory-acoustic mapping, then in order to maintain consistency in the output, such speakers must also maintain consistency in the input.

This gives rise to two opposing predictions for the role of the shape of the palate in the altered feedback experiment. From one viewpoint, if speakers with flatter palates have a more sensitive articulation-acoustics mapping, they might be faster and more complete compensators for several reasons. First, the less-quantal nature of this mapping means that they might have more flexible acoustics and therefore finer control over their output. Second, because of the greater effect of a given change in articulation on acoustics, they might also be able to achieve complete compensation with less deviation from their typical motor plan: in the same way that a given articulatory change has a greater effect on acoustics for flatter palates than for more domed palates, a given magnitude of acoustic change might be achievable with a smaller tongue re-configuration for a flatter palate. This part of the prediction comes with a caveat, though: the articulatory-acoustics map is many-to-one (Atal et al., 1978), so it is possible that for some changes in acoustics, a speaker with a more domed palate and a speaker with a flatter palate might have equally small changes in articulation to create the same change in acoustics, depending on the vocal tract.

Assuming that speakers with flatter palates do require less change in articulation to compensate more fully, then they would be able to compensate at a lower cost. When speakers change their motor plan to compensate, they may experience a mismatch in somatosensory feedback, because their tongue will be in a different place than when producing it under non-altered-feedback conditions. Speakers who need to change their tongue position *less* to create such an acoustic change will experience less of a somatosensory feedback mismatch than speakers who must change more, which could result in greater compensation at a smaller cost.

Finally, if speakers with flatter palates must constrain their articulatory feedback in regular speech to maintain acoustic consistency, they might be more attentive to their auditory feedback and thus more motivated to compensate for the altered feedback.

The behavior observed in Chapter 3 can also predict the opposite outcome. If speakers with flatter palates are less articulatorily variable, they might have a stronger and more precise representation of the motor plan required to produce a particular vowel. This might make them more likely to notice a difference in articulation, and less likely to deviate from their usual motor plans.

These hypotheses boil down to whether we expect that the domedness of the palate only affects the articulatory-acoustic mapping, or if its influence extends to an individual's *articulatory habits*. If palate shape only goes so far as the vocal tract sensitivity function, then we would expect people with flatter palates to have more control over their acoustic output. If the palate shape exerts a strong influence over an individual's articulatory habits, then we would expect speakers with flatter palates, who tend to be articulatorily consistent, to be resistant to the altered feedback, as they will be less willing to change their habits.

4.2 Equipment

FUSP

This experiment used the Feedback Utility for Speech Production (FUSP), a device that records speech and outputs an altered version in real time. The software was written by John Houde and is implemented in a collection of scripts written in C. For this experiment, a Python version of the FUSP controller was written by Ronald Sprouse.

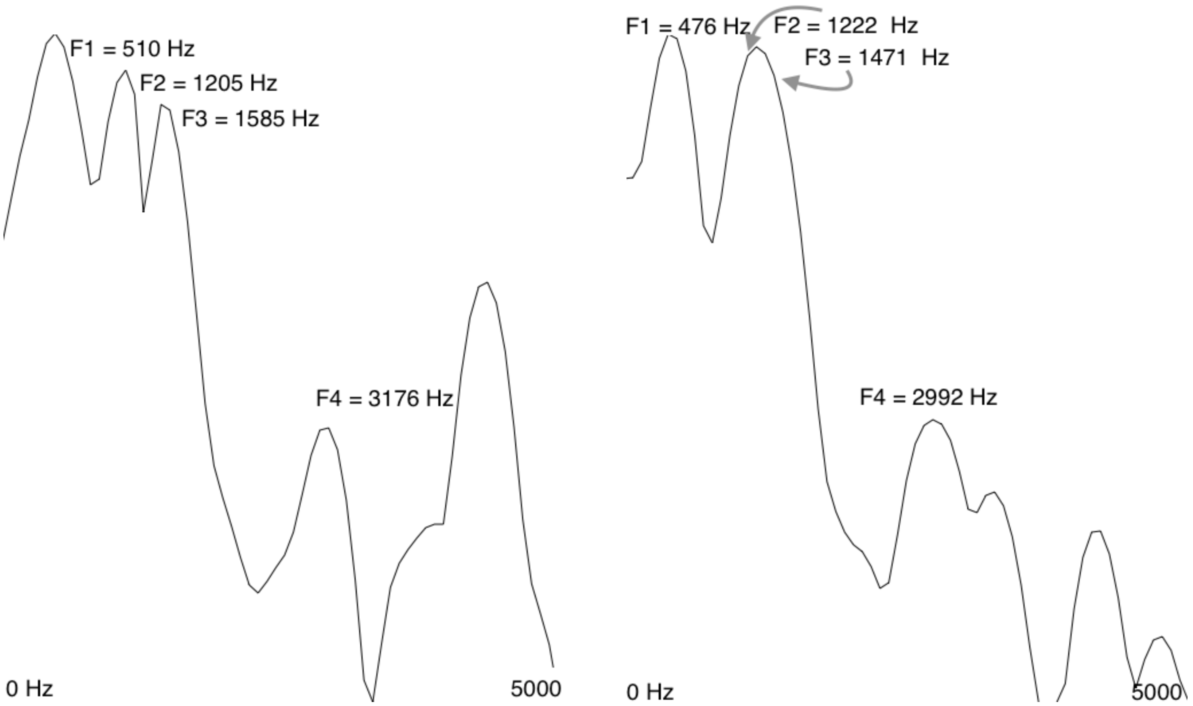
Speech audio was recorded with an AKG 535 EB microphone at a sampling rate of 48 kHz, and it was digitized with a Steinberg UR22 USB audio interface. This audio was sent to a designated FUSP computer that down-sampled the audio to 11025 Hz, performed the alteration, and sent it to the participant over research-quality earbuds (Etymotic ER4B MicroPro Binaural In-Ear Earphones).

FUSP analyzes the spectrum for each sample to select peaks that are likely to be formants. This is done on each frame independently. Once formants are determined, the perturbation (if any) is applied, and the frame is resynthesized and sent to the participant as the next frame is analyzed. While this allows FUSP to operate close to real-time (estimated 12ms delay), there is the potential for errors that are less likely to arise in other formant-tracking methods where time is not an issue.

The FUSP peak-picking method has several parameters that can be changed to tailor the device to a participant's voice, thereby improving accuracy. The first is the number of LPC coefficients the device uses to pick peaks. The greater the number of coefficients, the more sensitive the device is to smaller peaks in the spectrum, as well as peaks that sit close together. Fewer peaks means more smoothing and a greater chance that a smaller peak might be missed. The second is a high-pass filter, which emphasizes higher peaks in the spectrum that are otherwise low-amplitude and might be passed over by the peak-picker. Finally, a new parameter was added for this experiment, which sets limits for how low the lowest formant can be and how high the highest formant can be.

Peak picking presented a particular challenge for this experiment. In the third block of this experiment, the third formant from /r/ was altered. One of the hallmark acoustic characteristics of /r/ is that the third formant lowers close to F2. This increases the potential for F2 and F3 to be interpreted by FUSP as a single peak in the spectrum, and for F4 to be subsequently interpreted as F3. During /r/, F3 is typically around 1500 Hz in a male voice, and F4 is above 3000 Hz. By increasing sensitivity to smaller peaks higher in the spectrum, as well as setting a cap on the maximum value for F3, the chances of choosing the correct peak as F3 is substantially increased.

All participants provided a baseline sample where they heard themselves with no alteration. This was submitted to a MATLAB script that tested the different parameters. The parameters that maximized peak-picking accuracy were set before the first trial block.



(a) F2 and F3 sufficiently far apart to be discriminated as two different peaks by FUSP

(b) F2 and F3 register as a single peak in this spectral slice

Figure 4.1: Spectral slices from midpoints of two different productions of /r/ from the same speaker. F2 and F3 are sometimes far enough apart to be distinguished as separate formants (4.1a), but in other productions they are so close together that they appear as a single peak (4.1b).

Ultrasound

Ultrasound images of the tongue were recorded with an Ultrasonix SonixTablet using a C9-5/10 microconvex transducer operating at 107 fps. The ultrasound sent a synchronization pulse for each ultrasound frame that was also recorded and digitized with the same interface as for the speech audio. This pulse was used to align speech audio with ultrasound.

Subjects were fitted with a stabilization helmet from Articulate Instruments Ltd, which held the ultrasound probe in place under the chin.

Video

Participants were also video recorded in order to observe movement of the lips. Participants applied blue lipstick to allow for better extraction of lip shapes during processing. They were recorded with a Canon XF100 recording at 60fps. (Video data is not analyzed here.)

Palates

A cast of the palate and upper dentition was taken after the recording session concluded. The participant inserted a tray of dental alginate into his own mouth. The resulting impression was filled with dental stone.

4.3 Procedure

Stimuli

Three formants were independently altered on three separate blocks. This was done to ensure that if a speaker did not compensate for a particular formant, it could be determined whether the speaker did not pay attention to altered feedback at all, or just a particular formant.

Stimuli were displayed using a teleprompter so that speakers would look directly into the camera.

The words in the altered feedback experiment were 'head', 'hood', and 'heard,' and F1, F2, and F3 were raised in these three words respectively.

To obtain the speaker's vowel space, the words 'had', 'hawed,' 'hade,' 'hoed,' 'hid,' 'hud,' 'heed,' and 'who'd' were additionally elicited.

Experiment

First, the participant swallowed a small amount of water to obtain a trace of the palate with the ultrasound. The ultrasound recorded images throughout the experiment. Figure 4.2 shows the physical setup of the experiment.

There were four blocks in this experiment. OpenSesame (Mathôt et al., 2012) was used to present stimuli, interface with the ultrasound, and send commands to FUSP. On each trial, the stimulus was presented on the screen, and the participant said the word and pressed a key to advance to the next trial. At the same time, the participant's voice was recorded and fed back to the participant over headphones.

The first block served as a baseline for the participant's articulatory and acoustic vowel space. There were fourteen words, and the subject said each word twice. This block also provided a sample for the experimenter to tailor formant-tracking parameters to the participant's voice.

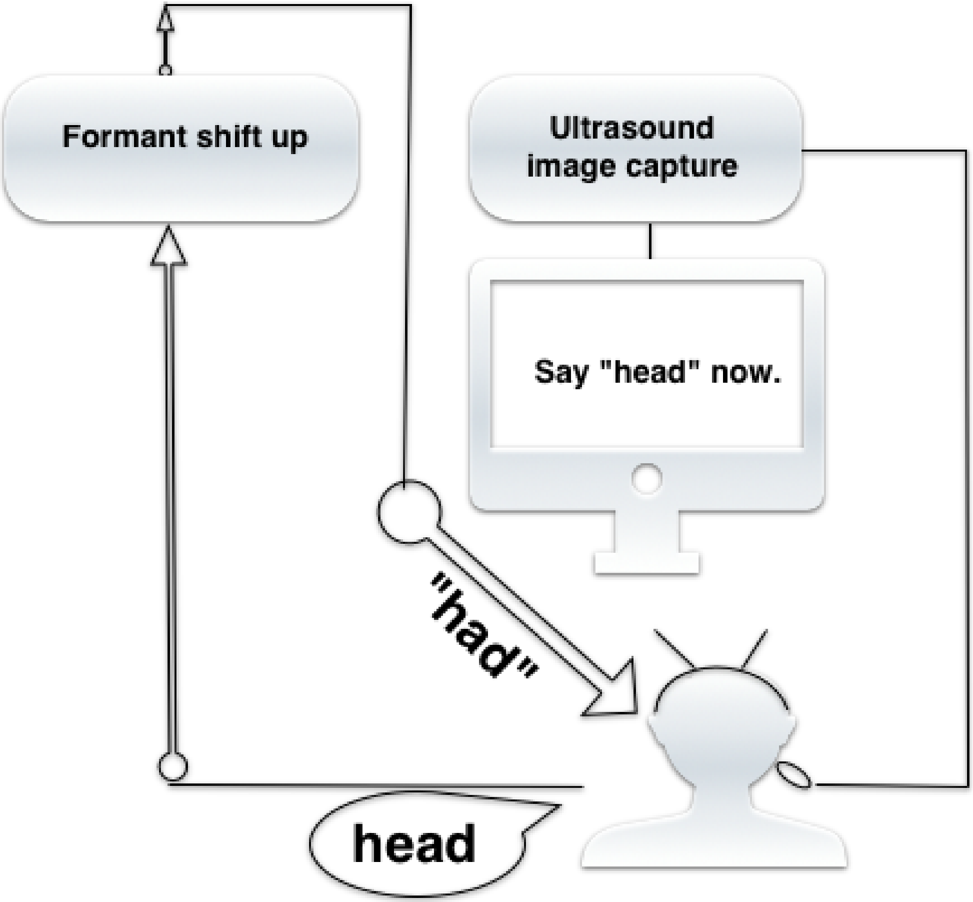


Figure 4.2: Experimental setup: the subject sat facing a teleprompter that displayed the stimulus. His produced audio was resynthesized with a shifted formant, which was fed back to him over headphones. A helmet held the ultrasound probe in place under the jaw.

The second, third, and fourth blocks each altered a single formant (F1, F2, and F3, respectively). Each block had four phases: baseline, ramp, hold, and washout. The first fifteen trials served as a baseline. Subjects heard their own feedback, but no alteration occurred. The next 30 trials increased the altered formant by a constant amount on each trial. During the hold phase (30 trials), the alteration to the single formant was held constant. During washout, alteration abruptly ended; these 15 trials were identical to the baseline phase.

During blocks 2 and 3, the ramp was +8 Hz and the hold phase held the perturbation at +240 Hz. During block 3, the ramp was +15 Hz and during the hold phase the perturbation was held at +450 Hz.

The shifts were chosen to maximize the absolute compensation, not merely the proportion of perturbation. Previous research Katseff et al. (2012) show that there is a limit to the proportion of perturbation that speakers are likely to compensate for, and this does not scale linearly. For lower formants, this proportion begins to drop off at about 150 Hz. For 50 Hz shifts, speakers are likely to compensate for 100% of the perturbation, but at 250 Hz, they might only compensate for about 50% of the perturbation. A 50 Hz compensation is still much smaller than a 125 Hz compensation. The present study investigates the articulatory strategies speakers use in compensating for these shifts. A shift of 240 Hz was therefore chosen to maximize the chances of capturing the articulatory changes with ultrasound, and also to capture differences between speakers in how much they will compensate.

The entire experiment, including the dental impression, lasted about one hour.

4.4 Analysis

This experiment examines the relationship between articulation and acoustics in responses to altered auditory feedback. The steps of the analysis include determining the timepoints at which to perform acoustic and articulatory analysis as well as whether and how subjects responded to the auditory feedback perturbation.

Acoustics

All analysis in the F1- and F2-altering blocks was done at the acoustic midpoint of the phone of interest. The audio that FUSP recorded, both of what subjects produced and heard, were converted to wav files, which were force-aligned to Praat TextGrids using the Penn Forced Aligner (Yuan and Liberman, 2008). For two subjects (328 and 329), a technical error made the FUSP audio unusable, so the input from the audio that had been recorded as part of the ultrasound acquisition (split from the same microphone signal that was sent to FUSP) was used instead and force-aligned in the same way. The time intervals in the TextGrids were used to divide each phone into seven equal parts using a Python library written by Ronald Sprouse. Formant calculations were done using ifcformant (Ueda et al., 2007).

For the F3-altering block, the timepoint used for acoustic measurement was found using an articulation-based method described below. Because of formant-tracking difficulties imposed by /r/, the rformant utility (based on the *formant* utility from Entropic Signal Processing System, (Sprouse and Johnson, 2016; Lee, 1988)) was used to track /r/ acoustics.

Ultrasound

For all blocks, the Penn Forced aligner (Yuan and Liberman, 2008) created a Praat TextGrid that was time-aligned with the audio. There were two methods for acoustic analysis: one for the F1- and F2-altering blocks (henceforth “vowel blocks”) and one for block three (the “r-block”). For the vowel blocks, the acoustic midpoint of the vowel (as determined by the alignment) was used for the acoustic analysis, and the closest ultrasound frame to this timepoint was also taken. For vowels, there is typically a long steady state portion in the middle of the vowel utterance, so the midpoint is representative of the articulatory strategy used for that vowel.

Unlike for monophthongal vowels, the acoustic time-course for /r/ is not symmetrical: over the course of /r/, F3 decreases. For some tokens and/or individuals, F2 and F3 approach each other, but for others, both F2 and F3 decrease (see again Figure 4.1 for reference). Using the time-point corresponding to the lowest F3 is not necessarily reliable, though, because most formant trackers do not reliably distinguish F2 and F3 as separate peaks.

It was found to be more reliable to find the *articulatory* point where there was the least movement during the /r/. This point was found by extracting the sequence of raw /r/ frames based on the Praat TextGrid that was created from the aligner. Each frame is represented as a matrix, where each element of the matrix represents a pixel with a brightness value. For each frame matrix in the sequence of /r/ frames, each frame was subtracted from the previous frame, resulting a matrix representing the difference between those two frames. The *norm* of this difference matrix, or the square root of the sum of the square of each element of the matrix, represented the magnitude of this difference. The lowest difference represents the least change from one frame to another, which is a proxy for decreased velocity of the tongue. This is the point where the /r/ approaches steady-state, and it represents the articulation that the subject has settled on. This frame was used for both a timepoint for acoustic measurement as well as further articulatory analyses for the /r/ block.

Palate curvature

Palate curvature was determined as per previous chapters (also Bakst 2015, after Brunner et al. 2009), by approximating a coronal slice of the palate to a parabola. The width and depth of the palate were measured between teeth # 15 and 16, and these measurements were used to calculate the equation for a parabola of the same curvature. The a term from the parabola equation $y = ax^2 + b$ is used to find the α term described in Brunner et al. (2009) according to the formula below.

$\alpha = \frac{4}{3\sqrt{|a|}}$

or, equivalently:

$\alpha = \frac{2}{3\sqrt{(depth/width^2}}$

4.5 Determining response to perturbation

A typical response to altered auditory feedback is a change in production in the direction *opposite* to the perturbation that was applied. If F1 is raised in the experiment, the expected response is a lowered F1 to counteract the perturbation. For example, if a target "head" is shifted towards "had," a subject is expected to shift his productions towards "hid" so that he hears "head" in his feedback. Median formant values from each phase of the experiment were calculated. If the median decreased between the baseline and the ramp or hold phase, the subject was determined to be a "compensator."

Subjects differ in how much they will compensate, and this is related to many factors, including whether the perturbation approaches a vowel boundary (Niziolek and Guenther, 2013), how much a speaker weights his somatosensory feedback against his auditory feedback (Lametti et al., 2012), or even personality traits like a speaker's optimism or sense of power (Dimov et al., 2012).

In the F1-altering block of the experiment, subjects widely differed in the point at which they stopped compensating for the perturbation, and in how much they compensated, acoustically speaking. In the F2-altering block, there was more compensation, and subjects were more consistent in adapting throughout the entire experiment. The actual acoustic perturbation in block 3 was not always successful, so this block will be considered on a subject-by-subject basis in Chapter 5. Group generalizations will be drawn for the vowel blocks in this chapter.

Because of these individual speaker differences in compensation, it is not possible to pick a particular trial number to compare across subjects. To determine *whether* subjects compensated in a given block, medians were calculated for each phase (baseline, ramp, hold, and washout) of the block. Subjects were considered to have compensated if the median value of the formant under investigation in *either* the ramp or hold phase was lower than the median value of that formant in the baseline.

Altered feedback data can be noisy due to frame-dropping from the FUSP computer during the experiment. Sometimes there is an error in the data, but a formant tracker may still calculate a reasonable-seeming formant value, generally due to random luck. To account for this and other formant tracking errors, every group of five successive trials was placed in a bin. The median formant value for that bin was used as a representative value for further analysis. Thus for each 90 trial block, there are 18 measurements (3 each from baseline and washout, and 6 each from ramp and hold phases).

Of the middle 12 measurements (i.e. during altered trials), the minimum value was determined and the trial matching this value was selected as the trial of greatest compensation for that block. The difference relative to the subject's average baseline measurement, taken as a percentage of the 240Hz total perturbation, was used to quantify amount of compensation. These measurements were done in Bark, which is an auditory-based scale, to enable comparison across subjects and blocks. Thus this 240 Hz shift was converted to Bark relative to a subject's baseline average.

Quantifying Tongue movement

The magnitude of tongue movement was determined for subjects who compensated for the altered feedback. Ultrasound presents a challenge for quantifying tongue movement because unlike other methods, such as point-tracking, there is no precise way to quantify the distance any particular point on the tongue has traveled. Given the success of determining the point of steady-state using the difference between successive ultrasound frames as described above for the /r/ block, a subtraction method was also used as a metric for how much individuals changed the shape of their tongues in comparison with their baseline in order to achieve compensation.

To obtain a baseline, midpoint frames from the baseline trials were averaged. The first trial was excluded, as subjects were often surprised at the start of the block, and this frame was not representative, either articulatorily or acoustically, of the following trials. An example of this average baseline is shown in Figure 4.3.

Figure 4.3: Average baseline image created from averaging baseline frames together.

The trial where the greatest acoustic compensation was determined (described above in section 4.5), and the frame matching the acoustic midpoint of the vowel in this trial was selected. The average baseline matrix was subtracted from the matrix representing the frame from the trial of greatest compensation. The algebraic norm of this difference was calculated to determine the magnitude of the difference between these two frames. This was divided by the norm of the average baseline frame. This was done to account for between-subject differences in how clearly a subject imaged and how much of the tongue was visible: depending on how well a subject images, there will be different ratios of white to black in the image.

