Muscle Tension Dysphonia as a Disorder of Motor Learning

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Kari Elizabeth Urberg-Carlson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Benjamin Munson, Adviser

April, 2013

Dedication

To my family. It really is done this time.

Abstract

Background: Adaptive learning has been demonstrated in many areas of motor learning. In speech, adaptive responses to auditory perturbation of fundamental frequency, formant frequencies and centroid frequencies of fricatives have been demonstrated. This dissertation presents the hypothesis that the motor changes observed in muscle tension dysphonia may be due to adaptive learning. To begin to test this hypothesis, an experiment was designed to look for evidence of an adaptive learning response to imposed auditory perturbation of voice quality.

Methods: 16 participants repeated the syllable /ha/ while listening to noise under a number of experimental conditions. The training condition presented a re-synthesized recording of the participants own voices, which had an artificially increased noise-to-harmonic ratio intended to simulate breathiness. A control condition presented speech babble at the same intensity. Catch trials in which the noise was turned off were included to test for evidence of motor learning, and trials where the participants repeated /he/ were included to test for evidence of generalization to untrained stimuli. H1-H2, a measure of spectral slant, was the dependent measure. A second experiment compared participants' performance on a task of auditory perception of breathiness to their response to the auditory perturbation.

Results: 12 of 16 participants showed statistically different values of H1-H2 between the training and control conditions. As none of the group differences between conditions were significant, this experiment was not able to demonstrate adaptive

learning. There was no relationship between performance on the auditory perception task and performance on the adaptive learning task.

Conclusions: Given the large body of evidence supporting the concept of adaptive learning in many domains of motor behavior, it is unlikely that behaviors that control voice quality are not subject to adaptive learning. Limitations of the experiment are discussed.

Table of Contents

List of Tables

List of Figures

Introduction

In recent years, adaptive internal models have received a great deal of interest from researchers as a theory to explain motor control of human speech. Houde and Jordan (1998) and Guenther (2006) used adaptive learning paradigms to investigate the behavior of the speech articulators under conditions of perturbed auditory feedback. Larson (1998) and Jones and Munhall (2000) investigated the behavior of the larynx during perturbation of feedback of fundamental frequency. To date, no investigators have examined the behavior of the larynx during perturbation of voice quality feedback. The voice disorder muscle tension dysphonia offers an opportunity to investigate adaptive learning of behavior related to voice quality. Adaptive learning may be able to explain the cause of this disorder, the etiology of which is currently poorly understood.

Muscle tension dysphonia (MTD) is a voice disorder in which a patient is dysphonic, but no biological cause for the dysphonia can be found. Its symptoms include a rough, strained or breathy voice quality, vocal fatigue, and can include pain during or after speech. At least three possible causes of MTD have been proposed: that the dysphonia is a compensation for a physical problem that has resolved, that it is caused by overuse/misuse of the voice, and that it is caused by psychological conflict within the patient (Van Houtte, Van Lierde & Claeys, 2011). Each of these proposed causes are supported primarily by clinical intuition and anecdotal evidence. Additionally, in each case, no mechanism has been proposed to explain how each cause results in the symptoms observed in MTD. The purpose of this dissertation is to propose

a theoretical model to explain how compensation for a physical problem with the larynx could lead to persistent dysphonia, and to begin to test this model experimentally.

Adaptive Internal Models

Adaptive internal models (AIM) are a control process that was originally used for the control of mechanical systems such as robots. They were developed in order to allow stable motor control in systems with long transmission delays (Widrow, 1986). They were later proposed as a way to explain motor control in biological systems in cases where movement times are shorter than the interval required for feedback motor control, such as throwing and speech. In a system controlled by AIM, the controller (the central nervous system, in the case of vertebrate animals) contains two types of models of the system that is being controlled (i.e. the body). An *inverse model* transforms the shape of a desired movement, into the specific muscle commands that will result in that movement. A *forward model* transforms specific motor commands into the expected sensory consequences of those movements. The predicted sensory information of the forward model can then be compared with the actual sensory information that results from the movement to determine whether the movement was successful. If the actual sensory information does not match the expected results, an error signal is generated. This error signal is used to generate feedback motor commands to correct the error. The error signal is also used to train the inverse model to improve the results of future motor commands. As the system gains practice, the actual sensory consequences become closer to the predicted sensory information, and the error signal is reduced. Eventually,

when the actual sensory information matches the predicted sensory information, there is no error signal and the motor behavior stabilizes.

Once it is trained, a system controlled by AIM can produce accurate movements that have a shorter duration than the transmission delay of the system. This is not possible with systems controlled with only feedback control. The advantage that an AIM system has over one that is pre-programmed is that it can adapt to changes in its environment. For example, Callan, Kent, Guenther, and Vorperian (2000) showed that a speech synthesizer controlled by the DIVA model, which is a model that incorporates AIM, (Guenther, 2006) was able to adapt to changes mimicking the changes that the vocal tract undergoes as children mature.

Biological systems have transmission delays that make feedback control of movement impractical for many movements. For example, in humans, the delay from the start of phonation to the time that a feedback motor command can respond to errors in auditory or somatosensory feedback is about 150 ms. By contrast, in English, healthy speakers typically produce about five syllables per second (Ferrand, 2007), making the average duration of a syllable about 200 ms. By the time a speaker hears feedback about the consonant at the beginning of a syllable, they would be producing the vowel that follows it. It is therefore not possible that speech could be controlled by pure feedback control at the rate and accuracy that is observed in most mature humans. AIM is appealing as a model of animal motor control because it allows a system to both learn new movements and respond to changes in the environment, behaviors that are observed in both human and non-human animals.

Studies of limb movements established the experimental paradigm used to test the theory that biological systems use AIM to control movement- adaptive learning. Adaptive learning experiments have three phases. In the baseline phase, the participant performs a task such as throwing balls of clay at a target (Martin, Keating, Goodkin, Bastian, & Thach, 1996) or moving a cursor to a target using a robot arm (Maschke, Gomez, Ebner & Konczak, 2004). In the learning phase, the sensory feedback the participant receives is altered--Martin et al. used prism glasses to shift the participants' visual field, and Maschke et al. used the robot arm to exert force on the arm controlling the cursor to push the arm off course. Over the course of many trials, the participants' accuracy on the task approaches baseline as they learn to compensate for the sensory perturbation. In the final phase the perturbation is removed, and participants initially behave as though the perturbation were still present, producing errors in the opposite direction to that of the perturbation. These overcompensation errors are taken as evidence that the participants are not simply using feedback control to compensate for the sensory perturbation, but that the feedforward commands have been altered, implying changes to the inverse model. Generalization of the adaptation to similar movements (such as untrained cursor targets) is taken as further evidence of changes to the internal model.

AIM Models of Speech Production

The DIVA model (see fig. 1), developed by Guenther (1998, 2006) is an example of a computer implementation of a system controlled by internal models specifically for speech production. The DIVA model is a mathematical model that

provides instructions for the Maeda speech synthesizer (Maeda, 1990). It consists of a speech sound map which, when activated, activates both a set of motor instructions (the inverse model) and a set of predicted auditory and somatosensory consequences (the forward model). The predicted sensory consequences are compared to the actual auditory and somatosensory signals, generating an error signal if they do not exactly match. This error signal can be used to make adjustments to the ongoing motor commands for feedback control, and can also be used to adjust the inverse model to change future feedforward commands.

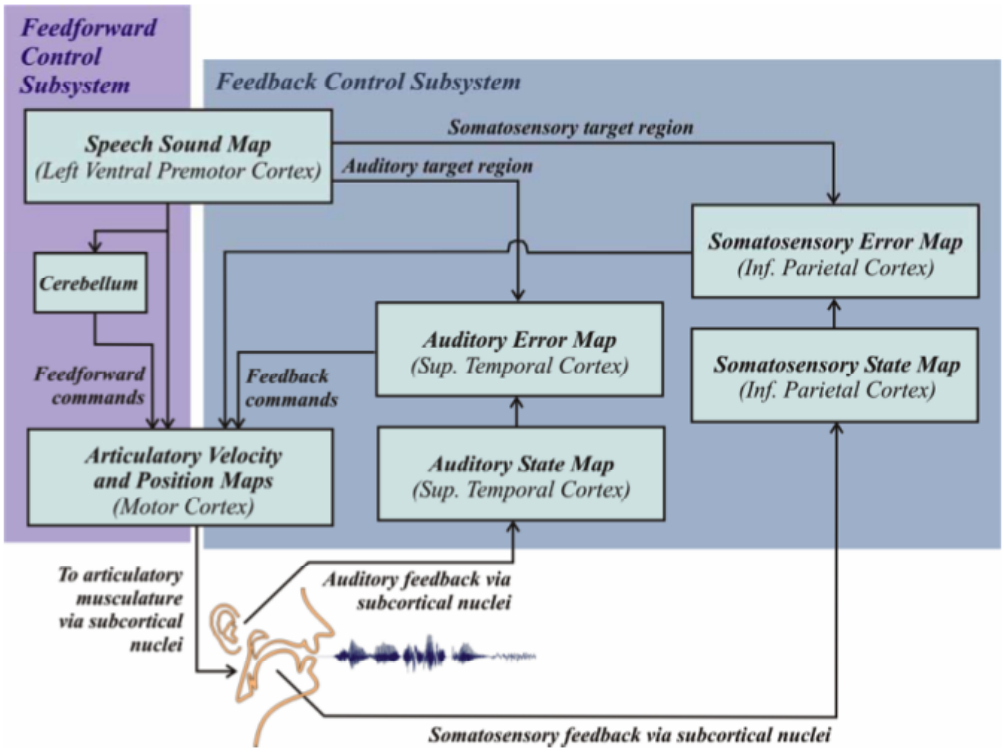


Figure 1. The Diva Model is a computer model of speech based on adaptive internal models (figure from Guenther, 2006, p. 282).

The mapping between speech sounds and motor commands in the DIVA model is tuned by running through a babbling phase in which the model hears the results of its

own productions and adjusts the forward model to reduce the error signal. In this way the DIVA model can learn to produce vowels. The DIVA model does not include mathematical models of the laryngeal system. The relationship between the position and tension of the laryngeal muscles and the resulting auditory signal is less well understood than the relationship between tongue and jaw position to the frequencies of the first two formants. Nevertheless, from a conceptual standpoint the process of adaptive change that has been shown to operate with the articulator muscles should also apply to the laryngeal musculature.

The DIVA model has been shown to be able to adapt and preserve its auditory targets when changes are made to its "vocal tract" that mimic the effects of growth in children (Callan et al. 2000). The 2006 version of the DIVA model includes maps between the functions of the model and proposed regions of the human brain that are believed accomplish those functions. The DIVA model can use those maps to generate a simulated fMRI image that has been compared to images created in human experiments, providing support for the predictions of the model (Tourville, Reilly, & Guenther, 2008).

There is also evidence from human subjects for adaptive inverse models as an explanation for motor control of human speech. Most of the research on adaptive learning in speech has manipulated either the fundamental frequency of the voice, perturbing the perception of pitch, or the first or first and second formant frequencies (F1 and F2), perturbing the perception of vowel height and backness. Formants are areas of the frequency spectrum of speech that are amplified by the resonant

characteristics of the vocal tract. Moving any of the articulators, such as the lips, tongue or jaw, will shift the formant frequencies in a characteristic manner. Shifting F1up makes it sound as though the jaw were lower, and shifting F2 up makes it sound as though the tongue were further forward in the mouth than its actual position. F1 and F2 provide sufficient information for the identification of most vowels in English. Shifting F1 to a higher frequency will make the vowel /e/ sound more like /æ/, shifting F2 to a lower frequency will make the vowel /e/ sound more like /o/.

Using the adaptive learning paradigm, Houde (1998) and Guenther (2006) have shown that participants adjust their articulators to compensate for artificially imposed changes in the first two formant frequencies (F1 and F2). When the perturbations are unexpected, there is a delay in compensation on the order of 150 ms. When perturbations are predictable, the adaptation persists after the perturbation is removed or masked. Houde showed that the adaptation generalized to untrained vowels and phonemic contexts. Aasland, Baum, and McFarlane (2006) used a plastic appliance attached to the roof of the mouth to show that participants adapted their productions of /s/ to compensate for the changed shape of the palate.

Adaptive learning demonstrates the close relationship between perception and motor behavior in speech production. Villacorta, Perkell, and Guenther (2007) showed that participants who did better on a task of F1 discrimination showed a larger degree of compensation to F1 shifts. Additionally, Schiller, Sato, Gracco, and Baum (2009) found that not only were motor behaviors affected by auditory perturbation, but perceptual categories were also affected. In an adaptive learning experiment they artificially shifted

the centroid frequency of productions of /s/, making the productions sound more like /ʃ/. The participants who heard the shifted auditory feedback adjusted their productions of /s/ to compensate for the auditory perturbation, as expected, however they also shifted their boundary of /s/ vs. /ʃ/ in a phoneme identification task following the auditory perturbation. The shift in perception was in the same direction as the direction of the auditory perturbation; tokens that had previously been identified as /ʃ/ were identified as /s/ following motor learning. This shows that adaptive learning can shift the perceptual targets as well as the motor behavior. Nasir and Ostry (2009) perturbed jaw movements in a way that affected proprioceptive feedback but not auditory feedback, and showed both that the perceptual boundary between the vowels /ɛ/ and /æ/ changed following motor learning, and that participants who showed greater motor learning also had larger boundary shifts. Mattar, Darainy, and Ostry (2013), in a similar experiment examining the link between motor behavior and perceptual discrimination in limb movements, showed that the perceptual shifts occurred more slowly than the motor shifts. This suggests that motor change can drive sensory change.

Brain imaging studies using functional magnetic resonance imaging (fMRI) have shown that some areas of the brain are more active when unexpected sensory perturbation is present than when it is not. Guenther (2006) showed that areas of the superior temporal gyrus were more active during auditory perturbation, while areas of the supramarginal gyrus were more active during unexpected somatosensory perturbation. Ghosh, Tourville, and Guenther (2008) showed that the superior temporal gyrus activity was present on trials where feedback motor commands were issued,

supporting the hypothesis that this activity is a sensory error signal used in the feedback motor control system. This strongly supports the AIM model of speech motor control, but is not yet conclusive evidence. The next step would be to show that the strength of the error signal is reduced and disappears as adaptation to auditory perturbation occurs and that it reappears after the removal of the perturbation.

AIM Models of Laryngeal Behavior

The behavior of the larynx primarily influences three aspects of the speech signal: loudness, fundamental frequency and voice quality. Of these, only fundamental frequency has been investigated in adaptive learning experiments. A century ago it was discovered that vocal loudness is affected by ambient noise; as increased background noise masks the voice and makes it sound quieter, the voice increases in loudness to compensate, and as background noise is reduced, the voice decreases in loudness (Lane & Tranel, 1971). The vocal response to loudness is always compensatory--in the opposite direction to the perceptual shift, as was observed in the articulatory response to F1 and F2 shifts described above.

The behavior of the voice in response to artificially imposed shifts in fundamental frequency is a bit more complicated. In cases where the fundamental frequency carries phonemic information as in Mandarin (Xu, Larson, Bauer, & Hain, 2004, Jones & Munhall, 2002) or prosodic information (Natke & Kalveram, 2001, Donath, Natke, & Kalveram, 2002), participants reliably shift their pitch to compensate for imposed shifts in pitch, although the compensatory shift generally does not completely compensate for the perceptual shift. Donath, Natke, and Kalveram (2002)

showed that the participants showed aftereffects when the pitch shift was removed, and Jones and Munhall (2005) showed that the effect generalized to untrained tones in Mandarin.

In studies where pitch did not carry phonemic or prosodic information, the majority of participants compensated for pitch-shifted feedback (Jones and Munhall, 2000). However, a small number of participants shifted their pitch in the same direction as the perturbation, in a following response rather than a compensatory one (Burnett, Senner, & Larson, 1997, Larson, Burnett, & Kiran, 2000). Larson (1998) showed that when pitch shifts were small, most participants compensated for the shift, but as the pitch shift grew larger, there were a larger number of following responses. This suggests that auditory targets are determined in the context of phonemic and prosodic requirements, and that in the absence of those requirements, the behavior of the system is less determined.

Perturbation of voice quality has not been investigated in any published studies to date. Voice quality differs from loudness and pitch in several important ways. First, in many languages voice quality does not transmit contrastive phonemic or prosodic information, although it does carry information about the gender, age, personality, mood and state of health of the speaker, among other things. Second, loudness and pitch are both basically monotonic processes that most people have a great deal of experience manipulating. If either loudness or pitch is shifted perceptually, there are motor behaviors that will compensate for the shift and most people are able to do so. This is not true of voice quality. If aperiodicity is added to the auditory feedback to simulate

roughness or breathiness, adjustments to motor behaviors will have a limited power to reduce and will be unable to eliminate the error signal generated by the perturbation. No amount of improvement in the voice will be able to reduce the amount of noise in the signal below the level of noise that has been added. In studies that examined pitch and formant shifts, compensatory behavior is rewarded by a reduction in the error signal. This would not always be true when voice quality is manipulated.

A third important difference is that the relationship between acoustic measurement of fundamental frequency, psychoacoustic measurement of pitch and the motor behaviors that produce changes in them is well understood. For the purpose of these experiments, pitch varies essentially along one dimension. This is not true of voice quality. As Kreiman (2010) points out, the perceptual judgments that are typically made of voice quality, such as roughness, breathiness, and strain, are poorly understood from a psychoacoustic perspective, and generally correlate poorly with acoustic measures of voice quality. It is not yet known whether these are independent dimensions of voice quality or whether they co-vary. Furthermore, the relationship between motor behavior and voice quality is not as well understood as it is with pitch. This makes objective measurement of voice quality problematic and places limits on our ability to measure change in motor behavior related to voice quality. Despite the difficulties associated with measuring adaptive learning in the context of behaviors associated with voice quality, this model of learning may help explain the cause of one of the most common voice disorders: muscle tension dysphonia.

Muscle Tension Dysphonia

Muscle tension dysphonia (MTD) is a frequently-occurring voice disorder that affects voice quality. It is a common cause of voice disorders in patients seeking help from an ENT, comprising 30% of the general voice population and 40% of patients who are professional voice users (Van Houtte, Van Lierde, Haseleer, & Claeys, 2009). The definition of muscle tension varies among studies. The *Classification Manual of Voice Disorders-I* (Verdolini, Rosen, & Branski, 2006) defines it as persistent dysphonia in the absence of (primary MTD) or out of proportion to (secondary MTD) physical changes to the phonatory system such as inflammation, mass lesions, neurological damage or tissue atrophy. The dysphonia can be mild or very severe and may be accompanied by pain while speaking or swallowing and fatigue after speaking (Verdolini, Rosen, & Branski, 2006).

The dysphonia, fatigue and pain associated with MTD are believed to be due to changes in motor habits associated with speech and voice production. A typical explanation of MTD is "it is considered to be a manifestation of excessive laryngeal musculoskeletal tension and associated hyperfunctional true and/or false vocal fold vibratory patterns, " (Dworkin, Meleca, & Abkarian, 2000). "Excessive musculoskeletal tension" and "hyperfunction" are not defined. Aronson (1990) considered resistance of the larynx or hyoid bone to manual manipulation or pain on palpation of the larynx accompanied by improvement in voice quality if the larynx is manually pulled lower in the neck to be diagnostic of MTD.

There is at least some indirect evidence that MTD is associated with increased neuromuscular activity. Redenbaugh and Reich (1989) and Hocevar-Boltezar, Janko, and Zargi (1998) measured higher levels of surface electromyography (sEMG) activity at several laryngeal and facial sites in patients with MTD than in normal controls. This suggests a higher level of muscle contraction in the extrinsic laryngeal musculature in MTD patients compared to non-patients. Both of these studies had a large number of measures compared to their numbers of participants, so the results should be interpreted with caution. Van Houtte, Claeys, D'haeseleer, Wuyts, and Van Lierde (2011) found no difference in sEMG activity between participants with MTD and participants with no symptoms of MTD.

At least three causes of MTD have been identified:

1. personality/psychological factors. Certain personality traits such as introversion, neuroticism and anxiety have been identified as associated with a diagnosis of MTD (Roy and Bless, 2000).

2. heavy voice use. "Overuse" and behaviors such as "pitching the voice too low" have been identified as causes of MTD (Koufman & Blalock, 1982). Verdolini (2005) points out that terms such as "overuse" and "misuse" are problematic, in that a given behavior is only classified as overuse or misuse if it results in a voice disorder.