Thus, for a matrix corresponding to the frame of greatest compensation F, and the matrix corresponding to the average baseline frame B, the following calculation was performed: $\frac{||F-B||}{||B||}$. This value was used as a measure of how much the tongue had changed shape from the baseline. Figure 4.4 shows a frame representing the difference F-B.

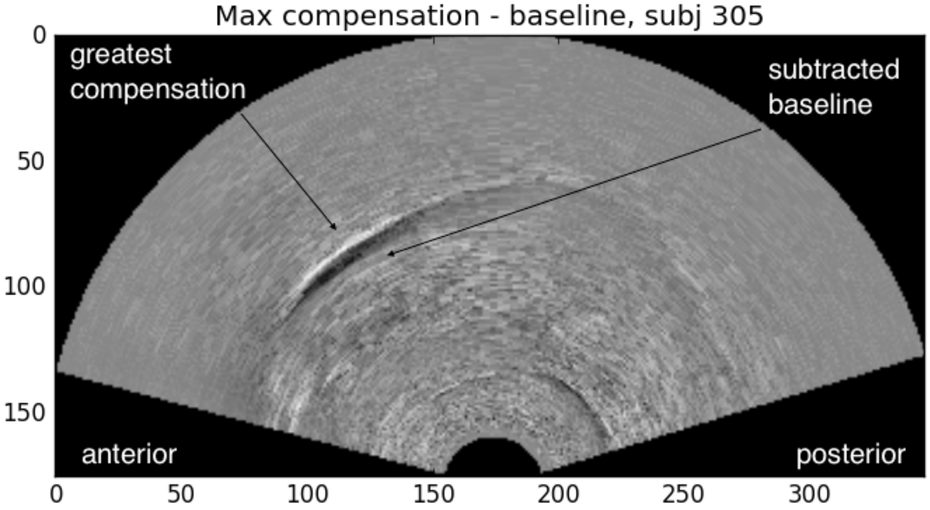


Figure 4.4: A graphical representation of how much the tongue has changed between the baseline and the trial representing the greatest compensation. The tongue is usually visible as a white line in an ultrasound frame. The black line indicates where the baseline has been subtracted away.

Principal Components Analysis

The vowel midpoint frame from each trial was selected using the TextGrids. These frames were compiled into a separate set for each block and subject, so each subject had three separate sets representing their articulations. Principal components analysis (PCA) was performed over each set. PCA is a statistical technique that finds independent axes along which the data varies. Principal components (PCs) are ranked by the amount of variation in the data they explain. This method thus finds patterns in tongue positions that can be used to discover alternations and variation in articulation, and it is particularly helpful for highlighting fine-grained differences in very similar articulations.

Ultrasound is emitted along a number of arrays along the probe. Each frame contains 416 rows and 69 columns representing these arrays. In this way each frame can be represented as a matrix of brightness values, where each element corresponds to a pixel, and the value of the element is the brightness of that pixel. PCA was performed over the raw ultrasound data frames. The PCs were visualized by showing the loading of the PC on each pixel.

Optimizing signal-to-noise ratio

One problem with this method is that PCA will find variation that can be accounted for anywhere in the ultrasound frame. The bottom of the "fan" contains information about the tongue muscles, but it does not on its own reveal the shape of the tongue. Further, because it is closer to the ultrasound probe, it tends to be more readily visible and brighter and higher-contrast in the ultrasound images. In this experiment, only the tongue shape is of critical importance here. A mask was applied to the data, tailored to each block to account for different depths at which the tongue appears for different phones and speakers. This was done by changing 125 to 200 of the bottom rows of the data to 0.

Ultrasound data tends to contain speckle, as shown in Figure 4.6. A filter was applied to the data to remove speckle and increase the contrast of the tongue. The filter increases contrast by determining the median brightness of a pixel within a certain radius. Speckle tends to be surrounded by darker pixels, so while a particular speckle pixel is by definition bright, the surrounding pixels tend to be dark, so the overall contrast in a region of speckle is dimmed by the filter. The brightness of the line representing the tongue is not random, and a given pixel along the line of the tongue is likely to be surrounded by other bright pixels, so this brightness is enhanced with respect to the speckle. This filtering gives a more insightful reading into how much variation is explained by the different components.

Determining PCs describing tongue movement

Frames for each subject and block were collected as described above. For each subject and block, a separate PCA was performed. There were 90 trials in each block, so each PCA contains 90 frames. The first PC explained between 17 and 23% of the variation in the data, while only the following five PCs explained more than 3% of the variation. This is possibly because the data contained very little patterned variation to begin with: over the course of 90 trials, speakers were only transitioning between two vowels. In some speakers, especially if there was not much variation to explain (i.e. they did not move their tongues very much), the first principal component was correlated not with an acoustic variable, but rather some global variable like exhaustion or probe movement.

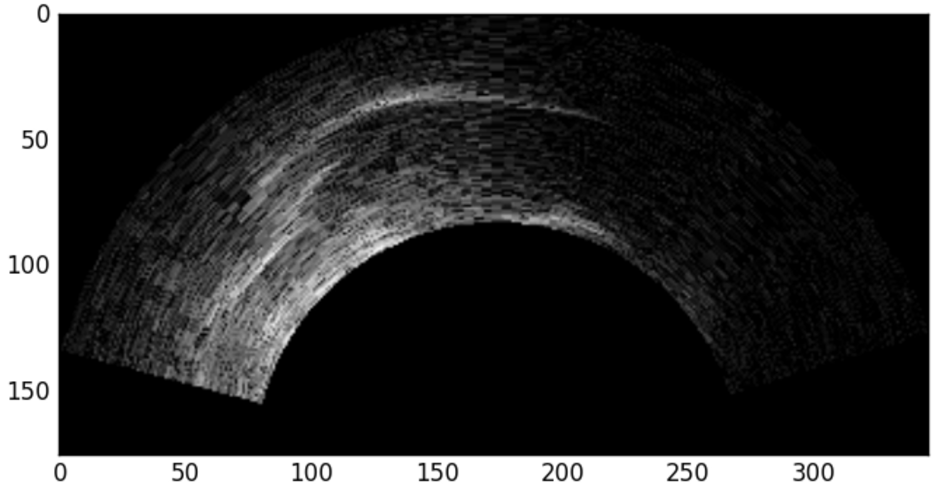


Figure 4.5: Sample ultrasound frame showing mask that was used to control PCA region of interest.

For each subject and block, a linear model was built to determine which principal components described movement that could be linked to differences in acoustics relating the formant under investigation with principal component values for that block.

Principal components whose correlation with an acoustic parameter reached near significance (for some subjects there was no PC that reached significance at the 0.05 level) were examined visually to determine the range of movement that resulted in the acoustic change, as well as to confirm that these PCs were plausible indicators of articulatory response to the perturbation (i.e. that they did not represent a covarying artifact). Figures were produced that showed the loadings of the PC on each pixel. Two samples are shown in Figure 4.8. Brightness or darkness is correlated with loading at the endpoints of an axis of variation described by the PC, so that very bright pixels correlate with strong positive loading on that PC, while dark pixels correlate with a strong negative loading. Thus in Figure 4.8, the first PC describes differences in height of the tongue blade, and the second PC describes differences in anteriority of the entire tongue.

This experiment collected three types of data: articulatory, acoustic, and anatomical. This section reports the results of this data and explores the relationships between them.

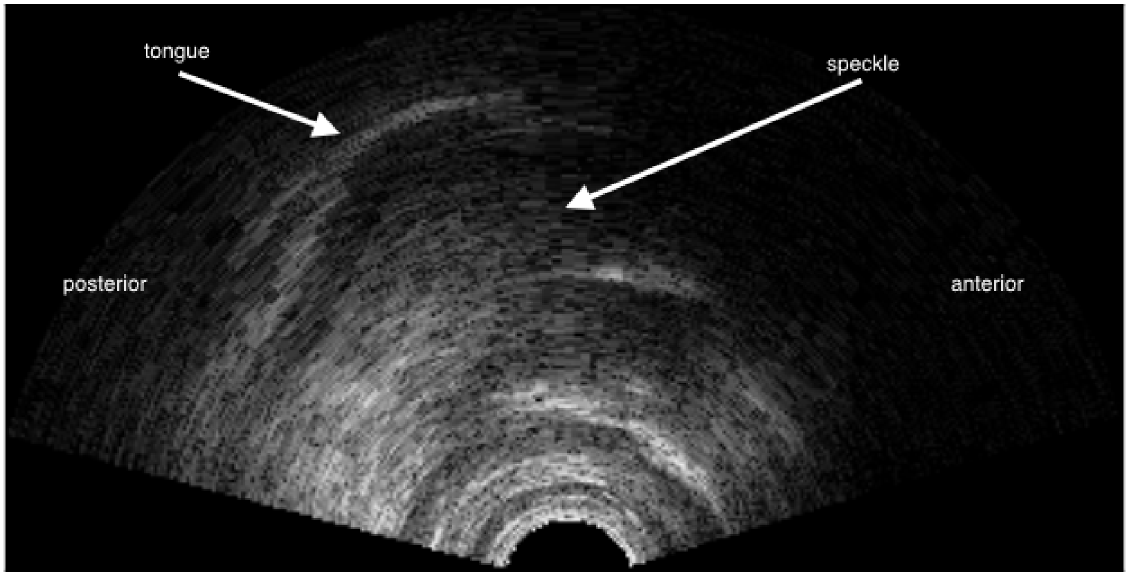
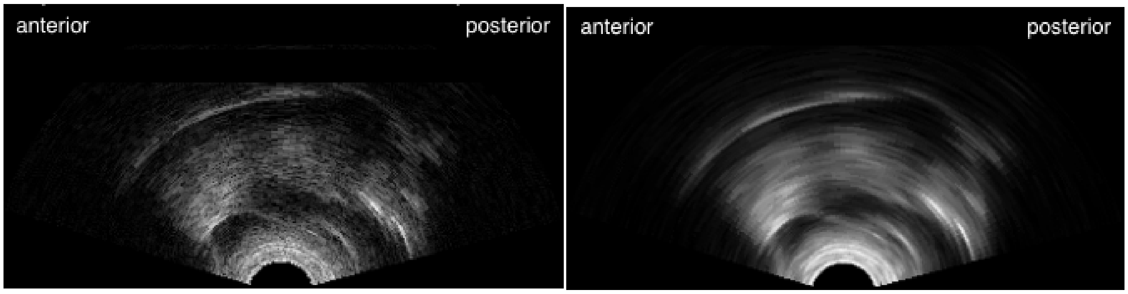


Figure 4.6: Sample ultrasound frame showing speckle.



(a) pre-filtering

(b) post-filtering

Figure 4.7: The same image shown before and after median filtering.

First, it was determined whether each subject actually compensated for the altered feedback or not. If they did, their data was included in subsequent analyses. If not, articulatory data was examined to see if there was any attempt at compensation that was not otherwise reflected in the acoustics.

For subjects who did compensate, further articulatory analysis was performed to determine how articulation was related to compensation. This was done to answer several questions related to compensation: first, how do speakers compensate, and second, how do speakers differ in how they search their articulatory-acoustic space in order to compensate?

Palate data was examined in the context of each of these stages to determine if there is any correlation between palatal doming and performance in the altered feedback task.

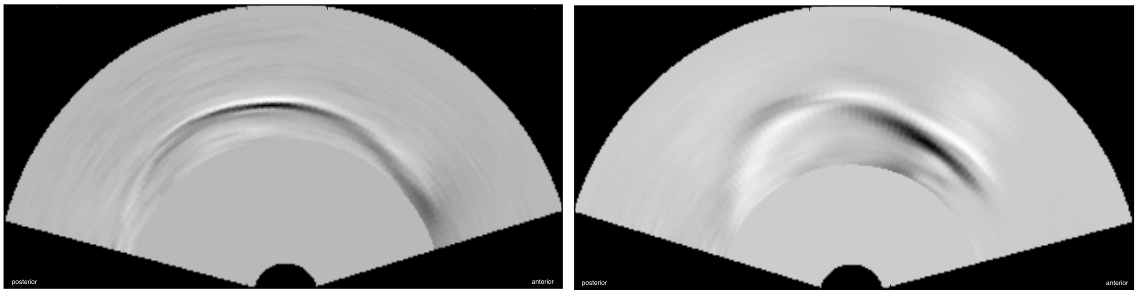


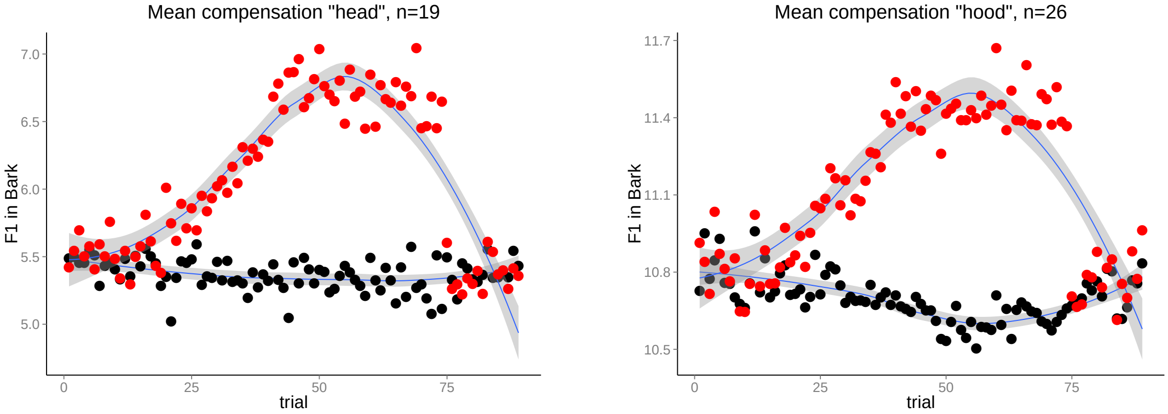
Figure 4.8: Examples of PCs that reveal axes of variation in two subjects.

4.6 Results

Acoustics

Compensation

Of the 30 subjects tested, 19 showed at least 10% compensation for the F1 perturbation, and 26 for the F2 perturbation. The plots in Figure 4.9 show the mean compensation in blocks 1 and 2 for all of the subjects who compensated.

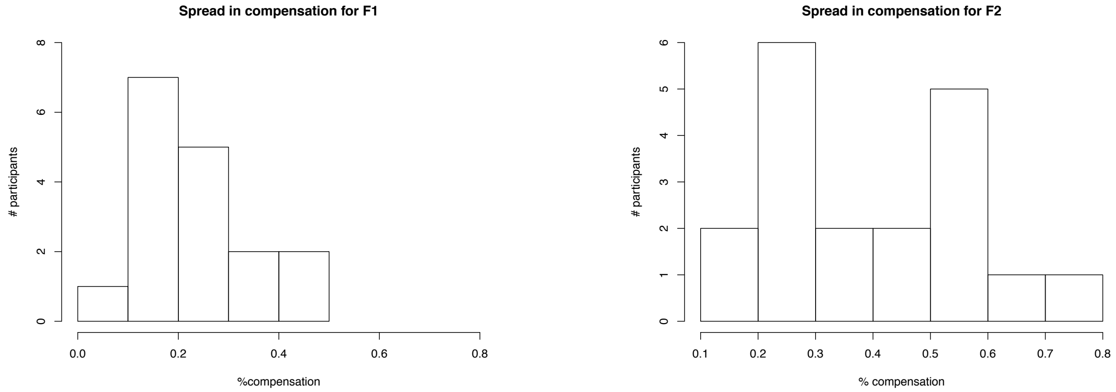


(a) Block 1 (b) Block 2

Figure 4.9: Acoustic results averaged across all speakers who compensated more than 10% in the experiment.

There was greater variability in how much speakers compensated in the F1-altering block than the F2-altering block. Figure 4.10 compares spread of the percentage of the perturbation which speakers compensated most.

Not everyone compensated for the entire block, and the trial where there was greatest compensation (where the produced formant reached a minimum) varied between blocks. In the F1 block, the median trial of greatest compensation was 47 (two trials into the hold phase), and in the F2 block, the median trial was 52, (seven trials into the hold phase). The



(a) “head” block

(b) “hood” block

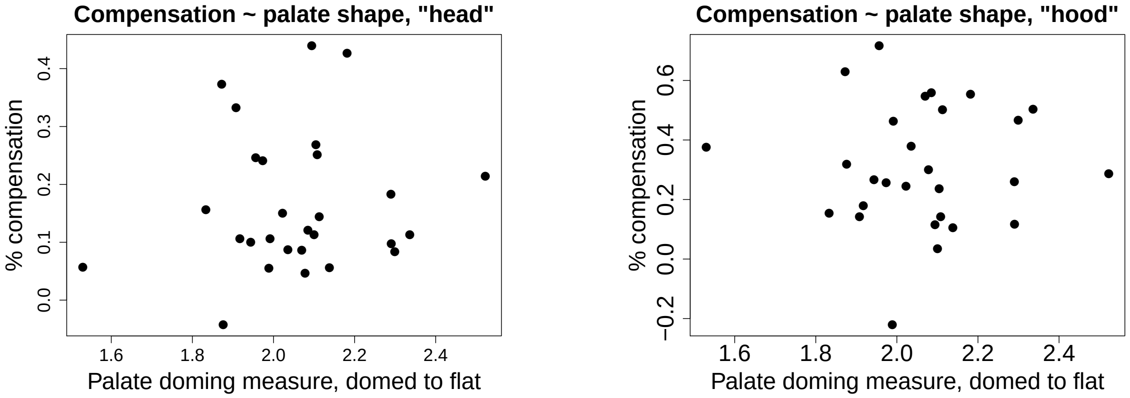
Figure 4.10: Differences in proportion of compensation between the two blocks.

amount of compensation also differed between blocks, and a subject’s compensation in one block did not predict his compensation in the other ($r = 0.14, p = 0.46$)

This experiment hypothesized that the shape of the palate would influence compensation. One hypothesis stated that people with flatter palates would be more complete compensators due to greater flexibility in the acoustic output that could be achieved with less tongue movement. The other stated that palate shape affects articulatory habits, and because people with flatter palates have less articulatory variability, they would have stronger representations of the motor plan required to produce the target. Such people would be less likely to stray from their normal motor plans and therefore less likely to be more complete compensators. Overall, no direct effect of palate shape was found between how much a person compensated and their palate shape. No relationship was found (F1-altering block, $r = 0.09, p = 0.62$; F2-altering block, $r = 0.005, p = 0.98$). Figure 4.11 shows the lack of relationship between compensation and palate shape for all subjects. Negative compensation indicates that a subject followed the perturbation by shifting the formant in the same direction of the perturbation.

Change in production of multiple formants

Some speakers compensate along more dimensions that the single altered formant: during a block that shifts F1, such a speaker might change both their F1 and F2. It was hypothesized that speakers would change the unaltered formant in a specific direction. For the F1-altering block, F1 is shifted up, and speakers compensate by shifting F1 down. It was expected that speakers might also increase their F2 to increase the distance between F1 and F2, but many speakers actually decreased their F2. Similarly, in the F2-altering block, where F2 is also increased, it was expected that speakers would increase their F1 to close the widened space between F1 and F2, but many speakers actually decreased their F1.



(a) Block 1

(b) Block 2

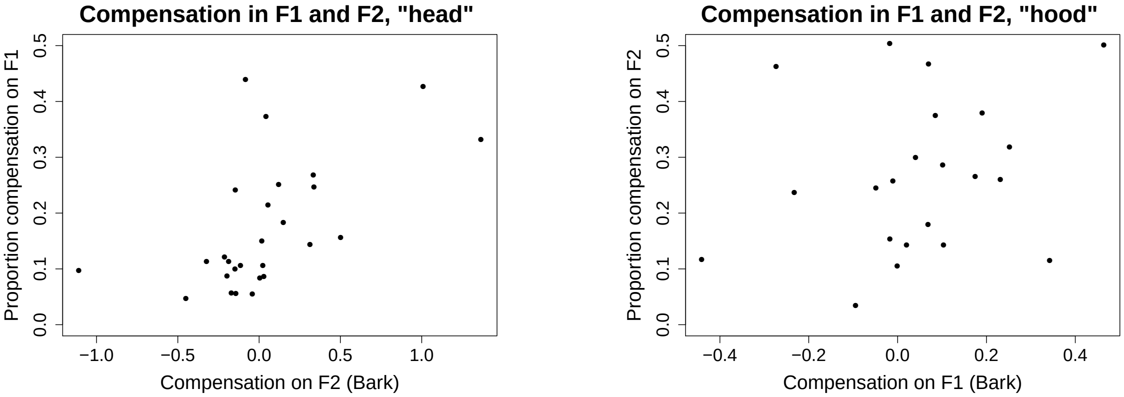
Figure 4.11: There is no relationship between a subject's palate domedness and how much he compensated during the experiment. Data from all subjects is shown here. Negative compensation indicates following the perturbation rather than compensation.

This discrepancy required further investigation to determine whether there was a correlation between compensation success and change on an unaltered formant.

The median formant value of this unaltered formant was calculated for the baseline phase. The value of this secondary formant from the trial corresponding to the greatest compensation (of the primary formant) was also calculated. The difference between the baseline average secondary formant and the value of the secondary formant at this trial of greatest compensation was calculated.

A correlation was found between the *percent* of total compensation on the altered formant and the amount of change produced in the unaltered formant for the F1-altering block ($r = 0.56, p = 0.002$), *but not for the F2-altering block* ($r = 0.002, p = 0.99$). The relationship between the percent compensation on the altered formant and the change in production of unaltered formant in shown in Figure 4.12. A version of the plots with subject numbers for by-subject comparison is given in Figure 4.13.

There are several questions that arise in between-block compensation. One is whether compensation in one block can be predicted by change in production on that formant in the opposite block. That is, if a subject is willing to shift F2 in the F1-altering block, are they "F2-shifters," or vice-versa? There was no relationship found in whether changing F2 in the F1-altering block predicted compensation on F2 ($r = 0.14, p = 0.62$), or whether changing F1 in the F2-altering block predicted the percentage of compensation in the F1-altering block ($r = -0.07, p = 0.79$). Another question that arises is whether change along the second formant in one block predicts change in the second formant in the other block, but no such relationship was found ($r = 0.24, p = 0.21$).



(a) Proportion compensation on F1 against amount change in F2 (Bark) in F1-altering block

(b) Proportion compensation on F2 against amount change in F1 (Bark) in F2-altering block

Figure 4.12: There is a relationship in the 'head' block between compensation on F1 and compensation on F2.

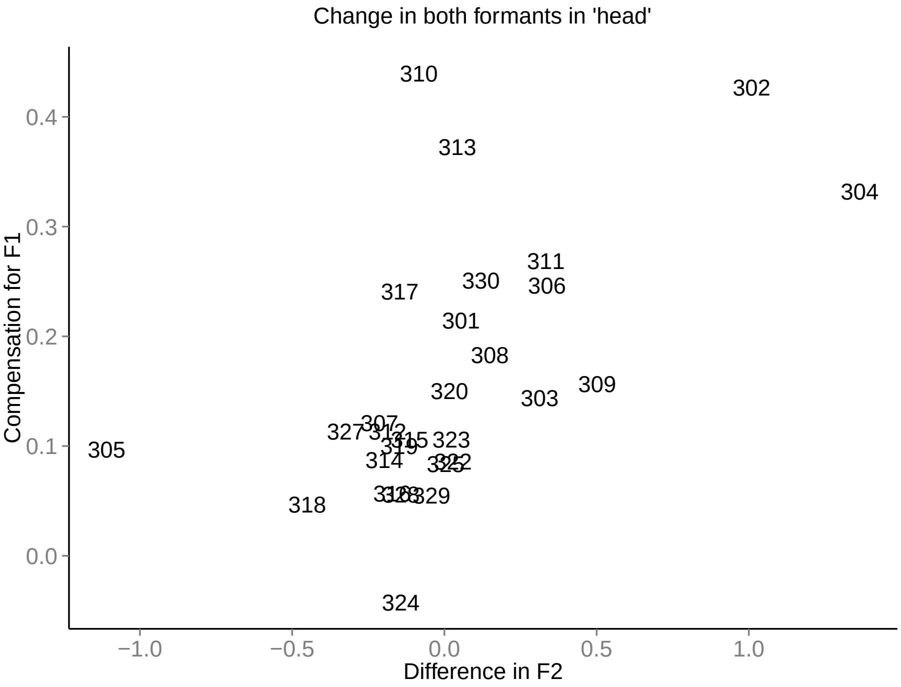
Articulation: quantitative variability

Two subjects who were included in the acoustics analysis were not included in further articulatory analysis because ultrasound data is not available due to technical error.

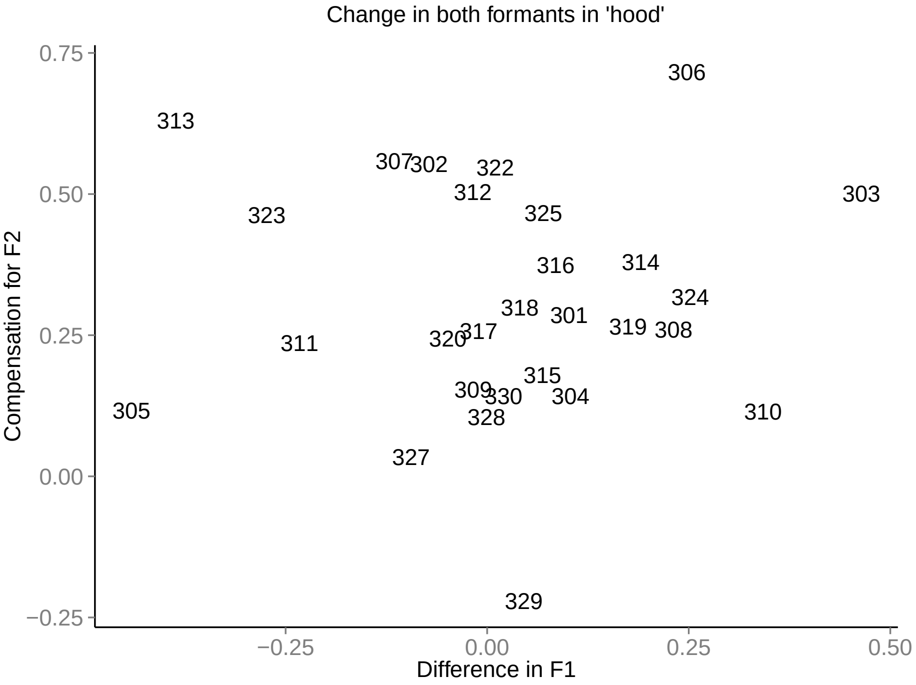
Speakers differ in how much the tongue moves during speech. As shown in the pilot study in Chapter 3, this articulatory-acoustic variability relationship can be explained at least in part by the doming of the hard palate. Speakers with flatter palates had reduced articulatory variability compared to people with more domed palates, likely to maintain acoustic variability. While this effect was significant, there does not exist a metric to measure the *size* of this effect. The analysis in this section considers whether this between-speaker articulatory and acoustic variability is related to a speaker's palate shape.

The articulatory variability measure was taken as described in the analysis section: the difference between the ultrasound frame representing the acoustic greatest compensation and the average baseline frame was taken, and the norm of this matrix (normalized by the norm of the average baseline) was used to quantify how much the tongue had moved. This was compared with the percent of the perturbation that a subject had compensated and the palate shape.

The previous section found no correlation between percent of perturbation compensated for and shape of the palate. There was also no direct correlation between the shape of the palate and how much the tongue moved between the baseline and the trial of greatest compensation ($r = 0.14, p = 0.47$).



(a) Proportion compensation on F1 against amount change in F2 (Bark) in the F1-altering block



(b) Proportion compensation on F2 against amount change in F1 (Bark) in the F2-altering block

Figure 4.13: There is a relationship in the 'head' block between compensation on F1 and compensation on F2. Subject numbers shown to facilitate comparison between blocks.

Interim Discussion: direct correlations

No correlations were found between palate shape and any of the variables measured, and there was only one correlation found between any variables in the data (positive correlation between proportion compensation in F1 and change in F2 in the F1-altering block). A larger analysis was done to determine if groups of variables could reveal different types of speakers in the data, and whether any of these groups was related to palate shape.

Group trends

No correlation was found between palate shape and any single measurement, but there are patterns in group behavior that reveal different kinds of speakers, and palate shape does pattern with some of these. In order to determine these patterns, multidimensional scaling (MDS) was performed over the data. A distance matrix taking into account six variables provided the data for the MDS. The six variables were three parameters taken from each of the two blocks examined so far: percent compensation on the altered formant, change in production of the secondary formant at the trial of greatest compensation, and the amount of change in tongue configuration at the trial of greatest compensation (times two for the second block). The data was reduced to three dimensions. Hierarchical clustering was also performed over the data to help reveal subjects who group together based on differences in their behavior as quantified by the distance matrix.

Dimension 1 was positively correlated with percent compensation on the altered formant in the F1- and F2- altering blocks, as well as how much speakers changed their F2 in the F1-altering block. It was also positively correlated with how much speakers moved their tongues in the F2 block. Dimension 2 was negatively correlated with compensation in the F2 block, positively with change in F2 during the F1-altering block, and negatively with articulatory change in the F1-altering block. Dimension 3 correlates negatively with both how much F1 changed during the F2-altering block and also articulatory change in the F1-altering block.

This can be summarized as follows: speakers registering high on Dimension 1 were strong compensators overall, and they tended to shift their F2 up during the F1 block. They also tended to have a greater overall change in articulation in the F2 block. Speakers registering high on Dimension 2 didn't compensate much during the F2 block, despite having comparatively larger shifts of F2 during the F1 block, and they also didn't change their overall articulation very much in the F1 block. Speakers high on Dimension 3 didn't change their articulations as much as other speakers in the F1 block, and they didn't shift F1 much in the F2 block. None of these dimensions correlated with palate shape ($r = 0.03, p > 0.49$).

Two clustering diagrams are shown here. Figure 4.14 shows the hierarchical cluster with palate shape excluded as a factor, and Figure 4.15 shows the hierarchy when palate shape is included in the distance matrix. Including palate shape as a factor does not have a large effect on how subjects are distinguished from each other; their compensation behavior in the experiment has a larger effect on how subjects are grouped and is independent of palate shape. The analysis shows where palate shape could have any (small) effect.

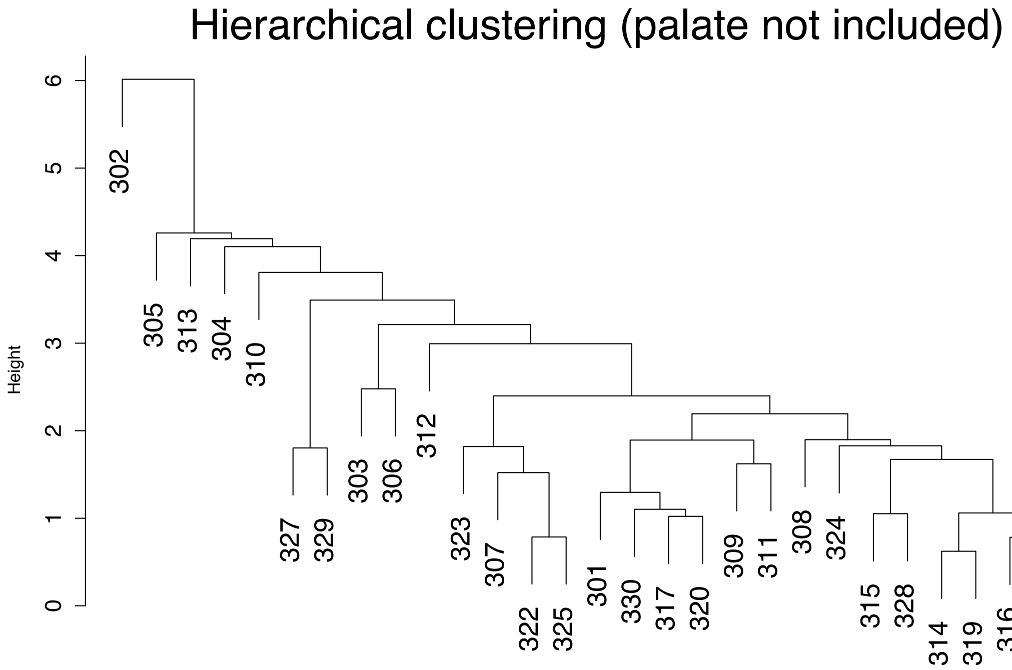


Figure 4.14: Hierarchical cluster showing clusters when palate data is not included as a factor.

A second MDS analysis over the same experiment data but with the addition of palate shape included as a factor revealed relationships between palate shape and behavior. The first dimension loaded exactly as above, again with no relationship to palate shape. Dimension 1 correlates positively with compensation in both blocks, increase in F2 during the F1-altering block, and change in articulation in the F2-altering block. It does not correlate at all with palate shape. Dimension 2 correlates positively with palate shape and articulatory change in the F1-altering block, and negatively with change in F2 in that block. This translates as people who have flatter palates also tended to have greater articulatory change but less change in F2 during the F1-altering block. Dimension 3 correlates positively with palate shape, and negatively with difference in F1 in the F2-altering block as well as change in articulation in the F1-altering block. Finally, Dimension 4 correlates negatively with palate shape and positively with compensation during the F2 block, suggesting that people with more domed palates may compensate more in the F2 block. This is curious because no correlation at all is found between these two factors outside of the MDS analysis ($r = 0.006, p = 0.98$). The correlations between the dimensions and factors are shown in Table

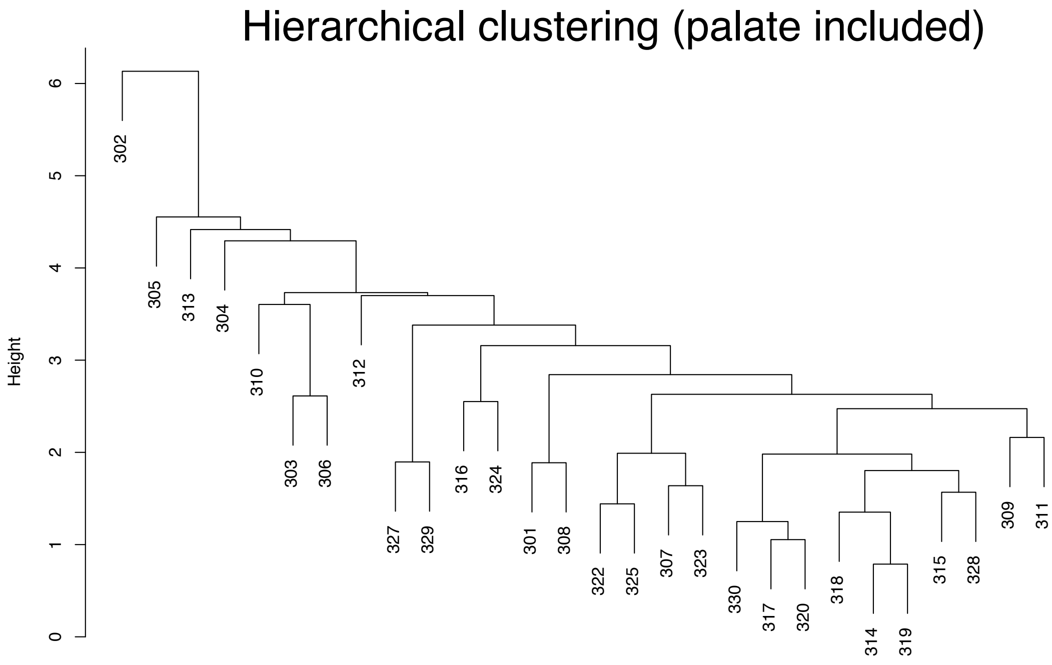


Figure 4.15: Hierarchical cluster showing clusters when palate data is included as a factor.

4.1 compares the two analyses (with and without palate shape included as a factor). The stress test in the MDS analysis showed that four dimensions best represent the data, as opposed to three in the first analysis. Table 4.2 summarizes these relationships in terms of positive or negative correlations.

These dimensions help to identify relationships and tendencies between the variables measured, but considering these dimensions outside of the context of the data can make some of these relationships appear contradictory. What the hierarchical clustering and MDS analysis show is that there are two main groups of speakers: one group is clustered at the center of the MDS dimensions and is particularly poorly described by Dimension 1, but the remaining three dimensions separate this large cluster into smaller groups. The other large group breaks into two groups that fall on the ends of Dimension 1, and they are either the very strong or very weak compensators. The members of this group are also differentiated by the three remaining dimensions.

Dimensions 2 and 3 may seem to contain contradictory relationships, but observing these two dimensions plotted against each other helps to make sense of them (Figure 4.16). No subjects fall into the corners of the graph, i.e. no subject represents more than one endpoint

D	alpha	Comp F1	Comp F2	Diff F2	Diff F1	Art F1	Art F2
1	0.03	**0.83**	**0.51**	**0.70**	0.12	0.24	**0.78**
2	0.03	0.21	**-0.57**	**0.59**	0.20	**-0.64**	-0.20
3	0.03	0.07	0.26	0.05	**-0.84**	**-0.46**	0.13

D	alpha	Comp F1	Comp F2	Diff F2	Diff F1	Art F1	Art F2
1	0.05	**0.78**	**0.58**	**0.70**	0.13	0.23	**0.80**
2	**0.53**	-0.09	0.35	**-0.63**	-0.29	**0.62**	0.16
3	**0.46**	0.18	0.00	0.13	**-0.62**	**-0.60**	0.08
4	**-0.52**	-0.34	**0.66**	-0.08	-0.27	-0.17	0.14

Table 4.1: Table of correlations showing relationships between dimensions from the MDS analyses and the factors measured: alpha value of the palate, compensation in the F1 block, compensation in the F2 block, change in F2 during the F1-altering block, change in F1 during the F2-altering block, magnitude of articulatory shift at trial of greatest compensation in F1- and F2-altering blocks, respectively. The top table reflects the first MDS analysis, where palate was not included as a factor in the analysis (three dimensions), and bottom table reflects the analysis where palate shape is included (four dimensions). Boldface indicates significant correlations.

D	alpha	Comp F1	Comp F2	Diff F2	Diff F1	Art F1	Art F2
1		+	+	+			+
2	+			-		+	
3	+				-	-	
4	-		+				

Table 4.2: Summary table of positive and negative correlations where palate was included as a factor in the analysis (second MDS analysis). Only significant correlations are shown.

of one of these two dimensions. Dimensions 2 and 3 separate out the outlier subjects into two "types" based on palate shape. One cluster, which includes subjects 313, 305, and 312, register high on Dim. 2, while 304 registers low. Subjects 310, 316, and 303 register low on Dimension 3, while 302 and 323 register high. These two dimensions suggest that palate shape can be linked with one of two types of behavior: either it affects how a subject compensates in the F1-altering block, or it affects how a speaker changes production in F1. In other words, Dimension 2 suggests that palate shape is correlated with a tradeoff between articulatory change in F1 and acoustic change in F2 in the compensation of "head," and Dimension 3 suggests that for other speakers, palate shape is correlated with positive relationship between the amount a speaker changes his tongue posture in the F1-altering block and how much he adjusts F1 in the F2-altering block. Neither of these dimensions correlates with how much a speaker actually compensates.

Dimension 4 did not shed very much light on the groupings here. It mainly served to separate several speakers in the reliable compensation group by how much they compensated in the F2 block, and it happened that the stronger compensators had more domed palates.

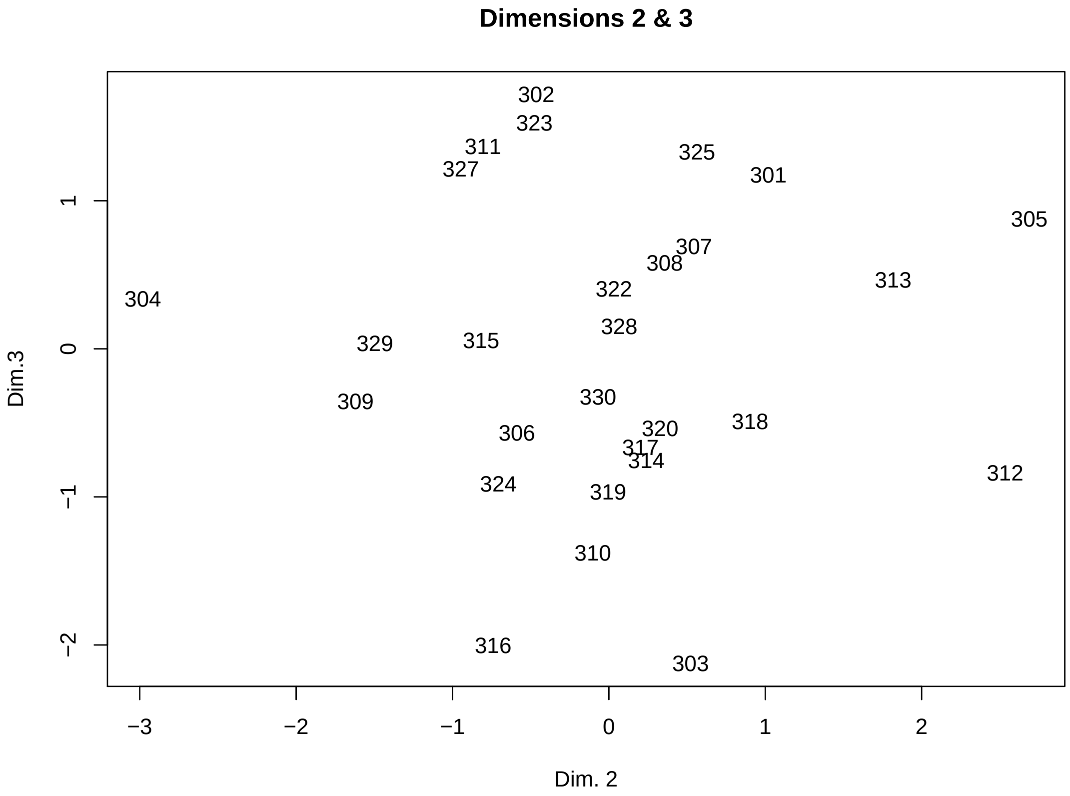


Figure 4.16: Dimensions 2 and 3 plotted against each other. Speakers who register at an extreme on one of these dimensions will not register high on the other.

Only 9 subjects landed near either extreme of this dimension, and two were separated out from the other seven as having flatter palates and less compensation in the F2 block. One of these two subjects had the flattest palate in the group, and the other had a palate that was barely on the flat side of average ($\alpha = 2.09$; mean in this group = 2.06). Although no effect of palate shape was found on the whole group, this effect was found within the group that registered high on Dimension 1 (reliable compensators).

Individual differences in Articulatory strategy

The section above identifies trends in different groups of subjects in this study. Most subjects fit into a large group that behave more or less the same, but several subjects registered strongly on one dimension. There were also two subjects who did not compensate on either block, 327 and 329, where the cluster analysis separated them into their own group for being consistently low compensators. The following section considers individual cases representing different types of compensation.