3. compensation for an underlying organic voice disorder. MTD often begins after an upper respiratory infection (Koufman and Blalock, 1982), or other organic cause. This suggests that MTD may be a motor strategy to compensate for an organic disorder (Verdolini, Rosen, & Branski, 2006, Van Houtte, Van Lierde, &

Claeys, 2011). The evidence that organic voice disorders can trigger MTD is only correlational. Belafsky, Postma, Reulbach, Holland ,and Koufman (2002) found that 94% of a sample of non-patients who had vocal fold bowing also had "abnormal muscle tension patterns" associated with MTD. These muscle tension patterns were measured with videoendostroboscopic examination, and consisted of findings such as medial approximation of the ventricular folds and anterioposterior contraction of the larynx. Muscle tension was inferred from the visual appearance of the larynx, and was not measured through palpation or electromyography. Koufman, Amin and Panetti (2000) found that 70% of a sample of patients with MTD had pH probe scores indicating the presence of acid reflux. This study had no control group, so it cannot be concluded from this study that rate of acid reflux in MTD patients is higher than the general population.

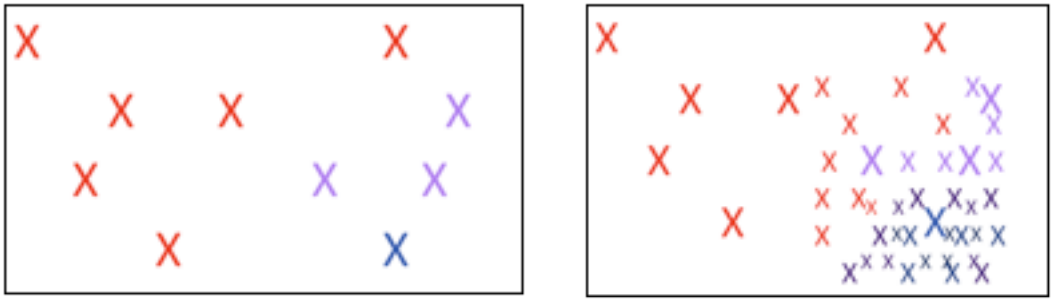
As of yet, no specific mechanism for how these causes result in MTD has been proposed. The theory of adaptive internal models (AIM) offers a possible explanation for how compensation for an organic voice disorder could result in MTD, and it provides testable predictions.

Adaptive Learning as an Explanation for Muscle Tension Dysphonia

To explain how adaptive internal models can explain MTD, we start with how this model predicts that learning occurs. A simplified version of learning an internal model was developed by Metta, Sandini, and Konczak (1999). In this experiment a robotic arm with camera vision was trained to reach for targets. It was designed to gaze at a target. It was initially pre-programmed with three possible arm movements, and the instructions to gaze at the hand after each movement. During this hand gaze, the

coordinates of the camera position were mapped to the coordinates of the arm with an artificial neural network. There was noise in the movements of the arm; each time it reached for the object, it missed by a small amount in a random direction. Then, the camera gaze mapped this new arm position into the set of possible arm movements. Because of this noise, the robot gradually explored and mapped its entire workspace. As the map became more detailed, the arm's accuracy of reaching became more and more accurate.

A similar process may occur with learning of vocal behavior in humans. Initially, when a child is born, the motor control system does not know how to move the speech articulators. As babbling occurs, the system begins to map motor behaviors to their auditory and proprioceptive consequences. These bidirectional maps between motor behaviors and their consequences are the forward and inverse internal models, which are coarse-grained in the early phases of development. Initially the baby makes large movements, which generate large error signals (fig. 2a). As the brain receives these error signals from the auditory system, new information is added to the internal models, which become more precise, allowing more accurate and complex behaviors to be generated (fig. 2b). Over time, as the child continues to practice and more behavior meets the perceptual criteria, fewer error signals occur. The system will approach a stable state in which the behavior does not change because there is no error signal. It is important to note that this is not a static state: the behavior is stable because the error signal is small, not because the behavior has become fixed in any absolute sense.



a b

Figure 2. An initial (a) and a stable (b) state of learning vocal behavior. Location in the x-y plane represents an n-dimensional vector of the muscle activation of the muscles of respiration, phonation, resonation and articulation. Color represents acoustic and proprioceptive feedback where blue represents a desired outcome and red a non-desired outcome.

The physical system that produces sound can be disrupted for many different reasons. Normal growth from infancy to adulthood introduces large changes in the length, mass, and structure of the vocal folds. Inflammation, hemorrhage, and scarring all increase the mass and stiffness of the vibratory portion of the vocal folds. This alters the pitch and amplitude of the voice and perturbs the periodicity of the auditory signal. Increased stiffness and mass of the vocal folds will affect proprioceptive feedback from the laryngeal and respiratory musculature. Tissue atrophy and neurological damage will also change the types of feedback that are received from speech gestures.

When the physical properties of the phonatory system change, motor gestures that produced acceptable feedback in the stable state will now generate an error signal (fig. 3). This error signal will trigger the adaptive learning system to start modifying the internal models, changing the motor instructions to try to return to the desired behavior.

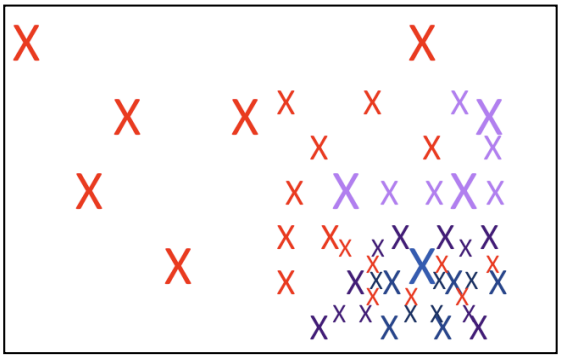


Figure 3. The mapping between muscle gestures and their perceptual outcomes has become perturbed because of physical changes to the larynx. Gestures that formerly would have produced a desired outcome now produce an error signal.

When the system receives error signals in response to a set of instructions that previously produced "good" voice, it tries to find a new set of gestures that will produce the desired result. A compensatory response could raise pitch to counter pitch lowering due to increased mass, increase breath flow to improve the stability of vibration, or use precise articulation to improve intelligibility. These responses could somewhat reduce the magnitude of the error signal. Compensatory responses based on somatosensory feedback could also include, for example, stiffening of the laryngeal musculature to stabilize the larynx in the case of essential tremor or medial compression of the laryngeal muscles in the case of vocal fold bowing or paralysis. Motor changes cannot remove roughness in a larynx with mass lesions on the vocal folds or breathiness in a

larynx where neurological changes prevent the complete closure of the vocal folds, so the error signal will be unable to guide compensation in these cases. A following response would lower pitch in the presence of increased mass or hypoadduct the vocal folds in the presence of breathiness. Factors affecting which response is selected in a particular case might include training, prior experience, voice models in the environment and chance.

There are two possible mechanisms that could account for the maintenance of MTD once the precipitating organic change has resolved. One possibility is that MTD is maintained by retraining of the forward model. Schiller et al. (2009) and Nasir and Ostry (2009) both showed that motor learning produces changes in the boundaries of perceptual targets. Furthermore, Schiller et al. showed that the direction of the perceptual shift would promote accepting the altered feedback as an accurate production. This suggests that if the feedback motor commands do not eliminate the error signal, the system may habituate to the error. In this situation, the forward model could begin to predict dysphonia as a result of the feedforward motor commands. The error signal is no longer produced because the resulting productions match the expected signal. The feedforward motor command that results in dysphonia and/or fatigue is therefore maintained. In this case, the ideal treatment would retrain the forward model so that it no longer predicts dysphonia. This will lead to the return of the auditory error signal in the presence of dysphonia. Once the error signal is restored, the feedback motor learning system will guide the system back to improved voice quality.

A second possibility is that MTD is maintained by reduction in response variability. This possibility hypothesizes that although the error signal is maintained, there is not enough variability in motor production for the learning system to guide the behavior back to its previous state. Initially, when the error signal is first detected, the system tries a large variety of behaviors to try to restore good voice quality. When an organic change to the larynx is present, nothing works well to reduce the error signal. Variability is reduced as whichever behavior produces the smallest error signal is reinforced. When the organic change goes away, small changes in the motor behavior do not result in improvements in the error signal. Large improvements are possible, but those motor behaviors are no longer attempted by the feedforward system. There is a local minimum in the error signal that has trapped the motor system. In this case, any intervention that increases variability in production will help the system escape the local minimum. Once the system increases the variability of motor behavior, the feedback motor learning system will guide the motor system back to improved production.

Brain imaging studies using fMRI might distinguish between these two possibilities. If the first possibility is true, then patients with MTD should have less increase in activity in the superior temporal gyrus when dysphonia is artificially increased than participants with healthy voices. If the second possibility is true, then patients with MTD should have similar performance to healthy controls when dysphonia is artificially increased. It is possible that both explanations are true, either in different individuals or in the same individual.

Predictions

The hypothesis that MTD is caused by disruptions to internal models caused by altered auditory and/or proprioceptive feedback generates testable predictions. The first step is to determine whether adaptive learning is observed in behaviors related to voice quality. The following questions will be addressed in this dissertation:

1. **Do participants change their vocal behavior when their perception of their voice changes?** Because no studies of adaptive learning have yet examined voice quality, this is the first question to be addressed.
2. **If participants do change their behavior, do those changes demonstrate adaptive learning?** In order to do so, the behavior changes that are observed would need to persist transiently once the perturbation in the auditory signal is removed.
3. **Do differences in perceptual discrimination of breathiness predict differences in how participants respond to perceptual perturbations?** Villacorta et al. (2007) found that participants who did better on a discrimination task of variation in F1 showed a larger response to the auditory perturbation task. This question has clinical implications, because if better auditory discrimination leads to more response to perturbation it would suggest that people who have trained their perceptual system more acutely, such as singers and actors, could be more likely to develop MTD than those who have not.

Although in this experiment the auditory feedback perceived by the speaker will be perturbed, the proproceptive feedback produced by the larynx will not. This will create a conflict between these two sources of feedback. Tremblay, Schiller, and Ostry (2003) showed that participants compensated for mechanical perturbations of jaw movement that did not affect the auditory signal. Partipants also compensated for jaw perturations during silent speech, in the absence of auditory information. Nasir and Ostry (2008) found the same effect in profoundly deaf adults, and Larson (2008) showed that compensation for auditory perturbations in F0 was greater when the vocal folds were anesthetized than when they were not. All of these experiments show that both auditory and proprioceptive feedback influence compensatory behaviors.

Individual speakers may vary in the relative extent to which they use proprioceptive and auditory feedback. There is currently no way to perturb proprioceptive feedback of vocal behavior without also changing the dynamics of the larynx in ways that would make acoustic and perceptual measures of voice quality unreliable measures of changes in motor behavior. Because Larson (1998) and Jones and Munhall (2000) were able to achieve changes in motor behavior of the larynx by perturbing auditory feedback of fundamental frequency without perturbing proprioceptive feedback, it is reasonable to presume that this is also possible for voice quality.