The goal of the qualitative study is to help explain different types of feedback responses in auditory feedback studies, and it is also an exploratory study to determine what speakers are doing articulatorily during the experiment, information that is not recoverable from acoustics alone. While many speakers compensate, some do not compensate at all. Further, some speakers will compensate on only formant, whereas others will adjust multiple formants.

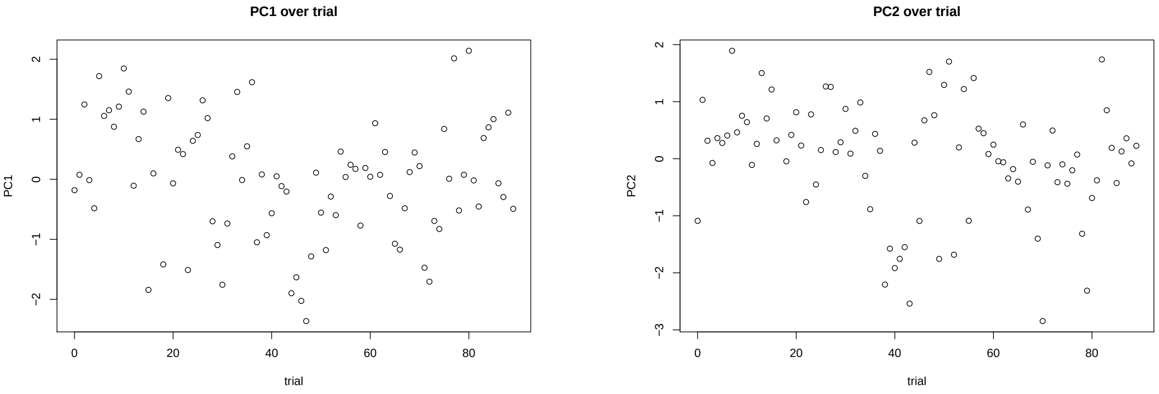
Caveat: Linking PCA and acoustics

As described in section 4.4, PCA was performed over each block on a by-subject basis. Principal components were visually inspected for quality of tongue movement. The components feature differences in brightness or darkness, depending on how strongly positive or negative, respectively, that pixel loads on a given PC. In the case of subjects who did not compensate, PCs were examined to see if there was tongue movement associated with an attempt to compensate. Linear models were built to identify PCs that were correlated with changes in F1 or F2, and further visual analysis was done to determine if those PCs revealed tongue movement. PCs higher than 4 or 5 (depending on subject) explained less than 2% of the variability in the ultrasound data and were therefore not used to explain variability in acoustics. In the figures representing PC loadings that follow, brighter pixels are associated with positive values and darker pixels with negative values.

A word on the relationship between the components found through PCA and acoustics: these components and their relationships with acoustics will vary based on a particular subject's articulatory and acoustic behavior. While it is reasonable to assume that the first couple of components (which by definition explain the majority of the variation in the data) might correlate strongly with the altered and unaltered formants respectively, such an assumption relies on the subject moving the tongue in two distinct ways that are not correlated with each other over the course of the block, and in ways that correspond with trial.

This assumption does not hold when subjects behave in unexpected ways. For example, in the in the F2-altering block, subject 304 had five consecutive trials where he tried a completely different strategy from the rest of the block. This ends up loading on a higher PC (2) because it is a large, reliable variation in the data. It also is strongly correlated with F2. This does not mean that this is the single PC that is strongly correlated with general F2 behavior over the course of the block. The subject tried an articulation distinctly different from what he had tried before, and it happened to cause a large change in F2 in the direction opposite compensation. Visual inspection of PCs prevents such false positives.

We can observe changes in the tongue that occur as a result of trial (Figure 4.17). In the case of 304, the values of PC1 change in a way that is associated with the changes in phase of the experiment: the values cross the 0-line during the ramp phase, and they cross again during washout; i.e. these changes are likely occurring as a result of the altered feedback. Crossing the 0-line indicates that the tongue is moving from an articulation described by one end of the PC axis to the other. However, PC1 is not the component that is most strongly associated with acoustic change, statistically speaking.



(a) PC 1 (b) PC 2

Figure 4.17: Two PCs correlated with F2 change for Subject 304. The PCs were standardized, and their z-scores are plotted on the y-axis.

The component that does correlate strongest with F2 change is PC2. There are only several trials in the middle of the experiment (39-44) where the speaker actually uses the articulatory strategy captured by PC2. These are the greatest changes in acoustics in the experiment, but here it does not follow that PC 2 explains the majority of the F2 shift in the experiment, or the way that this subject typically responded to the perturbation in the way described by PC2. It is possible that the speaker became aware of the experiment, stopped compensating, and then went back to compensating, or had some other reason he did not respond consistently over the block of trials.

A plot showing PC1 by PC2 for the entire block is shown in Figure 4.18. Unaltered and altered trials tend to be separated by PC1, but PC2 separates out the few trials in the middle where the speaker's tongue was more anterior.

This was also motivation to verify selection of PCs by how they change over the course of the trial rather than on acoustics alone: if speakers briefly choose a new articulation that causes a drastic change to the acoustics, this component may appear in a model as more correlated with the acoustic effect and falsely qualify as the component most closely associated with compensation. Therefore, linear models were used as a way to aid in the process of selecting PCs to consider, but any correlation or significance tests can not be used to compare one subject to another.

Compensators

Ultrasound imaging reveals speakers' strategies during compensation. In the F1-raising block, speakers hear "head" altered towards "had," and we expect them to shift their productions towards "hid" to compensate. Based on general relationships between formants and tongue positions, we might expect to see tongue raising in the blade. In the F2-raising

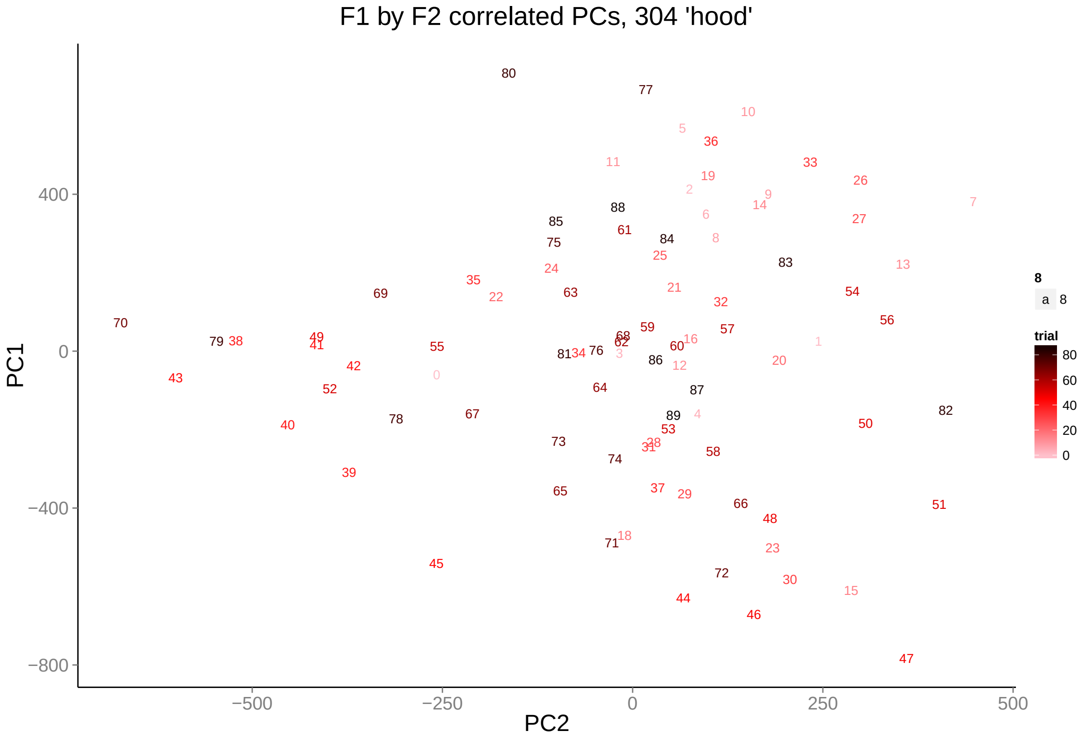

F1 by F2 correlated PCs, 304 'hood'

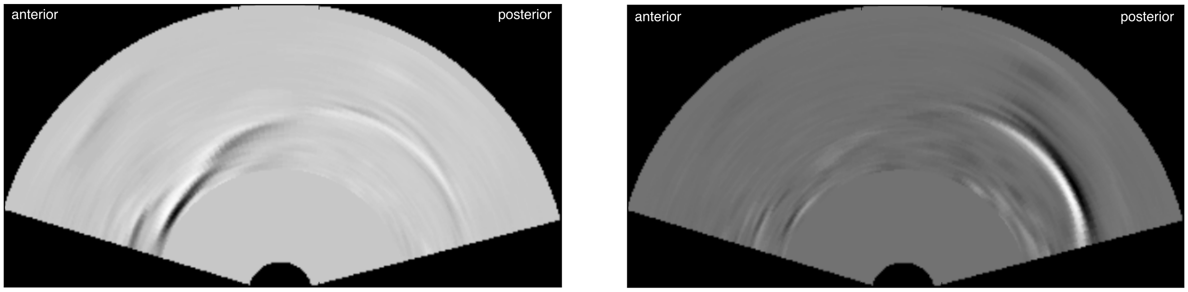
Figure 4.18: PC2 only describes five trials for subject 304.

block, ‘hood’ productions are shifted towards “hud,” and we might expect to see tongue retraction during compensation to lower F2.

Compensation on a single formant versus two formants

While there was a correlation between compensation on one formant and the other in the F1-altering block as described above, some subjects compensated for a large percentage of the perturbation without much shift on F2. For example, subject 310 compensates until trial 69, for a total compensation of 44%. PC1 shows a difference in anteriority of the tongue blade and is associated with change in F2. PC2 shows what seems most clearly to be a difference in root anteriority, but faintly it also shows differences in height of the tongue blade. These PCs are shown in figure 4.19 Figure 4.20 shows how the values of PC1 and PC2 vary over the block. The speaker spends most of the first half of the block with a high PC2, corresponding to a more anterior tongue shape, but ends up in negative PC2 space (more retracted). In contrast, PC1 rises and then falls between the start of the experiment and the trial of greatest compensation.

Subject 310’s F2 at the trial of greatest compensation is not much different from his F2 in the baseline. However, this only indicates his trial where there is maximum compensation, and it does not reveal how he searched his articulatory-acoustic space. Throughout the



(a) PC 1 (b) PC 2

Figure 4.19: PC1 and PC2 for Subject 310 in the F1-altering block.

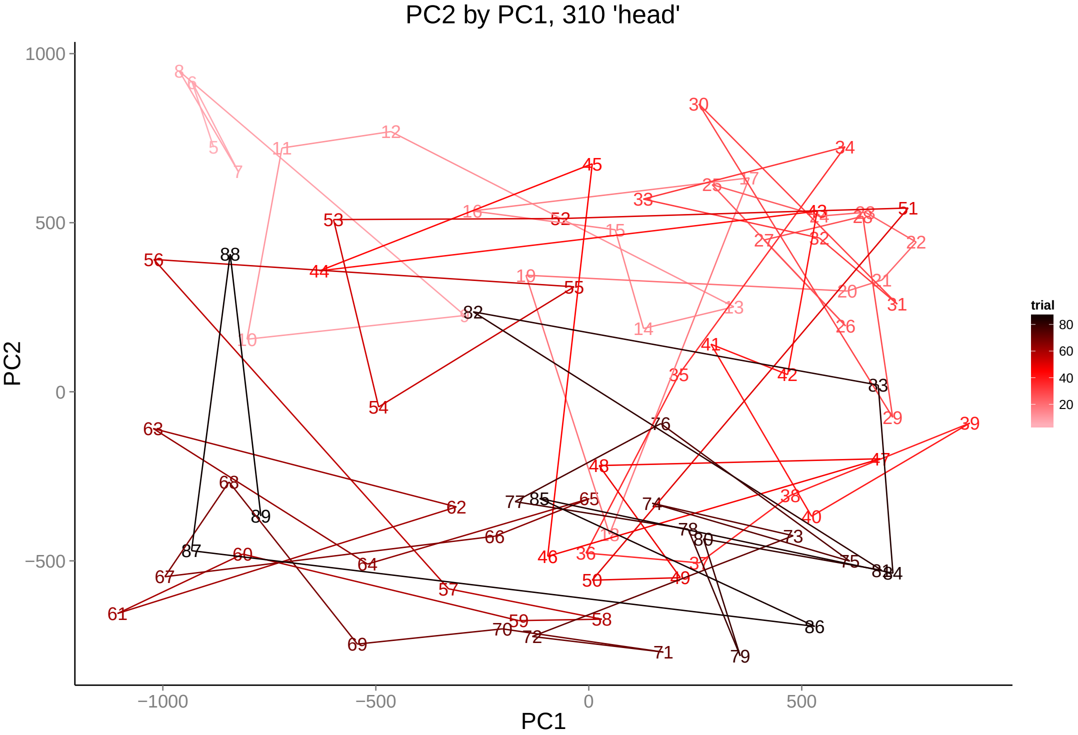


Figure 4.20: PC1 and PC2 for Subject 310 over the entire F1-altering block. The line connects trials in numerical order (with matching color scheme).

experiment, F2 rises and falls. At the trial of greatest compensation on F1, though, the blade is raised. On PC1, which describes posteriority, the value of the PC is -0.5 standard deviations from the mean, suggesting some anterior movement but not full advancement in comparison with the full range of movement throughout the block.

The plots below show how F2 changes throughout the block in correspondence with PC1. This shows how PC1 is likely the only tongue motion responsible for change in F2 until the middle of the ramp phase. Further investigation suggests that the speaker tries a new strategy that targets the tongue root, shown in PC3. The correlation of F2 with each PC1 and PC3 is shown in Figure 4.22.

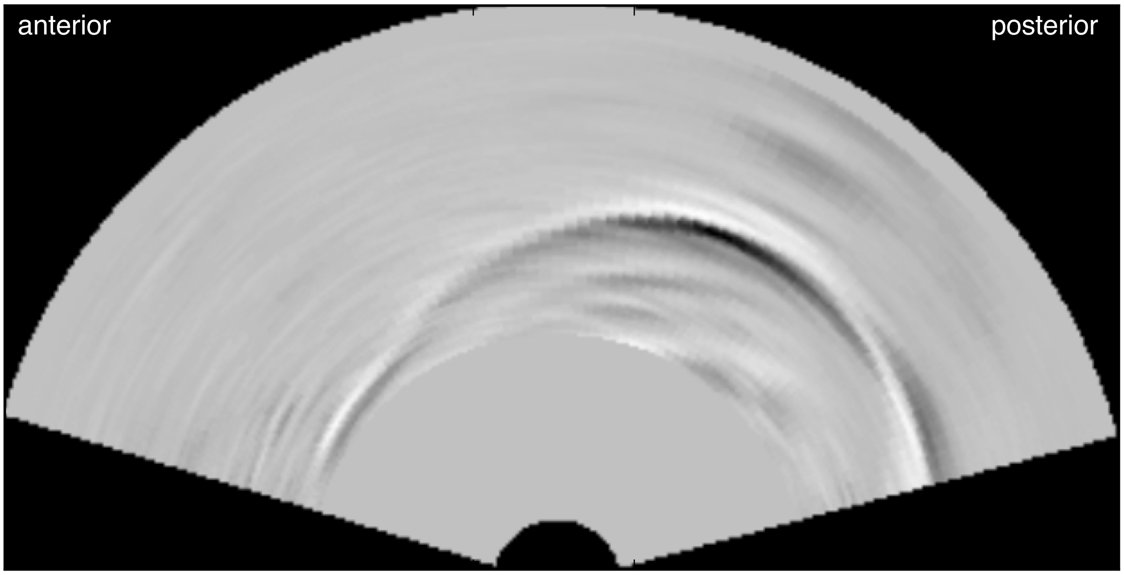
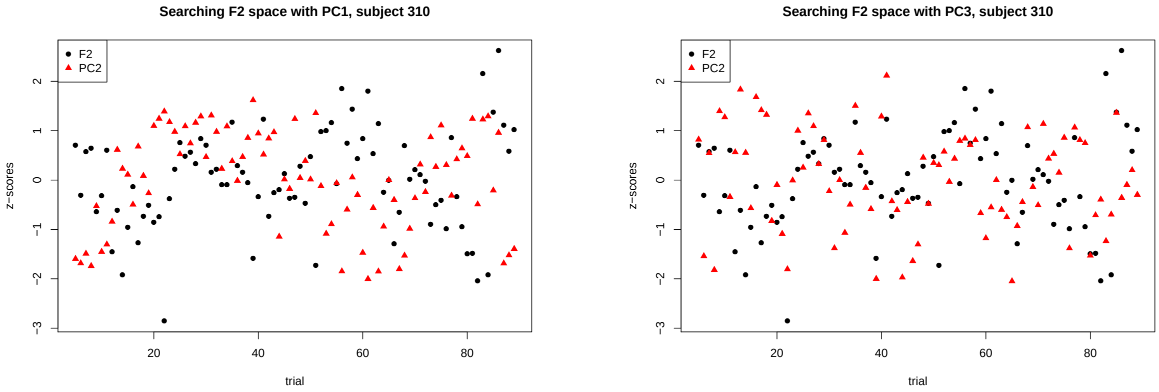


Figure 4.21: PC3, showing additional strategy that Subject 310 uses to adjust F2 during the F1-altering block.

Although Subject 310 did not ultimately change F2 during the trial of maximum compensation in F1, he tried two different strategies, captured by two separate PCs, to adjust F2. One of these PCs captures anterior-posterior movement of the tongue blade, and the other of movement in the entire tongue. The ultrasound reveals not only that he tried changing F2 twice, but that he used two different articulatory strategies to do so, indicating that he was probably searching his articulatory-acoustic space for the compensation that maximized the perception of “head,” ultimately deciding to shift F1 only.

One outstanding question regarding compensation is why some subjects change the production of two formants, even if only one is altered. The investigation above indicates that speakers may have access to multiple articulatory strategies for compensation that involve one more or more formants.

One hypothesis is that changing the production of the secondary formant maintains the distance or ratio of F1 and F2, which might be a more important perceptual cue to vowel category than a particular value for a given formant Nearey (1989). Further, it is possible



(a) PC 1 (b) PC 3

Figure 4.22: Two PCs correlated with F2 change. In each graph, the principal component and F2 are plotted on the same axes, so both were standardized. Their standardizations (z-scores) are plotted here.

that such a tongue posture change would be a smaller overall displacement than the movement required to compensate fully along one formant. If the same *perceptual* compensation can be achieved with a smaller deviation in tongue position and posture, speakers might choose this strategy to minimize somatosensory mismatch. Another explanation with the same effect is that subjects may reach a somatosensory mismatch limit in the primary formant's articulatory space, and so at higher levels of perturbation they might begin to change production of the secondary formant in order to enhance compensation. If this perceptual and/or somatosensory feedback hypothesis is correct, we would expect that subjects would *lower* F1 and *raise* F2 in response to the artificially raised F1 in the altered feedback.

Although he did not change F2 much during the F1-altering block, Subject 310 did shift F1 up by 0.34 Bark during the F2-altering block. His PC 1 shows change in anteriority at the root and differences in blade height, showing that he did change both formants simultaneously.

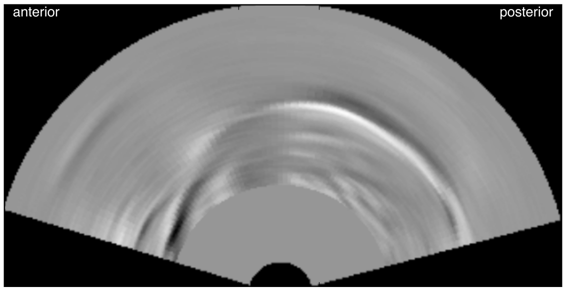
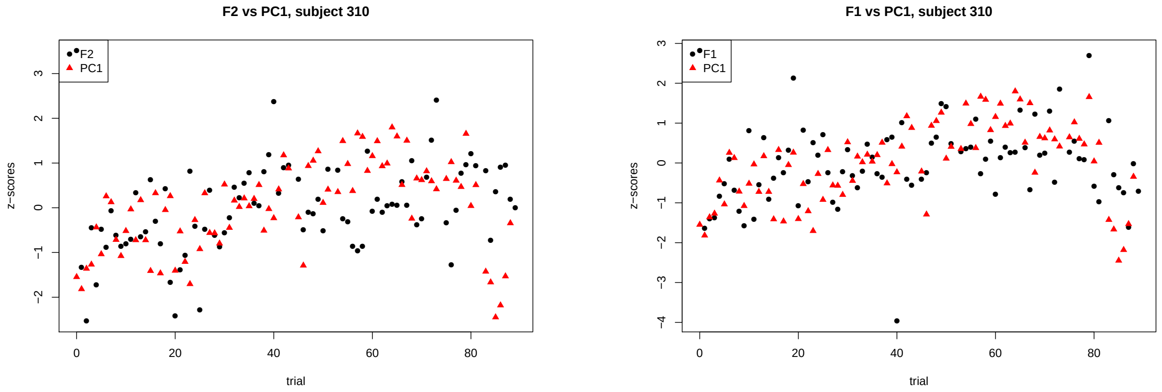


Figure 4.23: Subject 310's PC1 shows movement both in anteriority of the tongue root and height of the tongue blade.



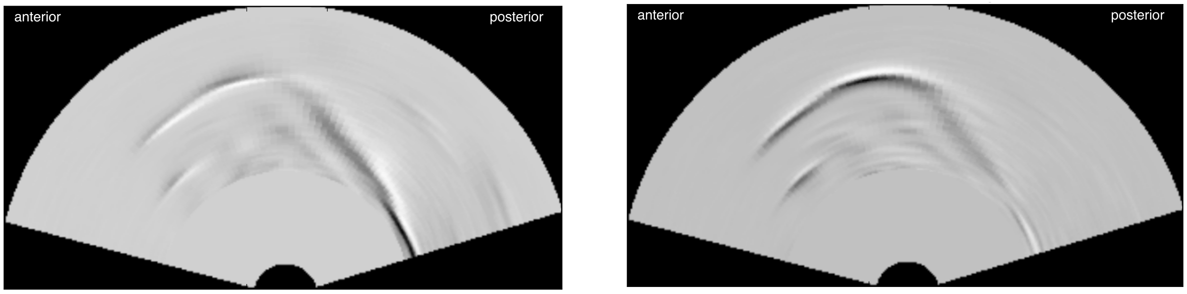
(a) Subject 310, F2 and PC1, F2-altering block (b) Subject 310, F1 and PC1, F2-altering block

Figure 4.24: PC1 correlates both with F1 and F2 change in "hood" block for Subject 310. In each graph, PC1 and a formant are plotted on the same axes, so both were standardized. Their standardizations (z-scores) are plotted here.

Subject 302 was one of only two who compensated significantly in both the F1 and F2 blocks (43% and 55%, respectively). The other subject was 313, who compensated for 37% and 63% of these shifts, respectively. Both of these subjects also raised their F2 as expected in the F1-altering block.

Subject 302 compensated for 43% of the perturbation and increased F2 by 1 bark. Unfortunately, the imaging was very faint, and none of the PCs correlates particularly well with acoustics. There is enough visible to infer parts of the articulatory strategy PC1 probably correlates with an overall probe drift or exhaustion. PC2 does not correlate well with either F1 or F2, but it rises during altered trials and falls in unaltered trials, indicating it probably changes as a result of the perturbation. PC2 seems to correlate with blade raising and possibly simultaneous anterior movement of the tongue root. These actions would respectively raise F1 and F2.

Subject 313 reached his maximum compensation early on in the ramp phase: trials 17 and 22 respectively in the F1 and F2-altering blocks, but he does not give up after this point; other subjects reach their maximum compensation and then return to their baseline articulations. PCs 1 and 2 respectively show tongue anteriority, correlated with F2 change, and tongue height, correlated with F1 change. These are shown in Figure 4.25. Figure 4.26 shows how PC1 and PC2 change over the course of the block (darker colors are later in the block). Even though the minimum value of F1 was reached earlier on in the block, the strategy changes over time. PC2 increases until the trial of greatest compensation, showing that the tongue body raises until that point and then lowers again. PC1 oscillates between positive and negative (anterior and posterior, respectively) over the course of the block, finishing posterior. I interpret this as the speaking finding a strategy (increasing the constriction by raising the tongue body) and reaching a limit, so he tries adjusting the anteriority of the tongue body to adjust F2, increasing the distance between F1 and F2.



(a) PC 1

(b) PC 2

Figure 4.25: PCs correlated with F2 change (PC1) and F1 change (PC2) for subject 313 in the F1-altering block.

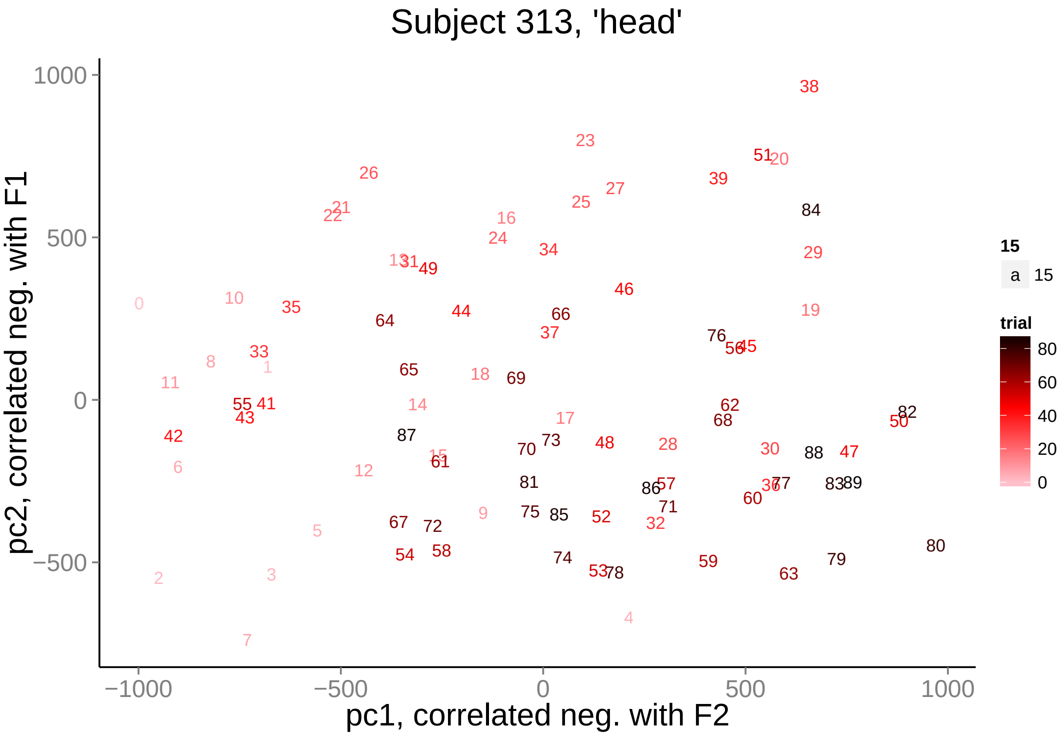
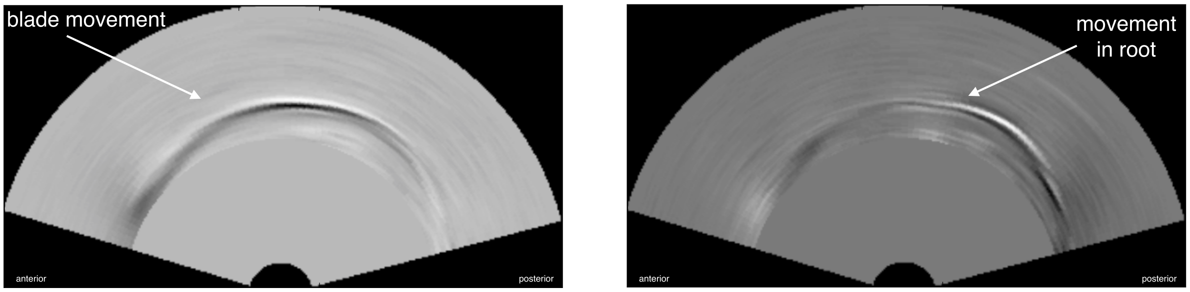


Figure 4.26: Subject 313's PC1 and PC2, "head"

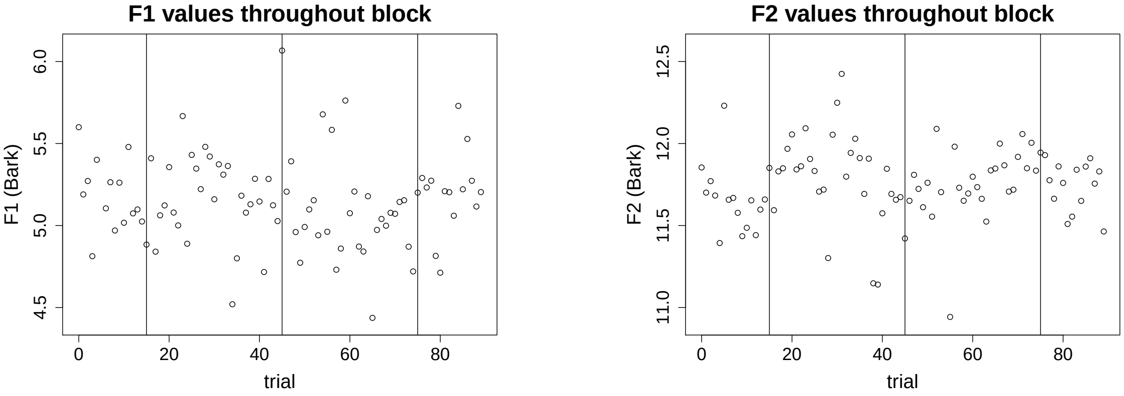
Consistent low compensation

Subject 305 compensated about 10% for both the F1- altering block and the F2- altering block. For the F1 block, subject 305 has two components that clearly account for change in articulation over the course of the experiment. The first accounts for tongue height and frontness, and the second for movement at the front of the tongue (blade raising).



(a) Blade movement (b) Root movement

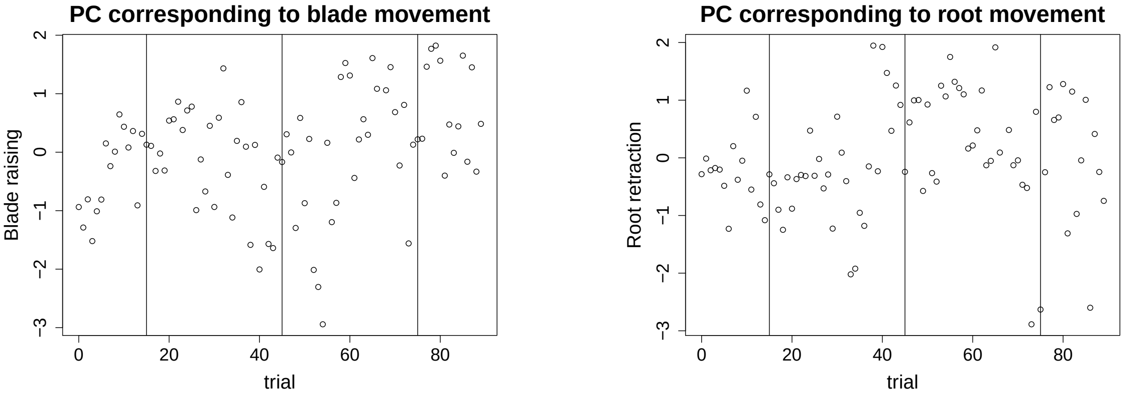
Figure 4.27: PC loadings for two components revealing directions of movement for Subject 305 in the F1-altering block.



(a) F1 change over trial (b) F2 change over trial

Figure 4.28: Changes in 305's *produced* formant values over the F1-altering block.

The tongue blade movement is encoded in PC 2, but there is also clear tongue root movement PC 4. PC 4 also appears to contain some blade raising that is simultaneous with root retraction. Compensation occurs between trials 30 and 60. During this time, the tongue root retracts; both F1 and F2 lower, which is surprising because F1 is not typically associated with blade lowering. It is possible that the faint blade raising in PC 4 is actually responsible for the F1 lowering, or that for this speaker retracting the tongue also results in F1 lowering. The tongue blade lowers until trial 40, but it rises again until 60. At about trial 45 (beginning of the hold phase), the speaker decreases the size of the constriction in order to further lower F1, possibly when he reaches a limit in tongue retraction. While the articulatory movement described by PC 2 in Figure 4.29a changes direction at trial 45, there is consistency in the direction of change described by PC 4 in Figure 4.29b. The speaker *changes* an articulatory component partway through to aid in the compensation.



(a) Blade movement over block

(b) Root movement over block

Figure 4.29: Changes in 305's tongue movements over the F1-altering block.

The speaker independently controls root retraction and blade raising/lowering, and that the strategies chosen are effective in compensation, but the speaker does not compensate very much for the F1 perturbation. However, of all the subjects he has the greatest negative shift in F2 during the F1-altering block. Although he compensated at the low end for the F1 perturbation, perceptually he may have compensated for much more of the perturbation than it appears here.

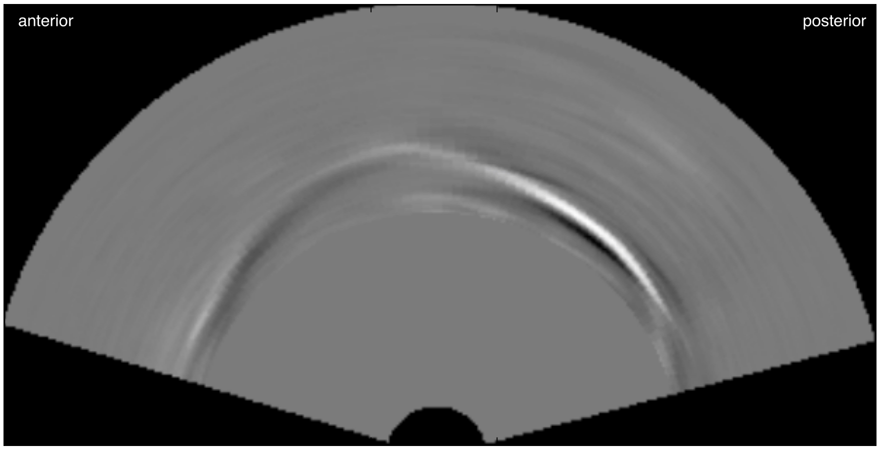
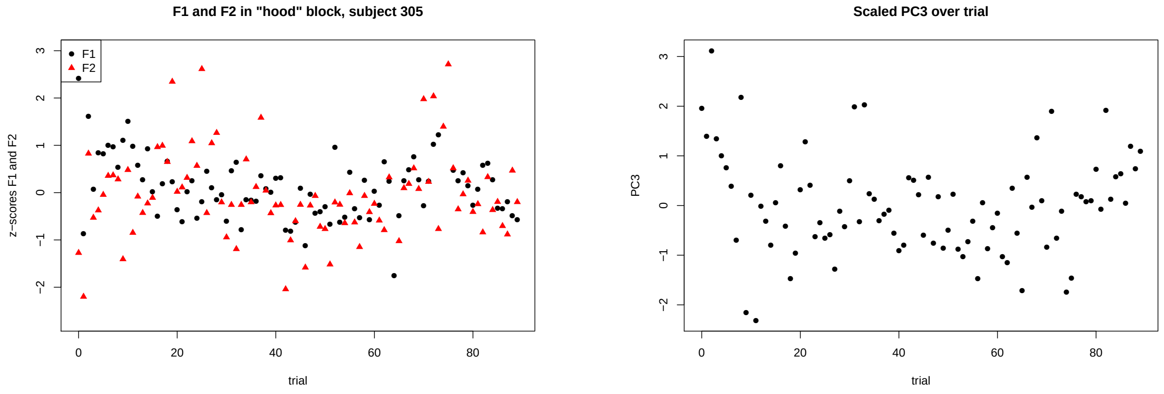


Figure 4.30: PC 2 representing articulatory strategies in the F2-altering block for subject 305.

In the F2-altering block, the same speaker also compensates for only 11% of the perturbation, and again, he does not stop compensating once he reaches this point. As in the other block, this speaker changes his production of the unaltered formant more than any other speaker. Against predictions, the speaker shifts F1 down. If he were trying to reduce the distance between F1 and F2, he would have shifted F1 up. Lip rounding (as in the vowel in the target word "hood") lowers all formants, so it is possible that by lowering F1 the speaker is increasing his perception of lip rounding, and that this gives an increased perception of /ʊ/ than the expected F1/F2 ratio.

PC3 describes the main strategy of raising the tongue body to decrease F1 and retracting the tongue body to decrease F2. Scaled F1 and F2 are plotted together in Figure 4.31a, and Figure 4.31 shows how PC3 changes over the block.



(a) F1 and F2

(b) pc3

Figure 4.31: Two PCs representing articulatory strategies in the F2-altering block. In 4.31a, the formant values were standardized so that they could be plotted together on the same axes without any distortion due to scaling.

Figure 4.31 shows (a) how F1 and F2 are adjusted simultaneously and (b) the values of PC3 over the course of the block. This speaker changes F1 and F2 together as part of the same articulatory strategy. Unlike other speakers who shifted F1 and F2 independently, 305 shifts them together throughout the perturbation period. This was the only speaker who showed this type of compensation pattern.

Non-compensation

Of the nine subjects who compensated 10% or less of the perturbation to F1, six changed their production in F2, and all of these six lowered F2. In the F2-altering block, only two subjects compensated 10% or less, one subject (327, displayed below) did change production in F1 (the other actually followed the perturbation and did not change F1).

Although many subjects did not compensate for more than 10% of the perturbation in any direction (9 in the F1-altering block and 1 in the F2-altering block), their tongues sometimes showed movement that appears to be in response to trial, indicating that these speakers might have been making an effort to compensate.

Subject 327 compensated the least overall, 11% for F1 and 3% for F2. This lack of compensation (Figure 4.32) is particularly striking because subjects overall tended to compensate for F2 more than they did for F1. His articulatory behavior in the F2 block is shown in Figure 4.33. This speaker did not alter F1, and only PC1 was found to correspond with acoustic change (i.e. only in F2), as shown in Figure 4.33a. The speaker is clearly changing the anteriority of his tongue, he just does not have much acoustic change. Any acoustic change is tightly correlated with PC1. F2 and PC1 are scaled and plotted together over trial in Figure 4.33b.

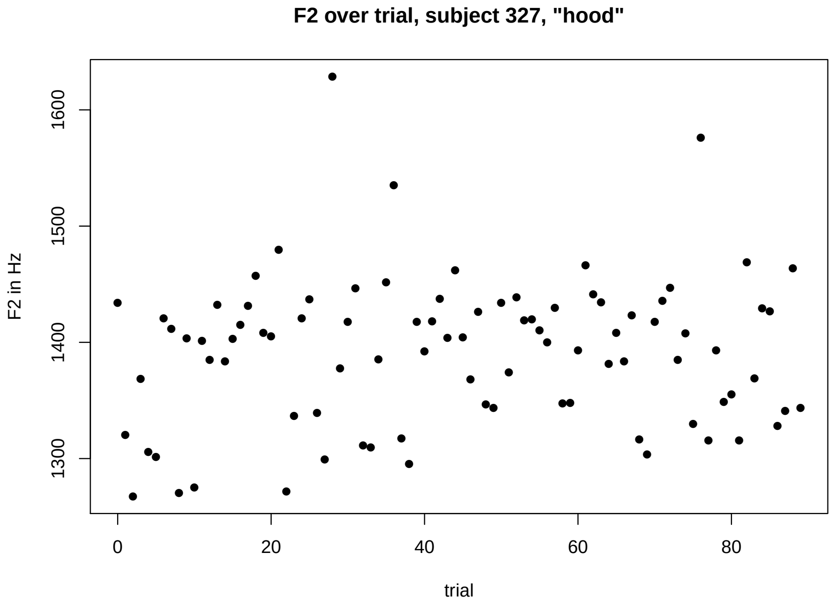
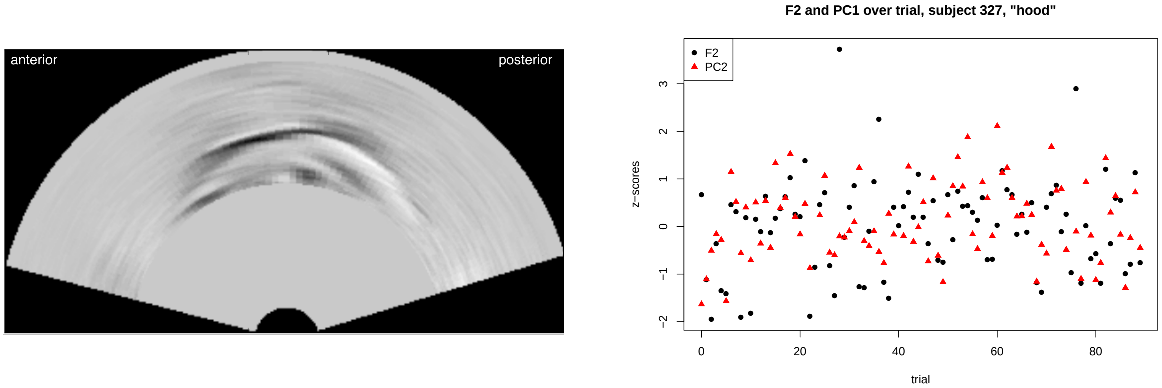


Figure 4.32: Subject 327 showed the least compensation in the F2-altering block.

Other subjects compensated on F2 *but not F1*, lowering F2 in response to the perturbation. This is in contrast with subjects who compensated on both F2 *and* F1. Subject 318 compensated the least for the F1 perturbation at 4%, but he lowered F2 by 0.45 Bark. (No subjects changed F1 only in the F2-altering block.) Despite the lack of compensation, PC1 still correlates with change in F1, and PC2 with change in F2. The loadings for these PCs are shown in Figure 4.34. The change here is subtle, but the graphs showing movement between correlated PCs and formant frequencies over trial shows that there is patterned change in F2. (*Note*: PCs were multiplied by -1 for visualization purposes in the plot in Figure 4.35. The sign itself is not meaningful.) It is clear that this speaker was adjusting F1 by raising and lowering the tongue blade, but it is possible that he found it easier to control F2.