Method

This experiment consists of two parts: a sensorimotor adaptation experiment, and a perceptual discrimination experiment. In the adaptation experiment, participants

repeatedly spoke the syllable ‘ha’ while hearing altered auditory feedback about their voice quality, which was intended to make their voice sound breathier than it really was. The purpose of the adaptation experiment was to determine whether participants change their vocal behavior when their auditory perception of their own voice changes. This experiment will test the prediction that acoustic measures of voice quality will change when participants hear altered auditory feedback of their own voices. Because voice quality does not carry contrastive phonemic or prosodic information in English, these acoustic changes may either compensate for the auditory perturbation, or they may be in the same direction as the perturbation. A compensatory response would result in a more pressed or strained vocal quality in response to perception of increased breathiness. A following response would result in a breathier voice quality.

In the perception experiment, participants listened to pairs of resynthesized vowels with different noise-to-harmonic ratios simulating different levels of breathiness. They responded by indicating which vowel of the pair sounded breathier. The purpose of the perception experiment was to determine whether differences in participants’ ability to discriminate levels of breathiness would account for some of the difference in their response to the adaptation experiment. Villacorta, Perkell and Guenther (2007) showed that participants with more acute discrimination of shifts in F1 showed greater shifts in F1 in an adaptation experiment. This experiment predicts that participants with more acute discrimination of levels of breathiness will show greater shifts in voice quality when their auditory feedback is perturbed.

In both experiments, recordings of vowels were inverse-filtered and re-synthesized using software developed by the UCLA Bureau of Glottal Affairs (Kreiman, Gerratt, & Antoñanzas-Barroso, 2006). Inverse filtering is a process that creates a mathematical model of an individual's vocal tract from the sound of their voice. This mathematical model can then be used to resynthesize a computer-generated voice that sounds like the individual. There are three programs in this software package: INVF, the inverse-filtering program; Synthesis, the voice synthesizer; and Sky, an acoustic analysis program. There are versions of this software package for Matlab and Windows. This experiment used the Windows version.

INVF inverse-filters a voice signal. It starts with a digital recording of a voice and separates the effects of the glottal source from the vocal tract filter using a method described by Javkin, Antoñanzas-Barroso, and Maddieson (1987). The output of the INVF program is a mathematical model of the vocal tract filter, which can be used by Synthesis to re-synthesize a voice that is acoustically and perceptually similar to the original recording.

The first step in the inverse-filtering process is to import a .wav file of a vowel. INVF requires the sound files to be sampled at 10 KHz, so the .wav files were resampled using Praat. A section of the .wav file that contains a steady-state section of the vowel is selected for analysis. Once the vowel is identified, the software uses a Fast Fourier Transformation to create a spectrum of the vowel (figure 4).

Figure 4. Fast Fourier Transform spectrum of a vowel (From Kreiman, Gerratt and Antoñanzas-Barroso, 2006, p. 14).

Once the spectrum of the vowel has been calculated, Linear Predictive coding is used to identify the formant peaks of the vowel (fig. 5).

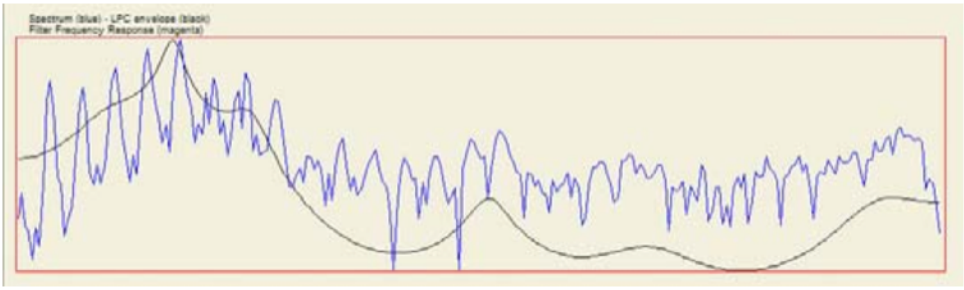


Figure 5. FFT spectrum with formants identified by linear predictive coding (from Kreiman, Gerratt and Antoñanzas-Barroso, 2006, p. 14).

The formant peaks provide an estimate of the vocal tract filter. With that information, the spectrum of the glottal source can then be estimated. If the inverse filtering has worked properly, the glottal source spectrum should appear to be a monotonically decreasing function (fig. 6). If the linear predictive coding identifies spurious formants, the calculated glottal source spectrum will have a different shape (fig. 7).

Figure 6. The calculated glottal source spectrum produced by INVF when working properly.

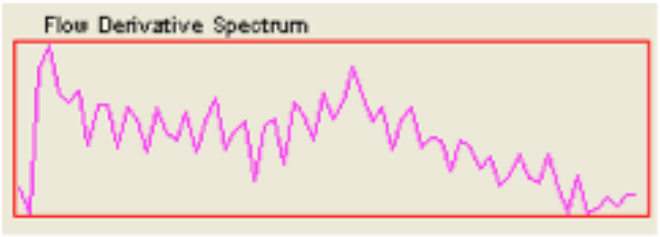


Figure 7. A calculated glottal source spectrum produced by INVF with spurious formants.

The inverse-filtering process is somewhat noisy and imprecise. There is no way to know from the output of the inverse-filter whether the result is accurate, except by re-synthesizing the voice and listening to the resulting sound to see whether it accurately reproduces the original voice. If spurious formants have been introduced to the model, they can be removed manually to improve the output. In practice, when preparing the stimuli for both experiments, the inverse-filtering and re-synthesis was often repeated many times using different segments of the original recording until the result was acceptable. Figure 8 shows an example of a spurious formant in the LPC envelope and the resulting flow derivative spectrum. Clicking the mouse on the spurious formant on the LPC envelope window changes the shape of the envelope, leading to the improved flow derivative spectrum in figure 9.

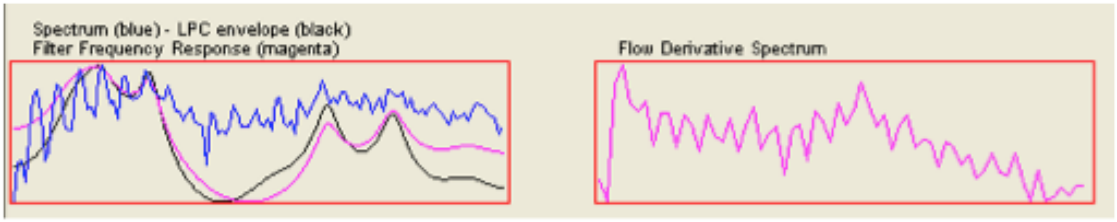


Figure 8. The LPC envelope and flow derivative spectrum including a spurious formant.

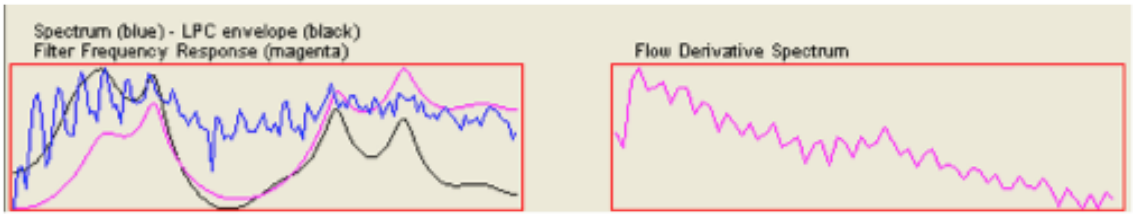


Figure 9. The LPC envelope and flow derivative spectrum of the same sample as figure 8, once the spurious formant has been removed.

The Synthesis program uses a source-filter model of voice synthesis. Once the shape of the vocal tract has been estimated by inverse filtering, Synthesis models the vocal tract response. In these experiments, the harmonic-to-noise ratio was manipulated to vary the level of perceived breathiness in the synthesized vowel. The output of the Synthesis program is an ASCII sound file. The Sky program was used to convert the ASCII files to .wav files. Finally, the .wav files were resampled again to be compatible with E-Prime, the software that presented the stimuli.

Participants

The human subjects committee of the institutional review board at the University of Minnesota approved the procedure for this study. Twenty participants were recruited, with the criteria that they be between the ages of 18 and 50, native speakers of North American English and have no history of speech, language or hearing

disorders. Participants filled out a questionnaire with information about their previous experience with voice training and their history of speech, language and hearing disorders. Of those who volunteered to participate, seventeen were female and three were male. Their mean age was 22, and they ranged in range from 18-35. Six reported some level of voice training, including one or more of the following: choir (4 participants), acting class (2 participants), singing lessons (2 participants) or speech class (1 participant). Two participants who completed the experiment were excluded from analysis due to their reported history of speech therapy for speech sound disorders as children. Two other participants who also completed the experiment were excluded from analysis due to equipment malfunction during the experiment.

Adaptation Experiment

After providing informed consent, participants sat in a sound-treated room and held an AKG C419 III PP lapel condenser microphone approximately 6 inches from their mouth. They repeated the syllable 'ha' four times, sustaining each production for approximately 3-4 seconds. If the participant's voice quality was judged to be rough by the experimenter, a licensed, certified speech therapist with expertise in voice disorders, they were asked to clear their throats, swallow, and repeat the syllable two more times. The productions were recorded on a Marantz CD recorder model CDR300. The participants then waited for 10-20 minutes while the stimuli for the adaptation experiment were prepared. The CD recordings were converted to .wav files using the CDex progam (Free Software Foundation, 2007). They were then resampled to 10 kHz using Praat (Boersma & Weenik, 2010), to accommodate the preferred sampling rate of

the inverse-filtering software. In the INVF program (Kreiman, Gerratt & Antoñanzas-Barroso, 2006), a steady-state section of vowel was selected and was inverse-filtered following the instructions in the software manual. Default settings were used, as described in the manual.

Once the inverse filtering was complete, the vowels were re-synthesized using the Synthesis software package (Kreiman, Gerratt & Antoñanzas-Barroso, 2006). Following the instructions in the manual, default settings were used, except that the length of the synthesized vowel was changed to .5 sec, and the option to use original pitch contours for synthesis was unselected. After re-synthesis, the experimenter listened to the vowel to judge its naturalness. If the result sounded unnatural, the process was repeated from the beginning of the inverse-filtering process until a natural-sounding result was achieved. Once a good synthesized vowel was achieved, the vowel was re-synthesized with four different noise-to-harmonic ratios (NHR's): -20 dB, -17 dB, -14 dB and -11 dB. These levels were chosen because they were perceptible to the researcher as small changes in breathiness. These synthesized vowels were converted from ASCII text files to .wav files using the Sky software package (Kreiman, Gerratt & Antoñanzas-Barroso, 2006). Using Praat, they were then resampled to 11025 Hz to be compatible with the e-Prime experiment software (version 1.2, Psychology Software Tools, 2006) and the mean intensity was set to 65 dB SPL. Figures 10 and 11 show examples of the long-term average spectrum of the added noise for two of the participants, calculated by subtracting the waveform of the -20 dB NHR stimulus .wav file from the waveform of the -11 dB NHR stimulus .wav file using Praat.

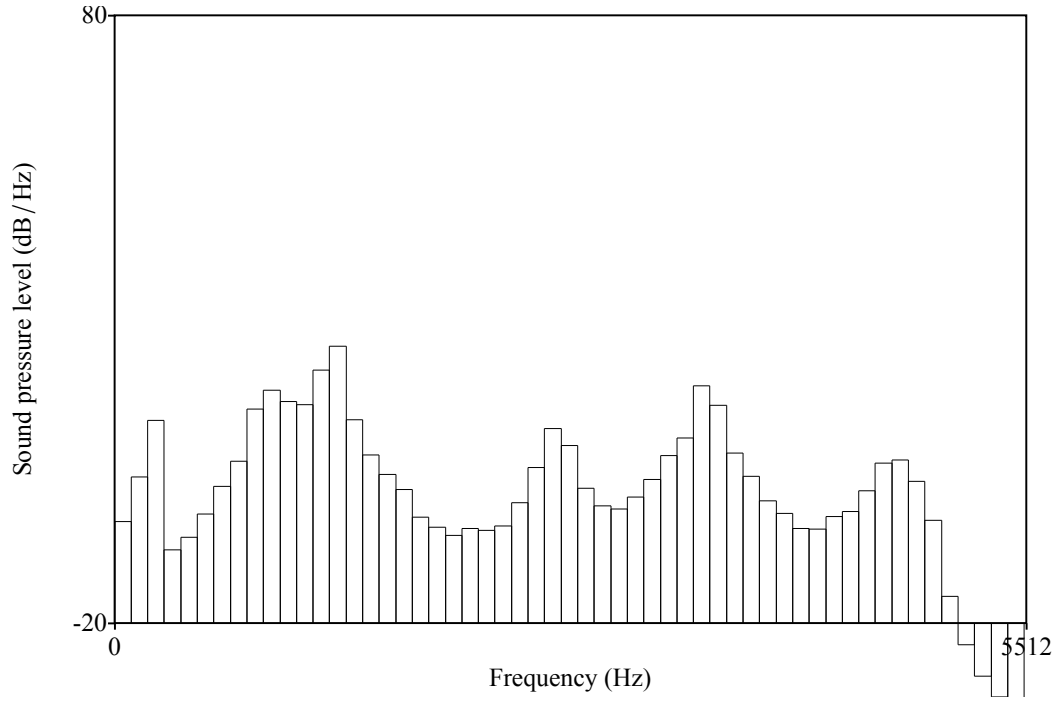


Figure 10. Long-term average spectrum of added noise for participant 1, calculated by subtracting the waveform of the stimulus .wav file with the highest NHR from that of the stimulus .wav file with the lowest NHR.

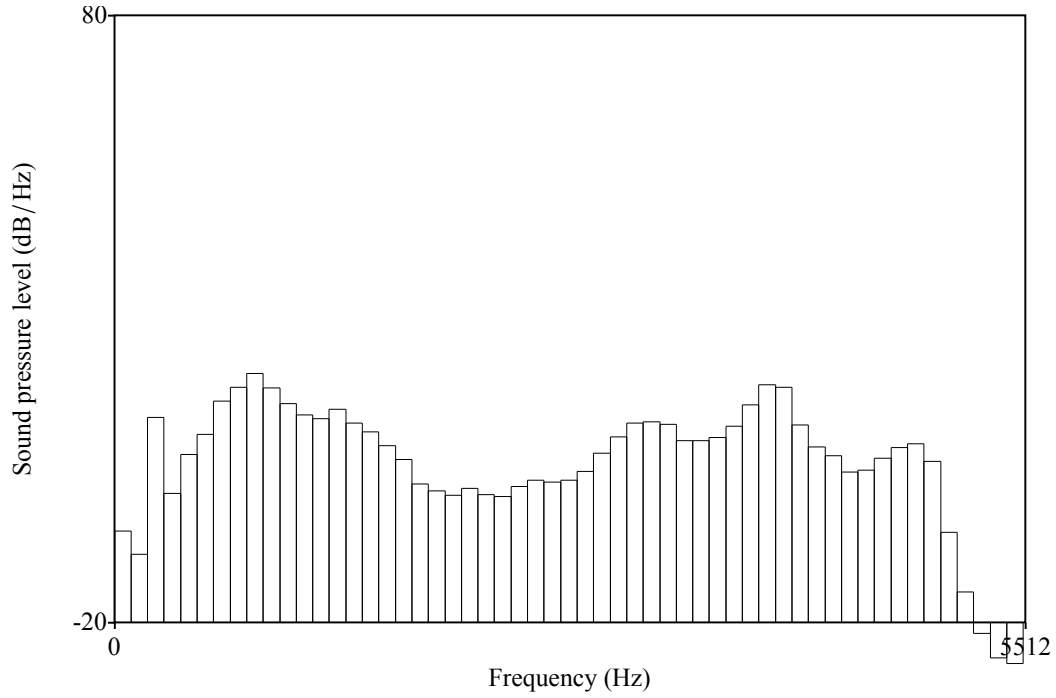


Figure 11. Long-term average spectrum of added noise for participant 3, calculated by subtracting the waveform of the stimulus .wav file with the highest NHR from that of the stimulus .wav file with the lowest NHR.

After the stimuli were prepared, the participants sat in a sound treated room wearing Sennheiser HD 280 pro headphones. The lapel microphone was clipped to their clothing and placed approximately 6 inches from their mouth. They were given the following instructions:

> In this study, you will be producing a series of /ha/ syllables. The goal is to produce them identical in duration, and similarly spaced apart. When you see the word “ha” appear on the screen, say “ha”. When it disappears, stop. Sometimes you will hear noise in the background. You can ignore the noise.
>
> Sometimes, instead of “ha”, the screen will say “hey” instead. The word “hey” will be in red to help you notice that it has changed. You will get a countdown to

help you start. The last "ha" will be in blue to help you know that it is time to stop. There will be two sets of practice items before the experiment starts.

Participants completed a set of ten practice items twice before the experiment started, with seven tokens of /ha/ and three of /he/. The experimenter remained in the room with the participants during the first set of practice items to ensure that the participant understood the task. The participants' productions were recorded on the left channel of the Marantz CD recorder described above. The stimuli, which were presented over the headphones, were simultaneously recorded on the right channel.

There were two phases to the experiment: a control phase and an experiment phase. The phases were counterbalanced, with half of the participants completing the control phase first, and half completing the experiment phase first. In the control phase, participants repeated the syllable /ha/ 280 times. The prompt to say "ha" was presented at a rate of once per second and a duration of .5 seconds. The first ten baseline trials were presented with no noise over the headphones. The remaining 270 control trials were conducted with a 1-second recording of speech babble repeated continuously, which was presented over the headphones at 65 dB SPL.

During the experimental condition, instead of speech babble the participants heard the stimuli prepared from their own voice sample at the beginning of the session, as described above. The stimuli were .5 seconds in duration, and were presented once per second. The participants repeated the syllable /ha/ during the period when the stimulus was presented. This condition contained the following segments in this order: 10 baseline trials with no noise, a ramping-up phase with 10 trials each at -20 dB, -17

dB, and -14 dB NHR, a training phase with 100 trials at -11 dB NHR, an adaptation phase with 10 catch trials with no noise randomly mixed with 50 training trials at -11 dB. After the catch trials, participants repeated 10 generalization trials of 'hey' with no noise. Then came the ramp-down condition with 10 trials each of -11 dB, -14 dB, -17 dB and -20 dB NHR. Then came a return-to-baseline phase of 30 trials with no noise. After the return-to-baseline trials, participants completed the perception experiment described below. Finally, an additional baseline condition of 30 additional trials with no noise was completed. Total phonation time for the entire experiment was approximately 5 minutes. Figure 10 shows a schematic of the order of presentation of each element in the experiment.

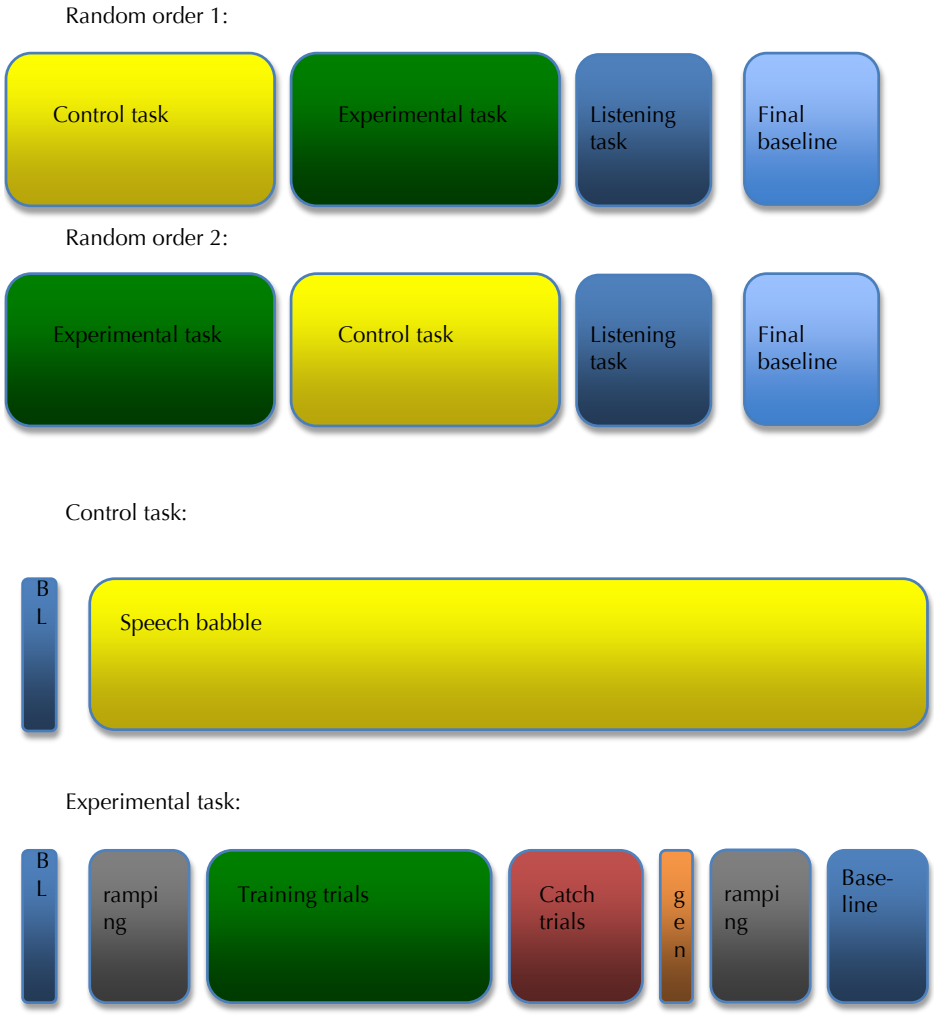


Figure 12. Order of presentation of experimental tasks.