4.7 Discussion

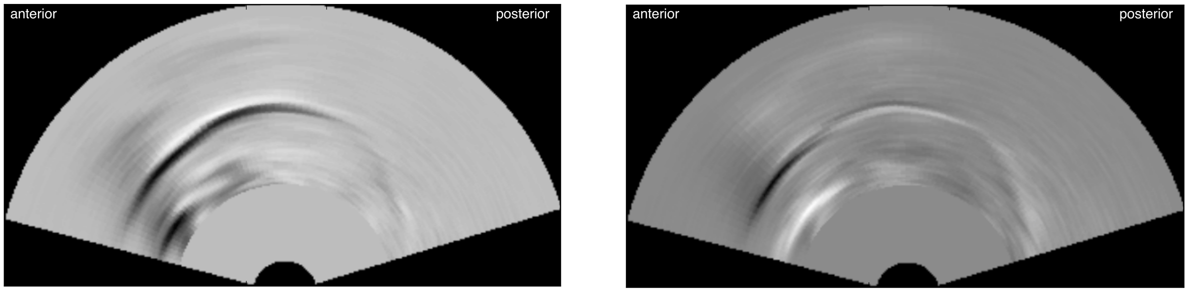
This experiment was conducted to answer several outstanding questions concerning how speakers compensate for altered auditory feedback. Chapters 2 and 3 showed that differences in the domedness of the palate influence articulatory and acoustic variability in normal



(a) Subject 327 F2-altering block, PC1

(b) Subject 327, relationship of PC1 to F2 over F2-altering block

Figure 4.33: Subject 327 only compensates for 3% of the perturbation in the 'hood' block, but ultrasound reveals that his tongue moved in response to the auditory perturbation. In 4.33b, F2 and PC1 were standardized, and their z-scores are plotted.



(a) Subject 318, PC1

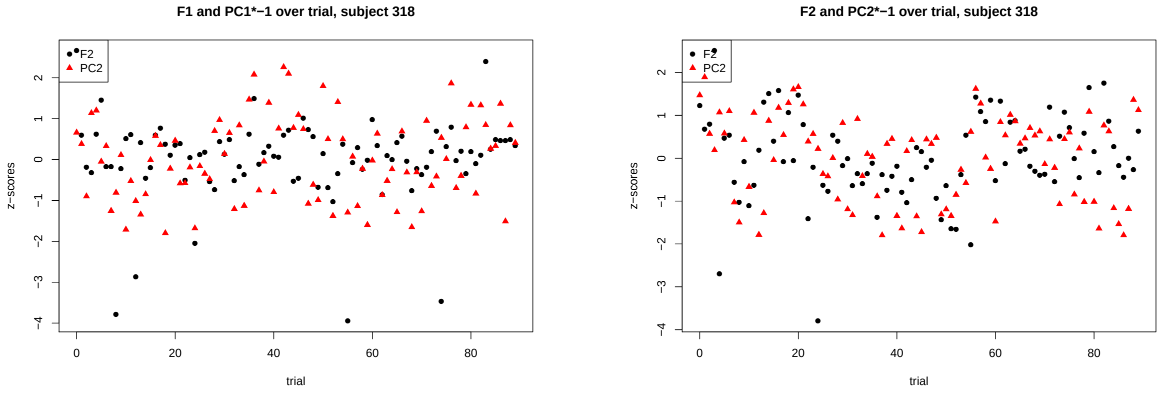
(b) Subject 318, PC2

Figure 4.34: Two PCs for subject 318, who changed F2 but not F1 in the F1-altering block.

speech, and this variable was investigated as a possible factor in how speakers respond to altered auditory feedback. The experiment used ultrasound imaging to expose the strategies speakers use during compensation and to determine any relationships between articulation and acoustics during compensation.

Differences in compensation behavior

In each block, only a single formant was perturbed, but the acoustic dimensions along which compensation occurs can vary across speakers. There was a significant difference between F1- and F2-altering blocks in how much speakers changed their production of the opposite formant. In the F1-altering block, speakers who compensated for a greater percentage of the perturbation in F1 were much more likely to also change the production of their F2, but in the F2-altering block, there was no relationship between how much speakers compensated



(a) Subject 318, F1 and negative PC1

(b) Subject 318, F2 and negative PC2

Figure 4.35: Subject 318 changed F2 but not F1 in the F1-altering block. In 4.35a, PC1 and F1 are plotted together, and in 4.35b, PC2 and F2 are plotted together.

on the altered and unaltered formants.

Simply looking at acoustic averages and percentages of perturbation that speakers compensate for does not explain how speakers compensate, and looking at acoustics alone obscures how speakers are responding to the task and what cues they are sensitive to. For example, subject 310 was clearly not insensitive to F2 feedback: he searched his F2 space and changed his tongue anteriority throughout the block, and the ultrasound supports this analysis. At the greatest compensation he did not have much difference in F2, but he clearly considered as an option for compensation. There were many such differences in articulatory behavior, but this only served to show that compensation could be achieved in different ways, and it did not explain why. The following discussion proposes possible reasons for differences in behavior (and lack of strong correlations with palate shape) as well as a few ways to test these possibilities.

Non-compensators

In the F1-altering block, three subjects compensated for F2 but not F1. In the F2-altering block, there were no subjects who changed their production of F1 in response to the perturbation but not F2. Subjects overall were more consistent in their compensation during the F2-altering block than in the F1-altering block. It is not possible to form a confident generalization to explain this asymmetry, but some possible theories are discussed below, taken together with the asymmetry in compensation variability between the two blocks.

Several subjects in the F1-altering block changed their production of F2 in response to trial *but not their F1*. The articulation for all three of these subjects shows differences in the tongue retraction, which is typically associated with F2. It does not appear that these subjects were attempting an articulation that they thought would alter F1. It is unclear why they would lower F2: if F1 was being raised, lowering F2 would increase the perception of a non-/ɛ/ vowel. In this block, F1 is raised, shifting the vowel towards /æ/. /ɛ/ is fronter and higher than /æ/, so if only F2 were to be adjusted, it is expected that subjects would *raise* their F2.

An alternate explanation is that some subjects in this experimental paradigm wish to maintain the distance between spectral peaks. This would explain compensating along both F1 and F2, but not on F2 alone. Changing production of both F1 and F2 would satisfy this goal if they are shifted together. If F2 is shifted alone, the subject actually hears F1 and F2 closer together, since F1 is being raised as F2 is being lowered, which would result in the perception of an even lower and backer vowel, at the opposite end of the vowel space.

Differences in F1/F2 adjustment

One question that has arisen from previous altered feedback experiments that I hoped to answer here is how and why speakers might change their production of a second, unaltered formant. Do speakers do this on purpose to maintain a ratio of F1 to F2, or are some speakers unable to change their articulation so that only one formant is altered? The results from this experiment showed that people who compensated more in the F1-altering block were highly likely to also change their production of F2 in that block, but the opposite was not at all found in the F2-altering block. There are many possible reasons for this. One is that speakers are just better at isolating change in F2 than in F1, which would account for the greater change in F2 in both blocks.

There seem to be fundamental differences in compensation in the F1-altering block and the F2-altering block. One reason for this might be that the maximum shift in both blocks was 240 Hz. In F1 space, the magnitude of this shift was between 1.76 and 1.93 Bark, and in F2 space, the magnitude was between 0.94 and 1.23 Bark. The *percentage* of a shift (i.e., not necessarily magnitude) that subjects will compensate for is negatively correlated with the size of the shift (Katseff, 2010). This is possibly because if the shift becomes too noticeable, speakers may stop attributing the mismatch between expected and observed feedback to their own speech production system; speakers may also begin to experience too great a somatosensory mismatch and be unwilling to stray too far from their typical motor plan. Thus a more likely alternative explanation for the discrepancy between F1- and F2-altering blocks is related to the fact that the F1 block shift was auditorily greater than the F2 block shift, as described above. It is possible that producing a change in the second, unaltered formant helps to increase the perception of the target vowel under extreme alteration. This finds evidence in the fact that more than one speaker searched his F2 space during the F1 block, as if searching for the optimum configuration.

The data from Subject 305 adds to this discussion. He consistently compensated for about 10% of the perturbation to the altered formant, but he strongly adjusted both F1 and F2 in both blocks. His consistency suggests that he may represent a different type of speaker for whom adjusting two formants is somehow preferable, possibly because it produces the desired perceptual effect better for that type of speaker, or because vowel perception is fundamentally different for this type of speaker.

Another possibility is related to individual differences in how F1 is controlled. As discussed in the background chapter, Johnson et al. (1993) found that speakers differ from one another in whether they are "tongue-movers" or "jaw-movers," or whether they adjust tongue height by moving their tongues independently, or by moving the jaw and allowing the tongue to travel as a passenger. Speakers who are primarily jaw-movers may have been inhibited by the ultrasound probe, or else may have noticed their upward jaw movement more because of the presence of the probe. Similarly, moving the jaw is a larger movement that is possibly more noticeable in general and thus more likely to be an obvious somatosensory mismatch. Jaw movement would only have affected F1 responses.

Speakers seem to have been more willing overall to modulate their F2 than F1: in the F1-altering block, greater compensation on F1 was correlated with more change in F2, but in the F2-altering block, greater compensation on F2 was not correlated with greater change in F1 in that block. At the other end of the spectrum, speakers who did not compensate at all for F1 did sometimes raise F2, which would result in a greater difference between F2 and F1 thereby helping to compensate for the perturbation. There were no speakers in the F2-altering block who only shifted their F1 but not their F2.

One question posed at the beginning of this dissertation was whether speakers had control over F1 and F2, or whether change in one formant was an unintended consequence of the articulatory strategy required to compensate for the altered formant. Principal components were visually examined for direction of movement. For some speakers, a single component was correlated with change in both F1 and F2, so it is unknown whether for these speakers F1 and F2 change were inextricably linked, or if speakers intended to change F1 and F2 together, or if they found an F1-altering strategy early on that also altered F2 and did not feel motivated to find a different strategy. In other speakers, though, formants were clearly linked to separate principal components, indicating that they were controlled independently; principal components are by definition orthogonal. As seen for speaker 310 in the F1-altering block, F2 was controlled by two different articulatory strategies over the course of the 90-trial block, but at the point of greatest compensation for altered F1, there was minimal change seen in F2. Therefore, for at least some speakers, change in F2 is not an automatic consequence of change in F1, suggesting there is likely a perceptual motivation for changing F2 at all.

There are several possibilities for why speakers may have behaved preferentially towards changing F2 instead of F1, and they are all grounded in differences in somatosensation and the idea that speakers may be compensating in ways that diminish their somatosensory mismatch. Speakers are more likely to produce compensatory articulations that feel overall less different from the articulation that they would normally use to produce the target word. Speakers are sensitive to somatosensory feedback (and mismatch of observed and expected somatosensation) during production. Experimentally-induced weakening of somatosensory feedback increases magnitude of compensation in altered auditory feedback (Larson et al. (2008)); the parallel conclusion is that somatosensory mismatch is a limiting factor in compensation for altered auditory feedback.

The findings from the study in this chapter corroborate past findings. MacDonald et al.

(2011) found that speakers respond to a single F1 or F2 formant shift by changing production of only the altered formant, but if both formants are altered, speakers will change production of both formants. They suggest that this is evidence that speakers have independent control over formants. This is supported by the articulatory evidence in this chapter, which showed that not only do speakers have independent control over formants, but that they might have multiple articulatory strategies to maintain such control. Although MacDonald et al. found that speakers will only change production of the altered formant, Katseff (2010) found that speakers may change the production of an unaltered formant as well.

One possibility for explaining the difference in compensation for F1 and F2 is that speakers may experience less of a mismatch in their expected versus produced somatosensory feedback if they adjust two formants rather than just the altered formant. As discussed earlier, speakers may perceive vowels based on a ratio or difference between F2 and F1. Speakers may be able to reach a ratio by changing the production of two formants with less of a somatosensory mismatch than adjusting for only the perturbed formant. If speakers can achieve the correct ratio of F1 and F2 with less acoustic change to any single formant, which may correspond to less articulatory distance traveled, there could be less somatosensory mismatch in compensation. This would explain the general trend seen in the F1-altering block, but further explanation is required to explain the asymmetry in F1 and F2 compensation as well as the asymmetry in production changes in the unaltered formant between the two blocks.

There are differences in the number of options for lowering F2. In the F2-altering block of this experiment, F2 was raised, and so the expected response for a subject was to lower F2. This can be done by either retracting the tongue. The latter option lengthens the entire vocal tract, which consequently lowers all of the formant frequencies, and which may explain why some speakers' F1 values lowered. With two articulatory options for lowering F2, speakers might have been able to compensate more for the shift, resulting in a smaller mismatch in somatosensory feedback for both articulators. This in combination with a smaller perturbation in auditory space may explain why speakers more consistently compensated for a greater percentage of the shift in the F2 block. Lip data were collected and will be considered in future analyses.

The double-articulator explanation does not account for the change in F2 production in the F1-altering block. Speakers who altered their F2 production in this block typically did so by *raising* F2. While the lip-rounding option lowers F2, the lips could potentially flatten and spread to shorten the vocal tract and raise F2. This would actually be a fairly natural articulation given that the lips are typically spread for [i]. By combining production changes in the tongue and the lips, the total displacement of any single articulator could be reduced. A more complete compensation for the auditory perturbation could therefore be achieved with less articulatory change (within any single articulator). One question that leads from this is whether somatosensory mismatch is additive.

There is one other possibility for why there was greater observed difference in F2 than in F1. In addition to possibly having a greater effect on acoustics for a given change in articulation, there is evidence that the posterior portion of the tongue encodes somatosensation *differently* from the front two-thirds of the tongue. Sakamoto et al. (2010) conducted an fMRI study that imaged the brain while different parts of the tongue were stimulated. They found that stimulation to the anterior part of the tongue correlated with activation in the right primary somatosensory cortex (SI). Stimulation in the posterior part of the tongue was additionally correlated with activation in anterior cingulate cortex. This study does not speak to the *sensitivity* of different parts of the tongue, but it is possible that information from different parts of the tongue is processed differently and might have bearing on how this information is processed for speech. Further study would be required to determine if this processing difference could actually account for differences compensation and speech behavior. I propose this as a possible direction of future study rather than as an explanation for the results seen here.

The anterior and posterior parts of the tongue also have different passive antagonist surfaces where they make contact. If speakers are trying to lower their F1, they will raise their tongue body, and they may make contact with the hard palate, inducing a somatosensory mismatch that would be quite obvious to the speaker. If speakers are controlling F2, they may not experience the same level of mismatch. In the F1-altering block, raising F2 in /ɛ/ requires moving the entire tongue body forward and is unlikely to result in any difference felt in the posterior portion of the tongue. Extreme retraction of the tongue to lower F2 might result in contact with the teeth, but it might be possible to compensate for a large enough percentage of the perturbation without such contact. Even if contact is made, the sensation in the palate of the tongue making contact is likely to be quite different from any sensation felt in the teeth due to such contact.

Palate shape and group dynamics

Overall, no influential role of the shape of the palate was found. One goal of this dissertation was to find out whether mouth geometry could help explain variation found in compensation responses. The relationship between palate shape and articulatory/acoustic variability as set out in Chapters 2 and 3 made predictions in two different directions depending on whether palate shape affects physical properties of the vocal tract that have secondary effects on behavior, or on the behavioral habits themselves. In other words, is the effect of palate shape (a) differences in the articulatory - acoustic mapping based on palate shape, or (b) fixed speaker habits that are informed by this mapping. During altered feedback, articulation is *remapped* to acoustics, so speakers' typical speaking habits don't have the effect that the speaker expects, and their internal map is no longer so precise.

These hypotheses are based on the fact that there is a quantal mapping between articulation and acoustics. Chapter 3 considers articulatory and acoustic variability *within* one phoneme. The experiment presented in here in Chapter 4 asks subjects to *change* phonemes: as the trials progress, the speaker hears his feedback with a vowel that has likely changed categories, e.g. from /ɛ/ to /ɪ/. While we observe a relationship between palate flatness and variability within-category, this does not speak to the change in tongue configuration required to change category. The quantal nature of the articulatory-acoustic mapping suggests that regardless of palate shape, phonemes will in general lie in zones where changes in articulation will not change much in acoustics, but that articulatory differences near the boundaries between these zones will result in rapid acoustic change. The predictions concerning palate shape refer to within-zone variability, and it is possible that palate shape does not much affect the regions of rapid acoustic change.

No direct correlation was found between palate shape and compensation, although it did correlate with other behaviors that grouped together. The MDS analysis distinguished groups of speakers. About one-third of the speakers were "reliable" compensators or non-compensators, marked by greater (or very small) rates of compensation in both blocks, and these were separated from each other based on idiosyncrasies in how they did or did not change an unaltered formant, or how much their tongue reconfigured its position between the baseline and the trial of greatest compensation. The other speakers were clumped in a large group that on the whole was minimally distinguished by the other dimensions. This main divide was not at all correlated with palate shape.

Palate shape did, however, correlate with the lower dimensions. These dimensions are somewhat abstract, and it is difficult to ascertain whether the palate actually had an influential effect. Dimension 2 describes two groups of people: one group shifts F2 less or negatively during the F1-altering block despite greater changes in tongue shape in that block, and those people have flatter palates. The other group shifts F2 up more despite changing tongue shape less, and these people have domed palates. It's possible that Dimension 2 is showing that for some vocal tracts and people, the palate changes a speaker's ability to control F1 independently of F2. Dimension 3 describes two groups where how much a speaker changes his tongue configuration in the F1-shifting block is correlated with how much he shifts F1 in the F2-altering block, with less change in both of these variables correlated with flatter palates. This may be suggesting that for some people, the palate does serve as an inhibitory articulation force, limiting his articulation in on block and his "unnecessary" movements in the other block (if he is able to independently and sufficiently change F2 in the F2-altering block). Finally, Dimension 4 finds a curious correlation between having a domed palate and stronger compensation on F2 only, but this only distinguishes speakers within larger groups and does not translate into a global generalization that strong F2 compensators tend to have more domed palates.

The clearest result is that there are many possible factors that affect how a speaker compensates. It is likely that they form a hierarchy. This hierarchy for sensitivity to somatosensory and auditory feedback has already been tested Lametti et al. (2012), and the experiments presented here suggest that palate shape may fit into this hierarchy, but at a lower level than factors influencing sensitivity to different kinds of feedback.

Chapter 5

Experiment 4: Altered auditory feedback in /r/: F3

A third experimental block in the same experiment described in Chapter 4 raised F3 to a maximum of 450 Hz during “heard.” The modeling in Chapter 2 made predictions for how speakers might compensate for shifts in F3, and Chapter 3 showed that during unperturbed speech, speakers control their articulatory variability in their /r/ production in order to maintain consistency in their F3. Despite measures taken to help FUSP distinguish F2 from F3, the equipment did not reliably track and raise F3 for the majority of participants, so group generalizations regarding palate shape can not be made. For four participants, however, there are reliable data. These four subjects are studied here individually.

The American English /r/ is characterized by a very low F3 that approaches F2. While this low F3 is an important perceptual cue that distinguishes /r/ from /l/, the relationship between F2 and F3 is also important (O’Connor et al., 1957), suggesting the strong possibility that speakers will also raise their F2 in response to the perturbation in addition to lowering F3.

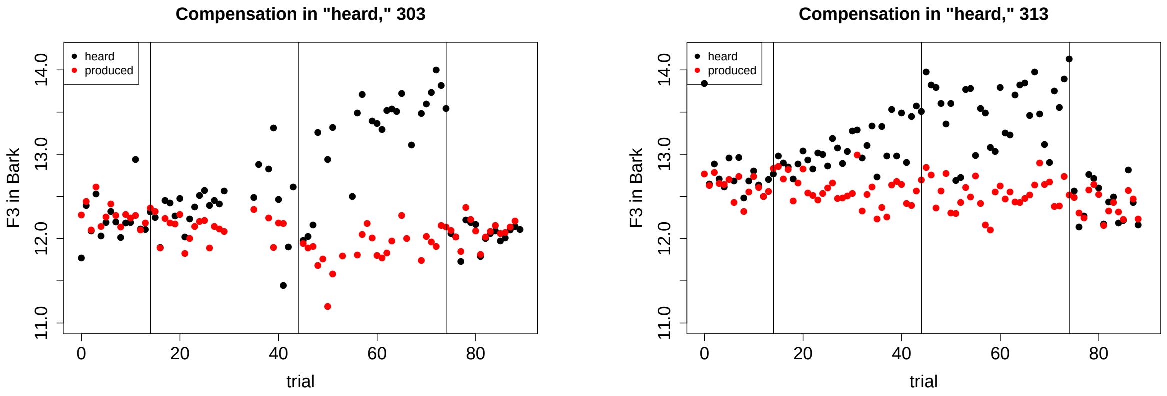
5.1 Acoustics

The acoustic results are summarized in Table 5.1. Figure 5.1 shows the produced and heard F3 values for all four subjects. Subjects did not compensate very much, between 10 and 34% of the perturbation, or between 0.17 and 0.51 Bark. F3 was raised by 450 Hz, or between 1.39 and 1.65 Bark, comparable to the difference in the F1-altering block.

Figure 5.2 shows the produced F2 during the F3-altering block for all four subjects. While all subjects compensated for F3, they are split in their F2 behavior. Two subjects, 313 and 329, additionally raised F2. This was probably done to close the gap between F2 and F3, which was widened by the F3-raising perturbation. Subject 322 compensated the least, and in fact shifted F2 more than F3.