Analyses

Once the experiment was completed, the CD sound files were converted to .wav files using CDex. The left channel, with the participants' productions was extracted using Praat software. The left channel was opened as a sound file, and the beginning and end of each vowel (excluding the /h/ portion of each utterance) was marked in a text grid according to the following rules: mark the beginning of the production just after the

start of voicing, mark the end of the production just before the end of strong formants. An example is shown in figure 13.

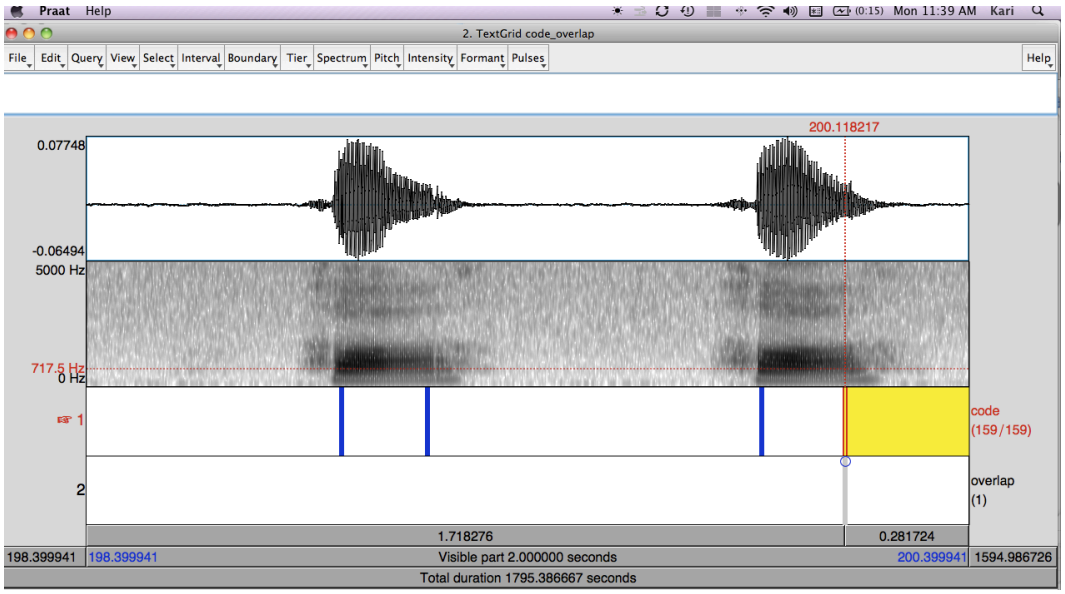


Figure 13. Markers were manually placed at the beginning and ending of voicing for each utterance.

The utterances were coded using both the track containing the participants' production and the track containing the stimuli. Each utterance was coded according to the type of noise that was present in the right track, as shown in table 1. Placement of the codes in the textgrid is shown in figure 14.

Table 1. Coding labels for each experimental condition.

Condition	**Type and level of noise**	**Code**	**Number of trials**
control baseline	none	b0	10
control	babble	c	280

First baseline	None	b1	10
1st rampup	.5 sec, -20 dB NHR	-20a	10
2nd rampup	.5 sec, -17 dB NHR	-17a	10
3rd rampup	.5 sec, -14 dB NHR	-14a	10
training	.5 sec, -11 dB NHR	-11	100
catch trials	none	ca	10 mixed in with 50 training trials
generalization	none	hey	10
1st rampdown	.5 sec, -11 dB NHR	-11	10
2nd rampdown	.5 sec, -14 dB NHR	-14b	10
3rd rampdown	.5 sec, -17 dB NHR	-17b	10
4th rampdown	.5 sec, -20 dB NHR	-20b	10
2nd baseline	none	b2	30
final baseline	none	b3	30

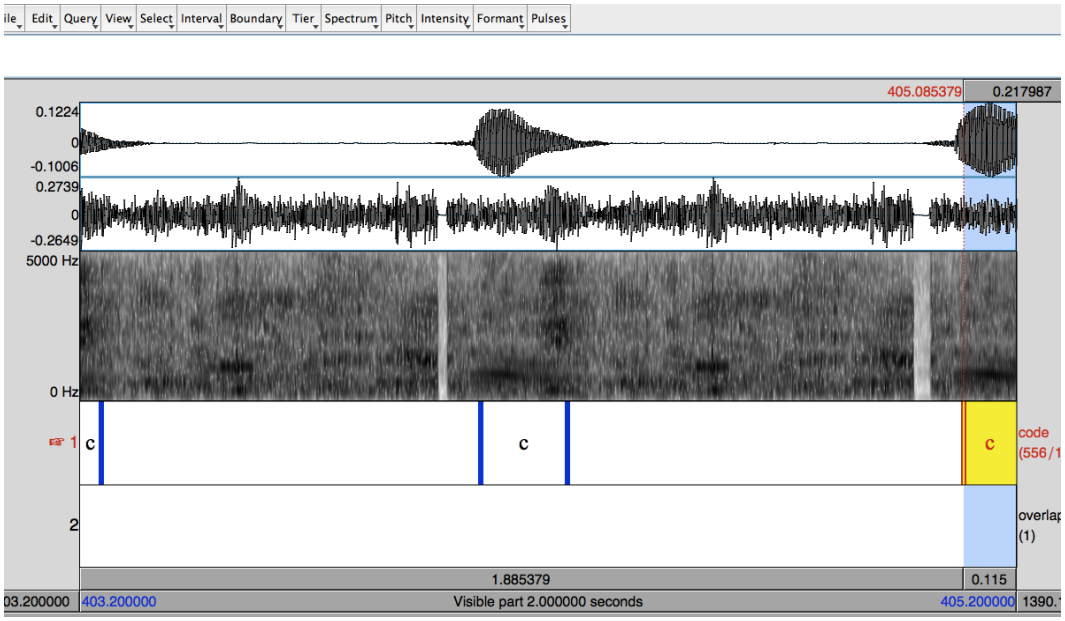


Figure 14. A coding marker placed in the textgrid during an interval of voicing, in this case indicating the control condition.

The participants' productions were extracted into separate soundfiles in Praat. Next, the middle 50 ms was extracted from each soundfile for analysis. The acoustic analysis was performed using VoiceSauce (Shue, 2011). VoiceSauce makes measurements of 25 acoustic parameters at 1 ms intervals. The frequency of a signal cannot be calculated within one period of the beginning or the end of the sound file, therefore VoiceSauce calculated the spectral slant, H1-H2c, for the middle 7 ms of each sound file. H1-H2 is the difference in amplitude between the fundamental frequency (the first harmonic, H1) and the first overtone (the second harmonic, H2). H1-H2c corrects for the effects of the vocal tract filter, which allows the measure to be compared across different vowels. These measurements were averaged to give an average H1-H2c for each token produced by each participant. Average values were then calculated for each condition calculated first within and then across participants. H1-H2

was chosen as a dependent measure because it has been shown to correlate with perceptual ratings of breathiness (Hillenbrand & Houde, 1996), and has been shown to be a good predictor of voice quality in languages where creaky and breathy voice is phonologically contrastive, as in Green Hmong (Andruski & Ratliff, 2000) and White Hmong (Garellek, Keating, Esposito, & Kreiman, 2013).

Finally, to assess how well the participants were able to follow the instructions for the experiment phase and vocalize only during the intervals where noise was present, on the trials with .5 second noise only, the portion of voiced production that was present in the intervals of silence between the noise were marked and coded ‘o’ on a separate tier as shown in figure 15.

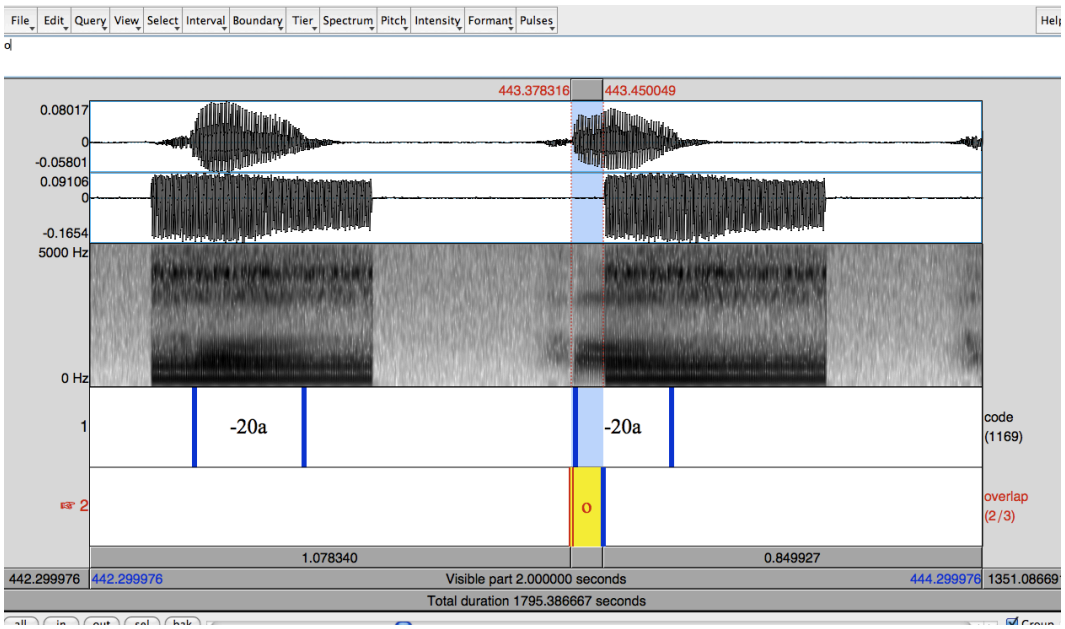


Figure 15. The blue section, coded as 'o' measures the length of time that the participant was voicing during the silent period between the intervals of simulated breathiness.

Perception Experiment

The purpose of the perception experiment was to determine whether variability in participants' perception of breathiness accounts for some of the variability in their adaptive response to auditory perturbation of their voice quality. For this experiment, two male voices and two female voices were recorded, inverse-filtered and re-synthesized by the same process as the adaptation experiment above. Stimuli were prepared for each voice at 5 noise-to-harmonic ratios: -25 dB (the default setting of the synthesis software), and -20 dB, -17 dB, -14 dB and -11 dB, the same levels used in the adaptation experiment. The stimuli were presented in pairs. For each speaker, each noise level was paired with every other noise level in both possible orders, resulting in 80 stimulus pairs. The stimuli were presented using e-Prime software over headphones as in the adaptation experiment. The participants listened to each pair of stimuli in random order, with each pair presented twice in each order. They pressed the number "1" on a keyboard if they thought the first stimulus was breathier, or the number "2" if they thought the second was breathier.

The participants' responses were compared to the synthesized noise-to-harmonic level. If the participant indicated that the stimulus with the higher NHR was breathier, the response was marked as correct. For example, if a stimulus with an NHR of -17 dB was presented first, and a stimulus with an NHR of -11 dB was presented second, a response of "2" would be coded as correct, while a response of "1" would be coded as incorrect. The number of correct responses for each participant was tallied in categories according to the number of NHR scale steps between the stimuli. Because there were

different numbers of stimuli pairs in each scale step category, percent correct was calculated based on the total number of stimulus pairs presented in each category. Percent correct in the 1 scale step group was used as the measure of perceptual discrimination, as it had the broadest range of responses.

Table 2. Noise levels of stimulus pairs presented in the perception experiment.

Scale steps	**1**	**2**	**3**	**4**
NHR of stimuli (dB)	-11 and -14 -14 and -11 -14 and -17 -17 and -14 -17 and -20 -20 and -17 -20 and -25 -25 and -20	-11 and -17 -17 and -11 -14 and -20 -20 and -14 -17 and -25 -25 and -17	-11 and -20 -20 and -11 -14 and -25 -25 and -14	-11 and -25 -25 and -11
Total # of stimulus pairs	64	48	32	16

Results

This experiment aims to answer three questions:

1. Do participants change their vocal behavior when their perception of their voice changes?
2. If so, do those behavior changes demonstrate adaptive learning?

3. Do differences in perceptual discrimination of breathiness predict differences in how participants respond to perceptual perturbations?

Question 1: Do Participants Change their Vocal Behavior when their Perception of their Voice Changes?

Looking at the 16 participants individually, the average value of H1-H2c in the training condition, in which the simulated breathiness was present, was compared to that of the control condition, which presented speech babble at the same intensity. Independent-samples t-tests between the training and control conditions of individual participants showed that 12 of the 16 participants did have significantly different values of H1-H2c across these two conditions. Table 3 contains individual *t* and *p* values for each participant. Those that changed their behavior will be referred to as *responders* and those with no significant difference in H1-H2c between the control and training conditions will be referred to as *non-responders*.

Larson (1998) found that, when there was no phonemic or prosodic context present, most participants compensated for a perceptual shift in F0 by shifting their F0 in the opposite direction as the perceptual shift, but some shifted in the same direction. Because breathiness and pressed quality do not carry contrastive phonemic or prosodic meaning in English, we expected that shifts in H1-H2c in both directions were possible, with some participants becoming breathier, and some becoming more pressed.

When participants in this study heard their voice with simulated breathiness, 7 shifted their voice quality to compensate for the auditory perturbation and their voice quality became more pressed (*compensators*), 5 shifted in the same direction as the

perturbation and their voice quality became breathier (*followers*), and 4 showed no significant difference in H1-H2c between the control and training conditions (*non-responders*). Of the responders, when comparing the difference in H1-H2c between the control and training conditions, all had values of Cohen's d larger than .25.

Table 3. Summary of individual participant data.

Subject	Sex	Initial quality	Response type	*p*-value of response to perturbation	*t*-value of response to training	Mean H1-H2c control	Mean H1-H2c training	Cohen's d	Effect size >.25	Scaled mean H1-H2c control	Scaled mean H1-H2c training	Perception % correct at 1 step difference	Total seconds error time
1	M	pressed	follower	<.001	4.92	-4.24	-3.67	-0.48	yes	0.24	0.35	65.63	14.17
3	F	breathy	non-responder	0.463	0.73	2.42	2.63	-0.08	no	4.41	4.70	62.50	7.22
4	F	breathy	follower	0.007	2.74	4.50	5.93	-0.28	yes	0.58	0.44	68.75	3.24
5	M	pressed	follower	<.001	19.91	-3.85	-0.92	-1.96	yes	0.01	0.77	87.50	5.58
6	F	breathy	non-responder	0.069	1.82	2.94	3.78	-0.26	yes	0.82	0.77	78.13	0.97
8	F	pressed	compensator	<.001	3.98	-0.85	-1.55	0.35	yes	0.52	0.12	79.69	2.58
9	F	breathy	compensator	<.001	27.18	3.51	-0.10	2.52	yes	0.10	1.03	73.44	8.02
10	F	breathy	compensator	<.001	13.26	3.71	0.82	1.32	yes	1.23	0.51	71.88	7.52
11	F	breathy	follower	<.001	10.36	1.74	4.09	-1.07	yes	0.26	1.97	70.31	4.18
12	F	breathy	non-responder	0.224	1.22	6.61	6.89	-0.11	no	0.22	0.28	70.31	7.44
13	F	breathy	non-responder	0.686	0.40	3.22	3.37	-0.04	no	0.61	0.59	81.25	7.18
14	F	breathy	follower	<.001	6.12	3.44	4.97	-0.62	yes	0.12	0.28	67.19	1.68
16	F	breathy	compensator	<.001	4.79	2.87	2.43	0.49	yes	1.76	1.33	89.06	1.29
17	M	breathy	compensator	<.001	15.50	2.22	-2.94	2.78	yes	0.07	2.41	75.00	3.54
18	F	pressed	compensator	<.001	7.60	-0.12	-1.90	0.74	yes	0.86	1.18	68.75	11.16
19	F	breathy	compensator	0.004	2.91	5.63	4.93	0.37	yes	0.30	0.39	78.13	12.39

Figures 16-18 show scatterplots of the values of H1-H2c for each token arranged in order of presentation for three participants who showed three different patterns of behavior (for the sake of clarity, all three participants selected to be shown here did the control condition first). Participant 9 (fig. 16) was a compensator, and was more pressed during the training condition than during the control condition ($t(429) = 27.18$, $p < .001$, Cohen's $d = 2.52$). Participant 1 (fig. 17) was a follower, and was more breathy during

the training condition ($t(401) = 4.92$, $p < .001$, Cohen's $d = -0.48$). Participant 3 (fig. 18) was a non-responder, and showed no significant difference between the two conditions ($t(425) = 0.73$, $p = .46$). Scatterplots for all 16 participants are included in the appendix.

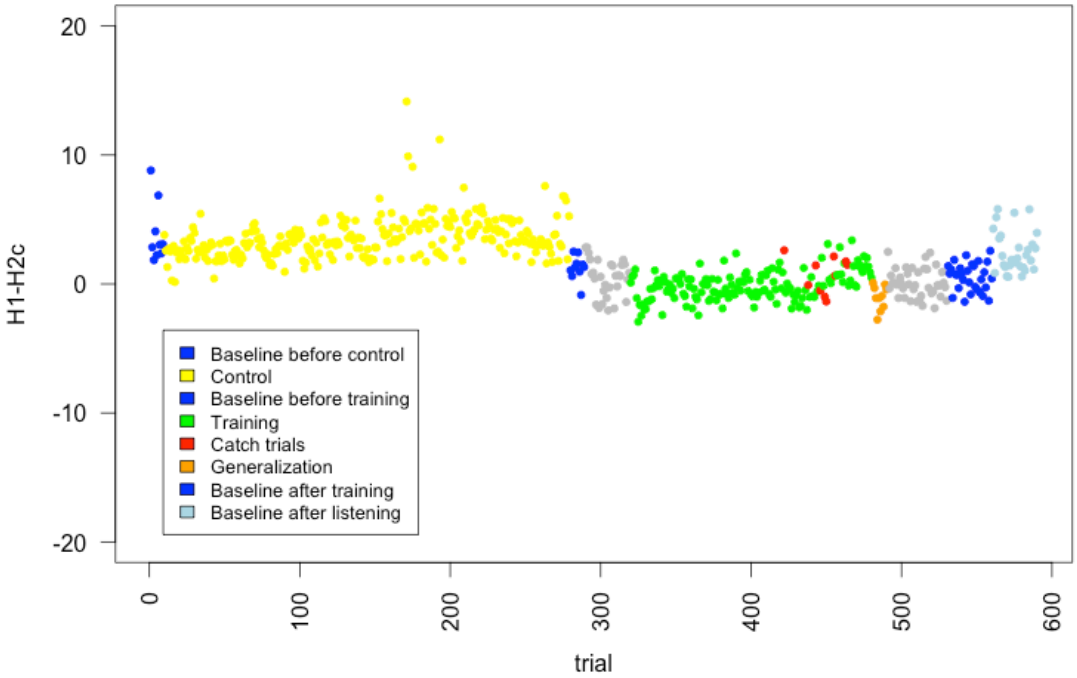


Figure 16. Value of H1-H2c by trial for participant 9. Values of H1-H2c are lower during the control trials than during the training trials.