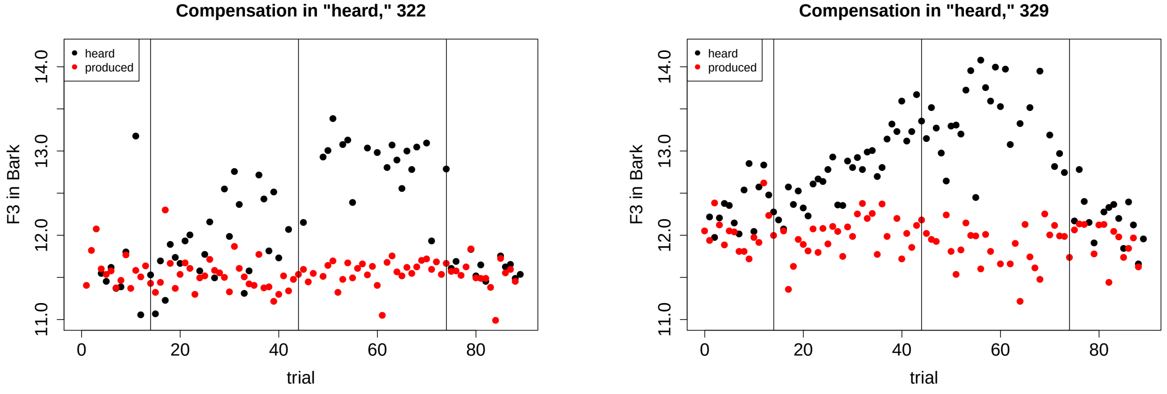
Subject	Alpha	Trial	Prop. F3 Comp.	F3 diff (Bark)	F2 diff (Bark)	Art.
303	2.11	49	0.34	0.51	-0.05	0.41
313	1.87	36	0.19	0.26	0.12	0.59
322	2.07	38	0.10	0.17	-0.19	0.45
329	1.99	67	0.24	0.38	0.12	0.37

Table 5.1: Summary of acoustic results for the F3-altering block. From left: subject, alpha (palate doming measure from previous chapters), trial of greatest compensation, proportion compensation for the F3 perturbation, the trial of greatest compensation, and how much they shifted F3 and F2 in Bark, and the articulatory difference at the trial of greatest compensation (as calculated in Chapter 4).



(a) Subject 303 F3

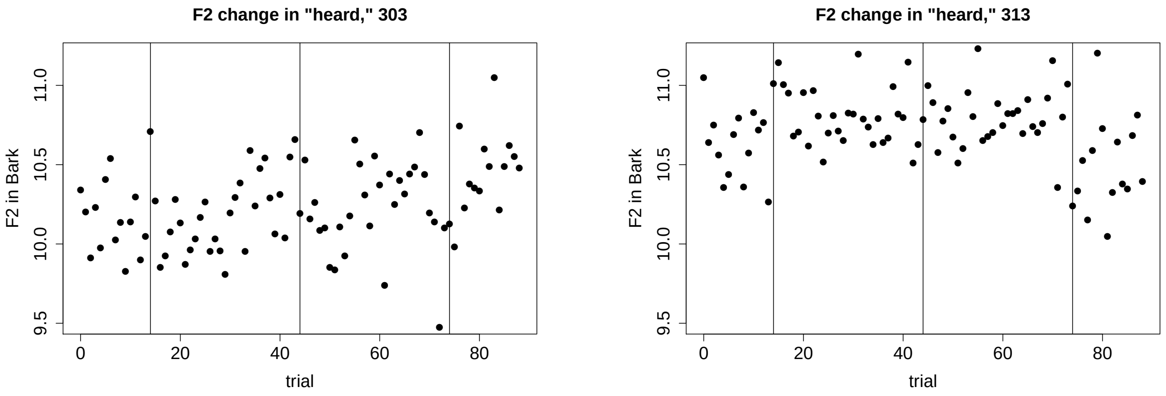
(b) Subject 313 F3



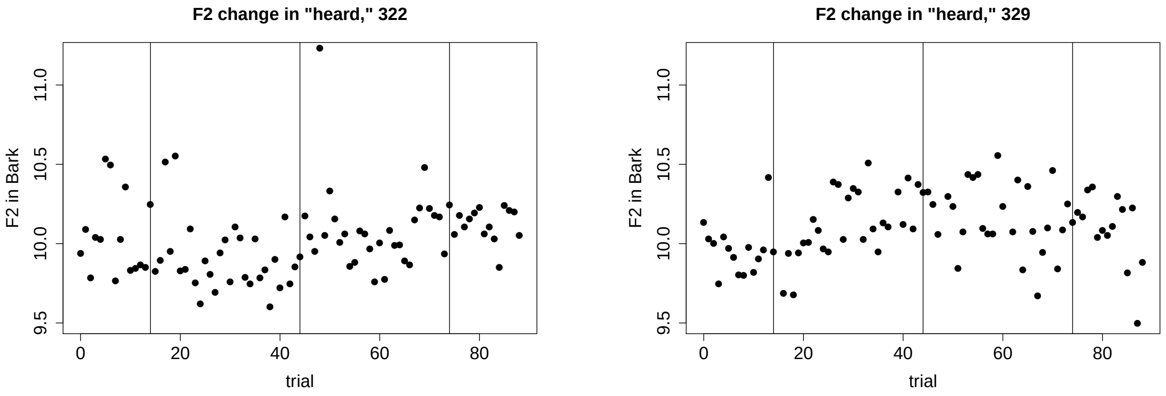
(c) Subject 322

(d) Subject 329

Figure 5.1: Produced and heard formant values for four subjects in the experiment. Vertical lines indicate phase boundaries: baseline, ramp, hold, and washout.



(a) Subject 303 F2 (b) Subject 313 F2

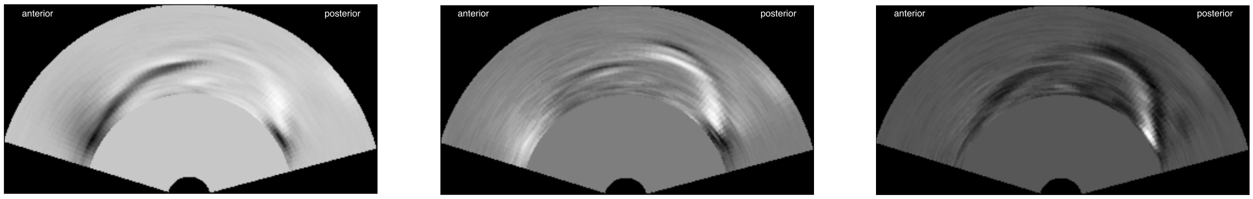


(c) Subject 322 F2 (d) Subject 329 F2

Figure 5.2: F2 values for four subjects in the experiment. Vertical lines indicate phase boundaries: baseline, ramp, hold, and washout.

5.2 Articulation

In this experiment, F3 was raised, and subjects were predicted to lower F3 to compensate. There are multiple ways to lower F3: rounding the lips, tightening the main constriction, and creating a greater constriction at the pharynx (which would not be visible by the ultrasound). For some of these strategies, other formants might be secondarily affected, which may explain observed differences in F2. For all four subjects, both F2 and F3 (respectively or together) were strongly correlated with at least one of the first three principal components. Each subject will be considered individually here. As in the previous two blocks, a separate PCA was carried out for each subject and block, so the principal components represent tongue shapes from this block only.

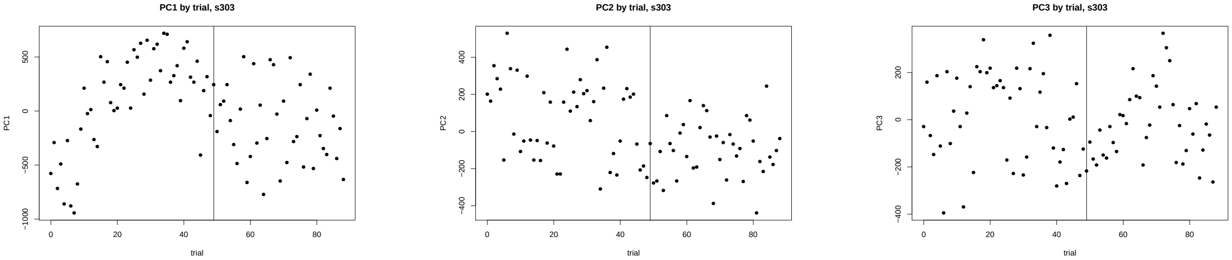


(a) PC1, F2 (b) PC2, F3 (c) PC3, F2

Figure 5.3: PCs and correlation with F2 or F3 for subject 303.

303

Subject 303 compensated the most for F3, and he changed his F2 the least of all three subjects on his trial of greatest compensation. PC1 and PC3 are both nonetheless strongly correlated with F2. Figure 5.3 shows loadings on the first three PCs. The values for these PCs are plotted against trial number (trial of greatest compensation marked with a vertical line) in Figure 5.4

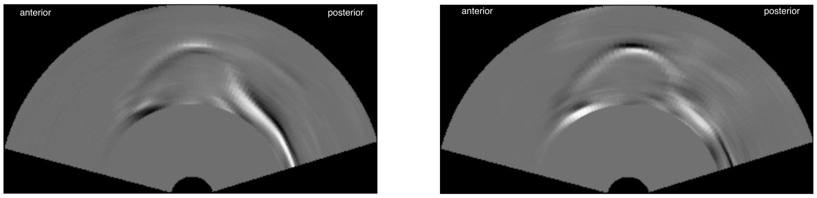


(a) PC1 (b) PC2 (c) PC3

Figure 5.4: PCs by trial subject 303.

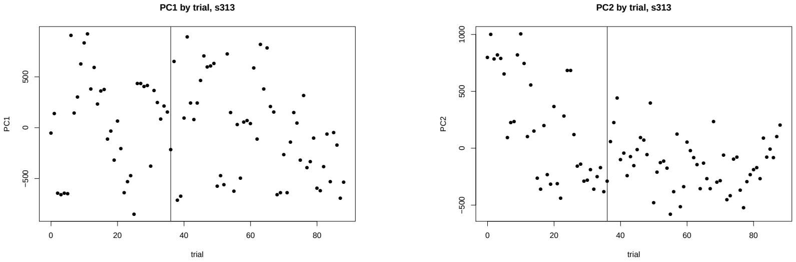
Figure 5.4b shows that over the course of the entire block, the constriction made by the tongue dorsum is retracting, and the blade of the tongue might be lowering; overall, the tongue is hyper-articulating an /r/. PC1 and PC3 shift in a single direction during the trials where the subject is compensating. PC1 and PC3 both show posterior movement, but in different parts of the tongue. By the trial of greatest compensation, he has actually undone the retraction described by PC1, which may explain why there is not a large difference in F2 at that trial.

Overall, the strategy for 303 is to hyper-articulate the /r/ posture and to retract the tongue, which we would expect to lower F3 and F2 respectively. The change in direction of PC1 and PC3 (corresponding to F2), as well as the change in F2 over the block, indicates that he is experimenting with the F2-F3 relationship over these blocks, and he appears to be able to control F3 independently of F2.



(a) PC1, F2 (b) PC2, F3

Figure 5.5: PCs and correlation with F2 or F3 for subject 313.



(a) PC1 (b) PC2

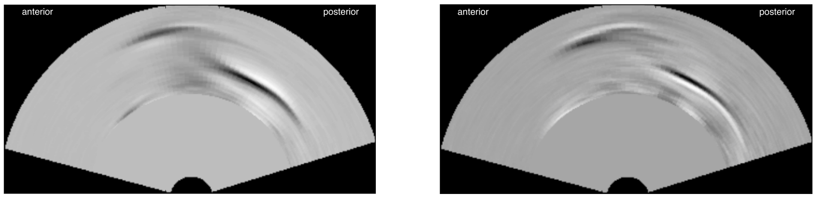
Figure 5.6: PCs by trial subject 313.

313

For subject 313, PC1 correlated with F2, and PC2 correlated with F3. PC1 shows retraction of the entire tongue body. PC2 shows a change in height at the point of greatest constriction. During the trials where he is compensating, he is steadily tightening the constriction, but the tongue retraction is less consistent: in the first half of these trials, the tongue is retracting, but by the last few trials where he is compensating, PC1 has risen again, corresponding with an anterior movement of the tongue. The reliable constriction-tightening suggests that he was aware of how to compensate for the F3 perturbation without searching the articulatory-acoustic space. The differences in F2-changing behavior suggest that he may be searching the F2 space to enhance the perception of an /r/ by bringing F2 and F3 closer together; he is one of the subjects who raised F2 at the trial of greatest compensation.

322

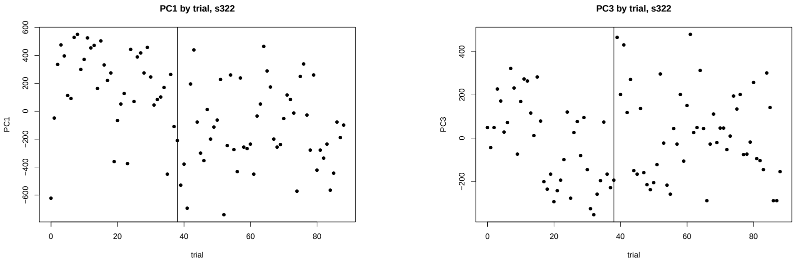
Subject 322's PC1 positively correlated with F3 and negatively with F2, and PC3 correlated negatively with F3. PC2 showed similar changes in tongue blade movement as PC1 but did not pattern strongly with F2 or F3. While the loadings on PC1 and PC3 look similar, they describe subtly different changes in tongue posture. In PC1, the difference between positive and negative values show a tilt at the highest point of the blade, where the positive loadings show that the blade has a steeper downward tilt than the negative loadings. In the tongue dorsum/root area, the positive values are more retracted and raised. Thus the



(a) PC1, F2

(b) PC3, F2

Figure 5.7: PCs and correlation with F2 or F3 for subject 322.



(a) PC1

(b) PC3

Figure 5.8: PCs by trial subject 322.

posture reconfiguration that is associated with tightening the main /r/ constriction is also associated with an anterior movement of the tongue root.

In PC3, the positive values show a higher blade than the negative values, but without the tilt seen in PC1. The tongue dorsum/root area shows similar differences to PC1, except that the values are reversed, where negative values show a more raised and retracted dorsum. Importantly, a greater length of the root loads strongly on PC3, where positive values indicate a more advanced root.

Given that the PCs represent very similar axes of variation, it is informative to see PC1 and PC3 scaled and plotted on the same axes. This is shown in Figure 5.9. The trial of greatest compensation was 38.

This figure shows that while both PC1 and PC3 have a general downward trend through the trial of greatest compensation (38), PC3 crosses the 0-line into the negative side of the variation described by the PC, indicating retraction. PC1 is generally positive (retracted) until several trials before 38, when the tongue blade tilts and becomes slightly flatter. The lack of consistency in these 38 trials suggests the speaker is trying different combinations of root movement and blade tilt to find the articulation that best compensates for the perturbation. The speaker did identify a strategy that would shift the acoustics in a compensatory direction, but he did not compensate very much, suggesting that he had reached either a physical limit or a limit in somatosensory feedback mismatch.

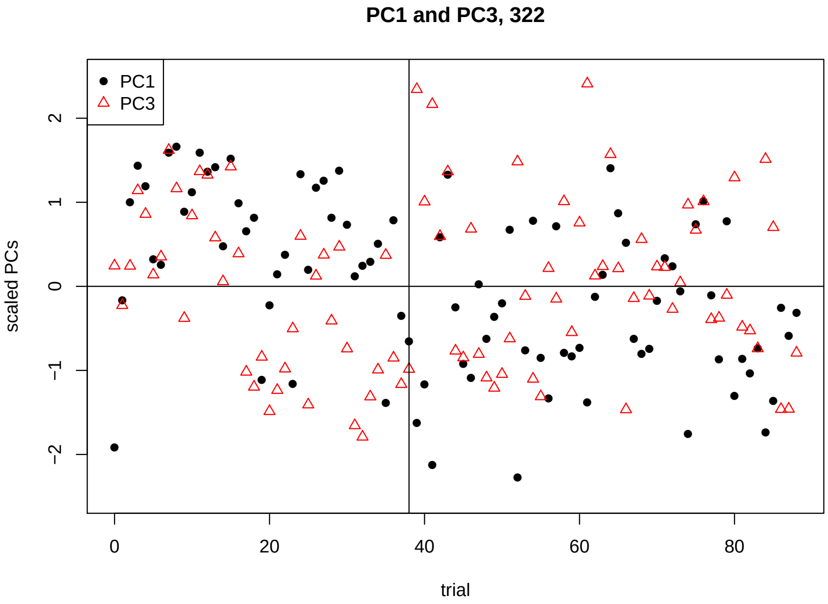
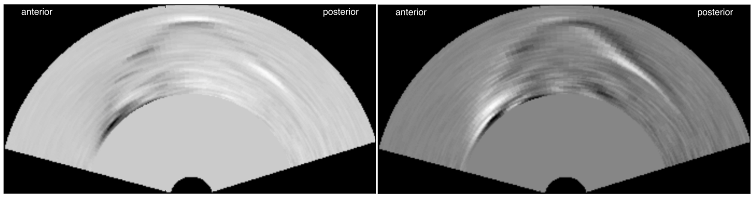


Figure 5.9: Scaled PC1 (black filled circles) and scaled PC3 (red open triangles) on same axes over trial. Vertical line at trial of greatest compensation (38). A line has also been drawn at $y = 0$ to aid in distinguishing positive and negative loadings for the PCs.

329

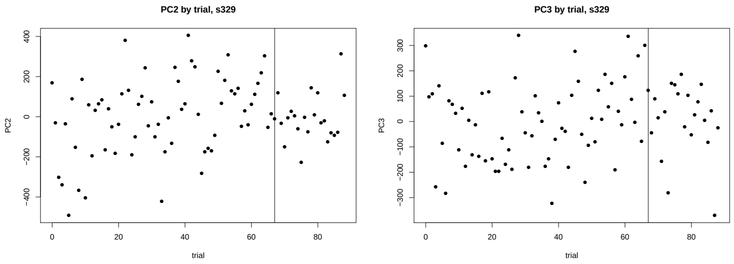
Subject 329 compensated for 38% of the perturbation and shifted up his F2 by a 0.12 Bark on the trial of greatest compensation, (67). The speaker's PCs are shown in Figure 5.10.



(a) PC2, F3 (b) PC3, F2 and F3

Figure 5.10: PCs and correlation with F2 or F3 for subject 329.

PC1 showed change in anteriority of a part of the tongue dorsum that did not consistently affect F3 or F2 and is not shown here; it also does not vary consistently over trials. PC2 is negatively correlated with F3; positive values correspond with F3 compensation. This PC describes a tightening of the constriction: the main point of constriction in the blade raises, the tongue front lowers, and the tongue back behind the constriction point lowers. Overall, the tongue is contracting and hyper-articulating the /r/. Both F2 and F3 are correlated with



(a) PC2 (b) PC3

Figure 5.11: PCs by trial subject 329.

PC3, which shows an overall advancing of the entire tongue. F2 is positively correlated and F3 is negatively correlated, so that movement along this PC changes the distance between F2 and F3. Both PCs generally grow more positive over the course of the compensation, and it appears additionally that the subject is trading off between weighting in the posture reconfigurations described by PC2 and PC3, as both PCs oscillate between positive and negative values.

5.3 Discussion

All four of the subjects discussed above respond to the perturbation by lowering F3. While /r/ tongue postures vary by individual, all four subjects compensated by tightening the point of constriction of their /r/. Subjects differed in their articulatory trajectories. Subject 322 switches between two similar strategies over the trials where he is compensating. The flip-flopping between strategies, combined with the general lack of success in compensation for F3 (10%, the lowest of the four subjects), suggests a lack of certainty over his own vocal tract sensitivity function or the best way to raise F3. Other subjects, like 313, can control F3 and F2 independently in a way that both maximizes F3 compensation and /r/ perception (with a raised F2). Subject 329's articulations over time grow more anterior and the constriction tighter, although he oscillates between stronger and weaker expressions of this movement. It is possible that, for example, 329 tries an articulation, finds it successful, attempts to recreate the same effect with less tongue movement, is unsuccessful, and repeats the cycle based on feedback. 313, on the other hand, abruptly switches direction along the axis represented by PC1, possibly indicating an asymptote in success of the previous articulation. One possible conclusion is that these types of behavior differences indicate differences in internal knowledge about how articulation and acoustics are connected, or possibly in how quickly speakers are able to incorporate somatosensory or auditory feedback. The abrupt change in 313's behavior suggests an anticipation in how the new articulation will affect acoustics, whereas 329's oscillatory approach suggests a possible greater dependence on feedback.

The subjects show different behavior for F2. 313 and 329 both raise F2, probably to minimize the difference between F3 and F2. 303 shows little change in F2 at the trial of greatest compensation, although over the course of the block he experiments with raising F2. 312 lowers his F2 more than he lowers F3. Raising F2 to minimize the difference between F2 and F3 was predicted. Subject 322 lowers his F2 more than his F3, and he has the greatest change in production of his F2. He also has the least compensation for F3 overall. There is not a clear reason to lower F2, but one possibility is that he was trying to retract the tongue to increase the constriction at the pharynx in order to lower F3, and this in turn lowered F2. Thus differences in F2 direction may partially indicate the primary constriction a speaker relies on to maintain a low F3. In contrast, Chapter 4 found that many speakers are able to control formants independently (see also MacDonald et al. (2011)). It is possible that it is more possible to control F1 and F2 independently of one another but less possible to control F2 and F3 independently, and this might be due to the physical location of F2 and F3 nodes and antinodes in the vocal tract. The articulatory changes required to affect F3 might also create constrictions at or near F2 nodes/antinodes.

It is difficult to ascribe a cause to the different kinds of F2 and F3 behavior seen here. Some subjects appear to be able to control F3 independently, indicated by PCs that correspond to either F2 or F3, though some speakers additionally have a component that corresponds to change in both formants. One possibility is that speakers reach a limit in how much they can change F3 independently and add another strategy that also happens to change F2. For at least two of the subjects, it is possible that this change in F2 was intentional to enhance the perception of /r/ by trying to close the gap between F2 and F3. For the two who lowered F2, it is possible that they were trying to enhance the perception of the lip rounding that often accompanies /r/. Speakers may have picked one strategy or another based on the cue that they rely on more strongly to assess the quality of their /r/.

These four subjects were able to compensate for at least some of the perturbation, and they all converged on decreasing the aperture size at the point of maximum constriction. This indicates that while speakers may differ in how they use or control F2 in their perception and production of /r/, they all had some prediction for what kind of articulation would result in a lowered F3, and lowering F3 was a primary objective for all subjects in compensation.

Chapter 6

General Discussion

This dissertation examined various ways speakers compensate when presented with an altered version of their own speech. Speakers vary widely in how *much* they compensate as well as how they try to settle on an articulatory strategy during adaptation. An interesting result is that even speakers who seem not to compensate at all in the acoustics are clearly trying to compensate, which was revealed by linking the ultrasound with acoustics: there were speakers who changed articulation in response to trial, even if the formant being manipulated by the speaker was not the formant perturbed in the experiment. In particular, the set of experiments here compares the role of mouth geometry in normal speech production versus online speech correction.

The modeling experiment in Chapter 2 showed that for /r/, the sensitivity of the mapping between articulation and acoustics was greater for flatter palates. The results showed that in general, the acoustics for flatter palates are more flexible because the mapping between articulation and acoustics is also *less quantal* than for more domed palates: for a given acoustic value, it is more likely that an articulation matching that acoustic value will exist for a flatter palate than a more domed palate. Put another way, more articulations result in the same acoustic values for more domed palates, so there is less control over acoustics for such palates.

This played out in the behavioral experiment in Chapter 3 in speakers' articulatory and acoustic variability. Speakers who had more sensitive articulatory-acoustic mappings (due to having a flatter palate) had reduced articulatory variability. All speakers had similar acoustic variability for /r/, and speakers with flatter palates probably constrained their articulations to constrain acoustic variability.

The combined results from Chapters 2 and 3 gave rise to two competing hypotheses for how speakers would compensate for altered auditory feedback. A higher-sensitivity articulatory-acoustics mapping, where the mapping is less quantal, might correspond to greater compensation: if a speaker has greater flexibility in reaching a particular acoustic target, then the speaker might be more likely to compensate. The opposite prediction was that such speakers have more practiced articulatory habits, so they might be less likely to stray from their motor plans and therefore would compensate less.

Neither hypothesis was correct: palate shape had no direct effect on how much a speaker compensated. *Whether* a speaker was a strong or weak compensator seems to be determined by a higher-level characteristic, such as how much a speaker weighs his auditory feedback against his somatosensory feedback, or the point at which he stops believing that the auditory feedback he hears is self-produced. However, palate shape did correlate with small differences in *how* speakers compensated. Two behavioral differences considered were overall change in the shape of the tongue at the trial corresponding to the greatest amount of compensation as well as how much a speaker changed the production of the unaltered formant.

The modeling chapter made specific predictions for how speakers would compensate in the /r/ block. There were only four speakers who could be analyzed in this block, so it is not very helpful to consider the role of palate shape. However, the modeling did predict articulatory strategies. Four articulatory parameters were examined: orientation of the tongue tip, tongue retraction, bunchiness of the tongue dorsum, and lip protrusion. The modeling predicted that the dorsum shape and position were most important, and indeed, all speakers primarily tried to compensate by increasing the bunchedness of their tongue and raising and retracting it to create a tighter constriction.

Overall, the shape of the palate seems to be a real factor that affects articulatory and acoustic variability, but it appears to control smaller differences between speakers in how speech is produced rather than what speech is produced. The experiments leave open several questions regarding differences in how speakers compensated and whether these differences have a basis in individual speaker variability or if they are due to experimental setup.

In all three blocks of the experiment in Chapters 4 and 5, subjects showed variation in whether and how they changed production of a formant that was not altered. There was a strong correlation between percentage of shift compensated for on the primary formant and change in production of the secondary formant during the F1-altering block. One explanation is that to produce a large enough shift in F1 to compensate, F2 shifts automatically as a result of the tongue movement required to produce the F1 shift. Another possibility is that as speakers may reach a limit for how much they can change F1 independently. During the perturbation, F1 is raised, shrinking the distance between F1 and F2. Shifting F2 up widens the distance between F1 and F2. If speakers are sensitive to some relationship between F1 and F2 Nearey (1989), either in addition to or instead of absolute formant frequencies, then shifting F2 might help in the perception of a more complete compensation.

(Katseff, 2010, p. 108) suggests that between-vowel differences in somatosensory feedback probably doesn't affect amount of compensation, or at least not as much as other differences between vowels. In her experiment, speakers compensated more for /u/ than for lax, non-corner vowels. Katseff reasoned that /u/ is more somatosensorily salient than the other vowels tested because high vowels have additional tactile information from the palate. Nonetheless, this more salient somatosensory mismatch did not prevent compensation in /u/ in her experiment. Therefore, differences in somatosensation are not likely the cause of the differences seen in this experiment.

In American English in particular, backness can be fluid. Speakers are known to vary in the backness of their back vowels, and the difference in F2 between front and back vowels is diminishing (Hinton et al., 1987). For this reason, it is surprising that speakers compensated as much as they did in the "hood" block. However, this fluidity might also have contributed to speakers' willingness to shift F2 in both blocks: if speakers' experience is that their own back vowels may vary in backness, they may be less sensitive (at a perceptual level) to somatosensory feedback involving the tongue root, or they might be more permissive in what tongue positions are acceptable for a production of /ʊ/.

The discrepancy between the F1- and F2-altering blocks with respect to changing production in the unaltered formant gains additional context from the F3-altering block. There were only four subjects, but all changed their production of F2 in addition to their F3 compensation. Speakers differed in how they changed F2: two subjects raised F2, probably to increase the perception of /r/ by bringing F2 and F3 closer together, but the other two lowered F2, possibly to increase the perception of the lip rounding that accompanies /r/. It is also possible that these speakers additionally rounded their lips.

All four subjects whose data is presented in Chapter 5 were able to compensate for the F3 perturbation, even subject 329, who did not compensate at all in the F1- or F2-altering blocks. Four subjects are not enough to form a generalization, but it might be expected that speakers would be more successful at finding a compensatory articulation just because the number of possibilities increases with formant number. The number of wavelengths that fit inside a tube that is open at one end (e.g. the vocal tract) increases by a half wavelength for every formant. The number of pressure antinodes thus increases for each formant. This is important because constricting a tube at a formant's antinode lowers the frequency of that formant. If there are more antinodes, then there are more physical places along the tube where a constriction can be made that will lower that formant. Thus for F3 there are the greatest number of possibilities for an articulation that could lower F3, increasing the chances that a speaker will find at least one articulation that can lower F3.

Future research

One major problem that prevented the ability to draw larger connections between the two vowel blocks was that there was possibly too much variability. Stronger connections between F1 and F2 strategies might have been drawn if speakers tended to be similarly successful in both blocks. Large formant shifts were chosen for this experiment to induce variability between speakers, but it turned out that so much variability prevented certain kinds of comparisons between blocks. A future iteration of this study would choose a smaller perturbation, or perturbations that were scaled between blocks to produce comparable compensation results.

Data revealing how subjects used their lips to help compensate has been collected but not yet analyzed. Analysis of this material may help reveal why and how subjects were better at compensating for F2 than F1. Data from other languages would also help. There were two hypotheses presented above concerning the difference in F1 and F2 compensation. One was that speakers may be perceptually less sensitive to differences in F2 given experience with American English back vowels, and the other is that there may be differences in somatosensation sensitivity in different parts of the tongue. Testing in other languages would help to distinguish between different hypotheses. If willingness to change F2 is language-dependent, we would expect different results based on the importance of F2 in distinguishing categories in that language. If differences in sensitivity are the cause, then there should be similar results across languages.

While only four speakers had useable data from the F3-altering block, they provided useful information. As in the F1- and F2-altering blocks, some but not all speakers also changed an unaltered formant (F2) in a direction that would bring F2 and F3 closer together to increase the perception of /r/. All four speakers did try a similar bunching motion to decrease the constriction size of their /r/. Speakers also retracted their tongues, but it is unknown whether they were trying to increase their pharyngeal constriction or if they were intentionally lowering F2. Further analysis of video data will show whether they also tried to change their lip curl or aperture.

Finally, cross-comparison with other languages will help reveal differences in how speakers change different formants. Altered feedback studies have been carried out extensively for English, and extension to other languages is a necessary future avenue of research to test whether the role of auditory feedback in speech motor control is universal or language-specific.

Bibliography

Abbs, James H., and Vincent L. Gracco (1984). Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech. *Journal of Neurophysiology* 51:705–723.

Aliu, Sheye, John F. Houde, and Srikantan S. Nagarajan (2008). Motor-induced suppression of the auditory cortex. *Journal of Cognitive Neuroscience* 21:791–802.

Atal, B. S., J. J. Chang, M. V. Matthews, and J. W. Tukey (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America* 63:1535–1555.

Bakst, Sarah (2014). Accounting for variation in Tamil retroflex articulation. Poster session presented at: Sound Change in Interacting Human Systems Workshop on Sound Change; May 28-31, Berkeley, CA.

Bakst, Sarah, and Susan Lin (2015). An ultrasound investigation into articulatory variation in American /r/ and /s/. In *Proceedings of the 18th International Congress of Phonetic Sciences*, ed. The Scottish Consortium for ICPhS 2015. Glasgow, UK: The University of Glasgow.

Baum, Shari R., and David H. McFarland (1997). The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America* 102:2353–2359.

Bladon, R.A.W., and F. Nolan. (1977). A video-flurographic investigation of tip and blade alveolars in English. *Journal of Phonetics* 5:185–193.

Boersma, Paul, and David Weenink (2009). Praat: doing phonetics by computer (Version 5.1.05)[computer program].

Brunner, Jana, Susanne Fuchs, and Pascal Perrier (2009). On the relationship between palate shape and articulatory behavior. *Journal of the Acoustical Society of America* 125:3936–3949.

Carre, R., and M. Mrayati (1988). Articulatory-acoustic-phonetic relations and modeling, regions, and modes. In *Speech production and speech modeling*, ed. A. Marchal and W.J. Hardcastle, NATO ASI Series. Dordrecht: Kluwer Academic Publishers.

Cui, F., D. Arnstein, R.M. Thomas, N.M. Maurits, C. Keysers, and V. Gazzola (2014). Functional magnetic resonance imaging connectivity analyses reveal efference-copy to primary somatosensory area, ba2. *PLoS One* .

Dailide, Saulius, and Aidas Dailide (2014). Pixelmator. Version 3.2 Sandstone.

Dart, Sarah N. (1991). Articulatory and acoustic properties of apical and laminal articulations. Doctoral Dissertation, UCLA.

Davidson, Lisa (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120:407–415.

Delattre, Pierre, and Donald C. Freeman (1968). A dialect study of American r's by X-ray motion picture. *Linguistics* 44:29–68.

Dimov, Svetlin, Shira Katseff, and Keith Johnson (2012). Social and personality variables in compensation for altered auditory feedback. In *The initiation of sound change: Perception, production, and social factors*, ed. Maria-Josep Solé and Daniel Recasens, 185–210. John Benjamins.

Espy-Wilson, Carol Y., Suzanne Boyce, Michel Jackson, Shrikanth Narayanan, and Abeer Alwan (2000). Acoustic modeling of american english /r/. *Journal of the Acoustical Society of America* 108:343–356.

Gick, Bryan, Khalil Iskarous, D. H. Whalen, and Louis Goldstein (2003). Constraints on variation in the production of English /r/. In *Proceedings of the 6th International Seminar on Speech Production*, ed. S. Palethorpe and M. Tabain.

Gu, Chong (2002). *Smoothing spline ANOVA models*. Springer Series in Statistics. New York: Springer-Verlag.

Guenther, Frank H. (1994). A neural network model of speech acquisition and motor equivalent speech production. Technical report, Boston University Center for Adaptive Systems.

Guenther, Frank H., Carol Y. Espy-Wilson, Suzanne Boyce, Melanie Matthies, Majid Zandipour, and Joseph S. Perkell (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* 105:2854–2865.

Guenther, Frank H., and Joseph S. Perkell (2004). A neural model of speech production and its application to studies of the role of auditory feedback in speech. In *Speech motor control in normal and disordered speech*, ed. B. Maassen, R. Kent, H.F.M. Peters, P. Van Lieshout, and W. Hulstijn, 29–50. Oxford University Press.

Hamann, Silke (2002). Retroflexion and Retraction revised. *ZAS Papers in Linguistics* 28.

Hinton, Leanne, Birch Moonwoman, Sue Bremner, Herb Luthin, Mary Van Clay, Jean Lerner, and Hazel Corcoran (1987). It's Not Just the Valley Girls: A Study of California English. In *The Annual Proceedings of the Berkeley Linguistics Society*, volume 13, 117–128.

Hirano, S., H. Kojima, Y. Naito, I. Honjo, Y. Kamoto, H. Okazawa, K. Ishizu, Y. Yonekura, Y. Nagahama, H. Fukuyama, and J. Konishi (1997). Cortical processing mechanism for vocalization wtih auditory verbal feedback. *NeuroReport* 8.

Houde, John F., and Michael I. Jordan (1998). Sensorimotor adaptation in speech production. *Science* 279:1213–1216.

Houde, John F., and Srikantan S. Nagarajan (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience* 5:1–14.

Inc., Arizona Software (2010). Graphclick. http://www.arizona-software.ch/graphclick/.

Ivry, Richard B., and Paul C. Lebby (1993). Hemispheric differences in auditory perception are similar to those found in visual perception. *Psychological Science* .

Johnson, Keith (2011). Retroflex versus bunched [r] in compensation for coarticulation. *UC Berkeley Phonology Lab Annual Report* .

Johnson, Keith, Peter Ladefoged, and Mona Lindau (1993). Individual differences in vowel production. *Journal of the Acoustical Society of America* 94:701–714.

Katseff, Shira, John F. Houde, and Keith Johnson (2012). Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Language and Speech* 55:295–308.

Katseff, Shira Eden (2010). Linguistic constraints on compensation for altered auditory feedback. Doctoral Dissertation, University of California, Berkeley.

Keating, Patricia A. (1990). The window model of coarticulation: articulatory evidence. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. John Kingston and Mary E. Beckman, 451–470.

Keating, Patricia A. (1991). Coronal places of articulation. In *The special status of coronals.*, ed. Carole Paradis and Jean-Francois Prunet, volume 2, 29–48. Academic Press.

Ladefoged, Peter, and Ian Maddieson (1996). *The sounds of the world's languages*. Blackwell.

Ladefoged, P. & P. Bhaskararao (1983). Non-quantal aspects of consonant production. *Journal of Phonetics* 11:291–302.

Lametti, Daniel R., Sazzad M. Nasir, and David J. Ostry (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience* 32:9351–9358.

Lammert, Adam, Michael Proctor, and Shrikanth Narayanan (2013). Morphological variation in the adult hard palate and posterior pharyngeal wall. *Journal of Speech, Language, and Hearing Research* 56:521–530.

Larson, Charles R., Kenneth W. Altman, Hanjun Liu, and Timothy C. Hain (2008). Interactions between auditory and somatosensory feedback for voice f0 control. *Experimental Brain research* .

Lawson, Eleanor, James M. Scobbie, and Jane Stuart-Smith (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15:256–268.

Lee, Chin-Hui (1988). On robust linear prediction of speech. *IEEE Transactions on Speech and Audio Processing* 36:642–650.

Li, Min, Chandra Kambhamettu, and Maureen Stone (2005). Automatic contour tracking in ultrasound images. *International Journal of Clinical Linguistics and Phonetics* 19:545–554.

MacDonald, Ewen N., David W. Purcell, and Kevin G. Munhall (2011). Probing the independence of formant control using altered auditory feedback. *Journal of the Acoustical Society of America* 129:955–965.

Maeda, Shinji (1990). Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In *Speech production and speech modeling*, ed. W. J. Hardcastle and A. Marchal. Kluwer Academic Publishers.

Maeda, Shinji, and Kiyoshi Honda (1994). From emg to formant patterns of vowels: The implication of vowel spaces. *Phonetica* 51:17–29.

Manzara, Leonard (1993). The tube resonance model speech synthesizer. Master's thesis, University of Calgary.

Mathôt, S., D. Schreij, and J. Theeuwes (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods* 44:314–324.

Mielke, Jeff, Adam Baker, and Diana Archangeli (2010). Variability and homogeneity in American English /r/ allophony and /s/ retraction. *Laboratory Phonology* 10:699–729.

Mitsuya, Takashi, Ewan N. MacDonald, Kevin G. Munhall, and David W. Purcell (2015). Formant compensation for auditory feedback with English vowels. *Journal of the Acoustical Society of America* 138:413–424.

Mrayati, M., R. Carre, and B. Guerin (1988). Distinctive regions and modes: a new theory of speech production. *Speech Communication* 7:257–286.

Namasivayam, Aravind Kumar, Pascal van Lieshout, and Luc De Nil (2008). Bite-block perturbation in people who stutter: Immediate compensatory and delayed adaptive processes. *Journal of Communication Disorders* 41:372–394.

Nasir, Sazzad M., and David J. Ostry (2006). Somatosensory precision in speech production. *Current Biology* 16:1918–1923.

Nearey, Terrance M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America* 85:2088–2113.

Niziolek, Caroline A., and Frank H. Guenther (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience* 33:12090–12098.

O'Connor, J. D., L. J. Gerstamn, A. M. Liberman, P. C. Delattre, and F. S. Cooper (1957). Acoustic cues for the perception of initial /w, j, r, l/ in english. *Word* 24–43.

Perkell, Joseph, Melanie Matthies, Harlan Lane, Frank Guenther, Reiner Wilhelms-Tricarico, Jane Wozniak, and Peter Guiod (1997). Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication* 22:227–250.

Pile, E.J.S., H.R. Dajani, D. W. Purcell, and K. G. Munhall (2007). Talking under conditions of altered auditory feedback: Does adaptation of one vowel generalize to other vowels? In *International Congress on Phonetic Sciences, Saarbrücken.*

Reznik, Daniel, Ori Ossmy, and Roy Mukamel (2015). Enhanced auditory evoked activity to self-generated sounds is mediated by primary and supplementary motor cortices. *The Journal of Neuroscience* 35:2173–2180.

Sakamoto, Kiwako, Hiroki Nakata, Koji Inui, Mauro Gianni Perrucci, Cosimo Del Gratta, Ryusuke Kakigi, and Gian Luca Romani (2010). A difference exists in somatosensory processing between the anterior and posterior parts of the tongue. *Neuroscience Research* 66:173–179.

Saltzman, Elliot L., and Kevin G. Munhall (1989). A dynamical approach to gestural patterning in speech production. *Haskins Laboratories Status Report on Speech Research* 99/100:38–68.

Smith, Caitlin, Michael Proctor, Khalil Iskarous, Louis Goldstein, and Shrikanth Narayanan (2013). Stable articulatory tasks and their variable formation: Tamil retroflex consonants. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association*, 2006–2009. Lyon, France.

Sprouse, Ronald L., and Keith Johnson (2016). The berkeley phonetics machine. In *17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016): Understanding Speech Processing in Humans and Machines*. International Speech Communication Association (ISCA).

Stevens, Kenneth N. (1972). The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data. In *Human Communication: A Unified View.*, ed. Edward E. David and Peter B. Denes, 51–66. New York: McGraw-Hill.

Stevens, Kenneth N., and Sheila E. Blumstein (1975). Quantal aspects of consonant production and perception: a study of retroflex stop consonants. *Journal of Phonetics* 3:215–233.

Story, Brad H. (2005). A parametric model of the vocal tract area function for vowel and consonant simulation. *Journal of the Acoustical Society of America* 117:3231–3254.

Story, Brad H., Ingo R. Titze, and Eric A. Hoffman (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America* 100:537–554.

Tan, Huiling, Cian Wade, and Peter Brown (2016). Post-movement beta activity in sensorimotor cortex indexes confidence in the estimations from internal models. *The Journal of Neuroscience* 36:1516–1528.

Thibeault, Melanie, Lucie Menard, Shari R. Baum, Gabrielle Richard, and David H. McFarland (2011). Articulatory and acoustic adaptation to palatal perturbation. *Journal of the Acoustical Society of America* 129:2112–2120.

Tremblay, Stéphanie, Douglas M. Shiller, and David J. Ostry (2003). Somatosensory basis of speech production. *Nature* 423.

Twist, Alina, Adam Baker, Jeff Mielke, and Diana Archangeli (2007). Are ‘covert’ /ɹ/ allophones really indistinguishable? *University of Pennsylvania Working Papers in Linguistics* 13.

Ueda, Yuichi, Tomoya Hamakawa, Tadashi Sakata, Syota Hario, and Akira Watanabe (2007). A real-time formant tracker based on the inverse filter control method. *Acoustical Science and Technology of the Acoustical Society of Japan* 28.

Vorperian, Houri K., Ray D. Kent, Mary J. Lindstrom, Cliff M. Kalina, Lindell R. Gentry, and Brian S. Yandell (2005). Development of vocal tract length during early childhood: A magnetic resonance imaging study. *Journal of the Acoustical Society of America* 117:338–350.

Watanabe, Akira (2001). Formant estimation method using inverse filter control. *IEEE Transactions on Speech and Audio Processing* 9:317–26.

Wilson, Stephen M, Ayse Pinar Saygin, Martin I Sereno, and Marco Iacoboni (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience* 7:701–702.

Wood, Sidney (1979). A radiographic analysis of constriction locations for vowels. *Journal of Phonetics* 25–43.

Yuan, J., and M. Liberman (2008). Speaker identification on the scotus corpus. *Journal of the Acoustical Society of America* 123.

Zhou, Xinhul, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland, and Ann Choe (2008). A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *Journal of the Acoustical Society of America* 123:4466–4481.