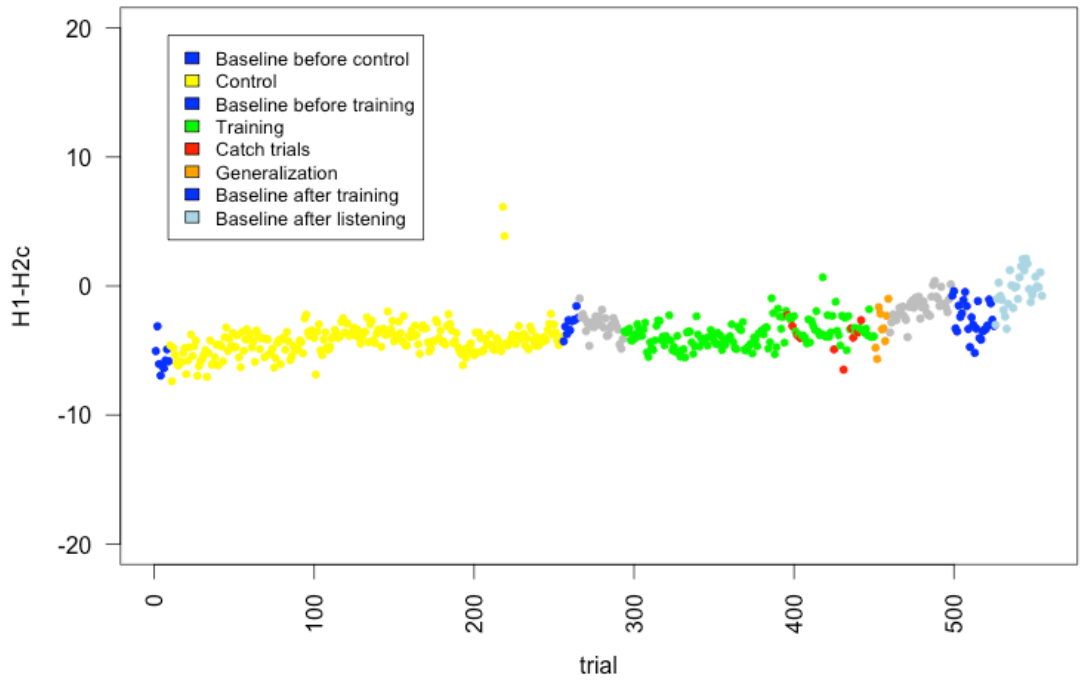


Figure 17. Value of H1-H2c by trial for participant 1. Values of H1-H2c are higher during the training trials than the control trials.

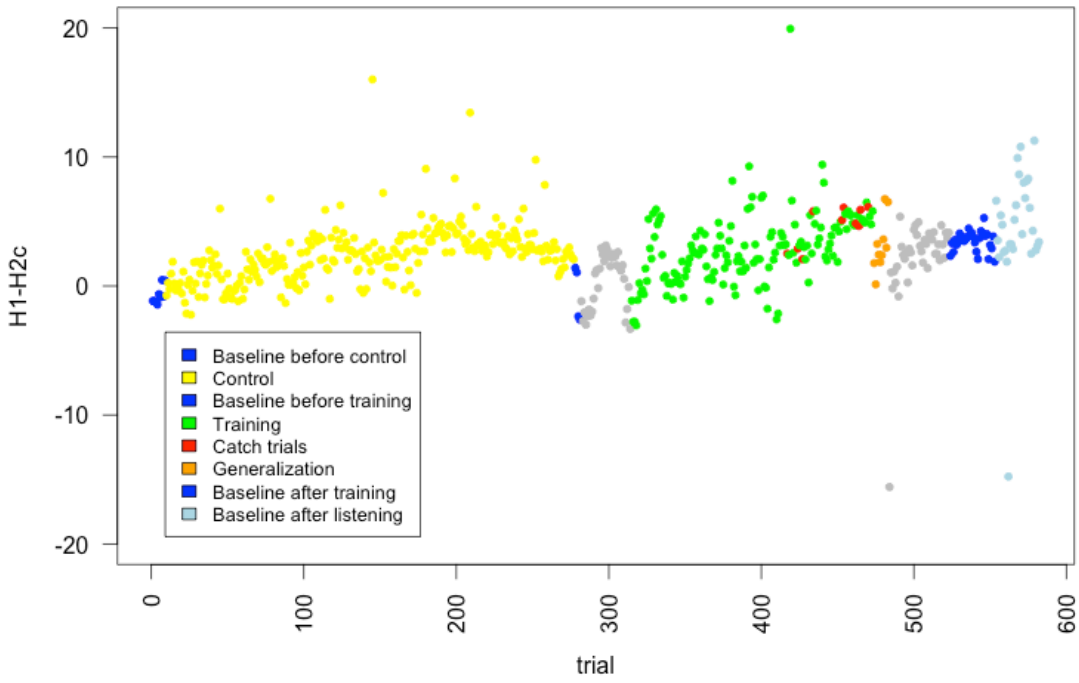


Figure 18. Value of H1-H2c by trial for subject 3. There is no significant difference in values of H1-H2c between the training trials and the control trials.

Group analysis.

When all 16 participants were analyzed together as a group with Friedman's ANOVA by ranks, none of the differences in H1-H2c between any of the experiment conditions were statistically significant: $F[7,105] = 0.265$, $p = 0.966$. Figure 19 shows the values of H1-H2c averaged across participants (the ramping conditions are not included in this figure). None of the conditions were significantly different from any of the others.

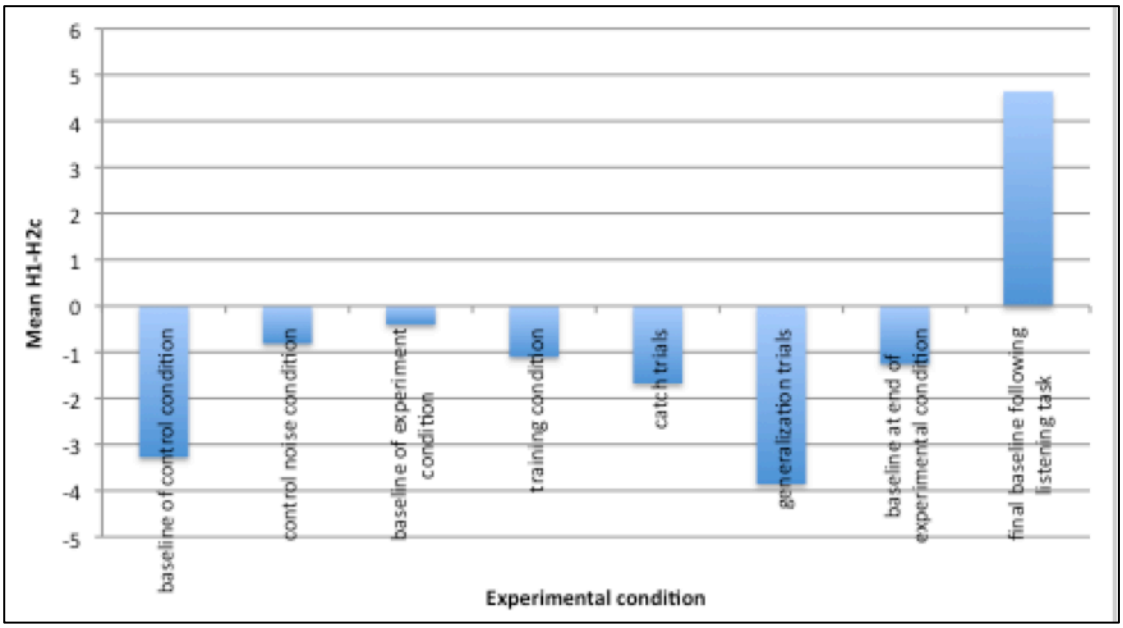


Figure 19. Mean value of H1-H2c across subjects by condition. None of the differences are statistically significant.

Because roughly half of the responders were compensators and half were followers, the change in H1-H2c of the two groups would tend to cancel each other out when averaged across the group. A scaled value of the change in H1-H2c was calculated for each subject to look at the magnitude of change, independent of the direction of change. The scaled value measured the absolute value of the difference between the mean of the initial baseline condition and the mean of each other condition. This measures the magnitude of change from baseline for each condition. It was calculated as follows: $\bar{x}'_{condition} = \left| (\bar{x}_{condition} - \bar{x}_{baseline})/\bar{x}_{baseline} \right|$ where $x_{baseline}$ refers to the first baseline measure: the baseline before the control condition for odd-numbered participants, and the baseline before the experiment condition for even-numbered participants. Comparing the two conditions where we expected the greatest difference,

the control condition vs. the training condition, a paired-samples t-test showed no significant difference between these two conditions ($t(15) = 1.57$, $p = .136$). Figure 20 shows the mean scaled value of H1-H2c across subjects by experimental condition. Figures 21-23 show boxplots of the mean values of H1-H2c by condition for the three response types: compensators, followers and non-responders.

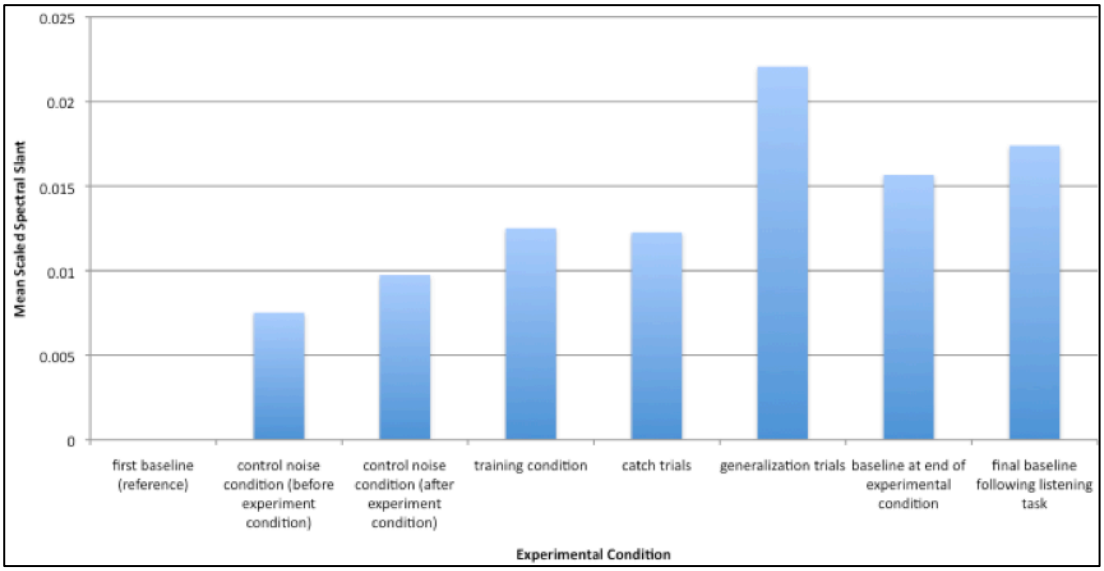


Figure 20. Mean values of H1-H2c across subjects. The values are scaled as the magnitude of change in either direction from the first baseline condition. None of the differences were statistically significant.

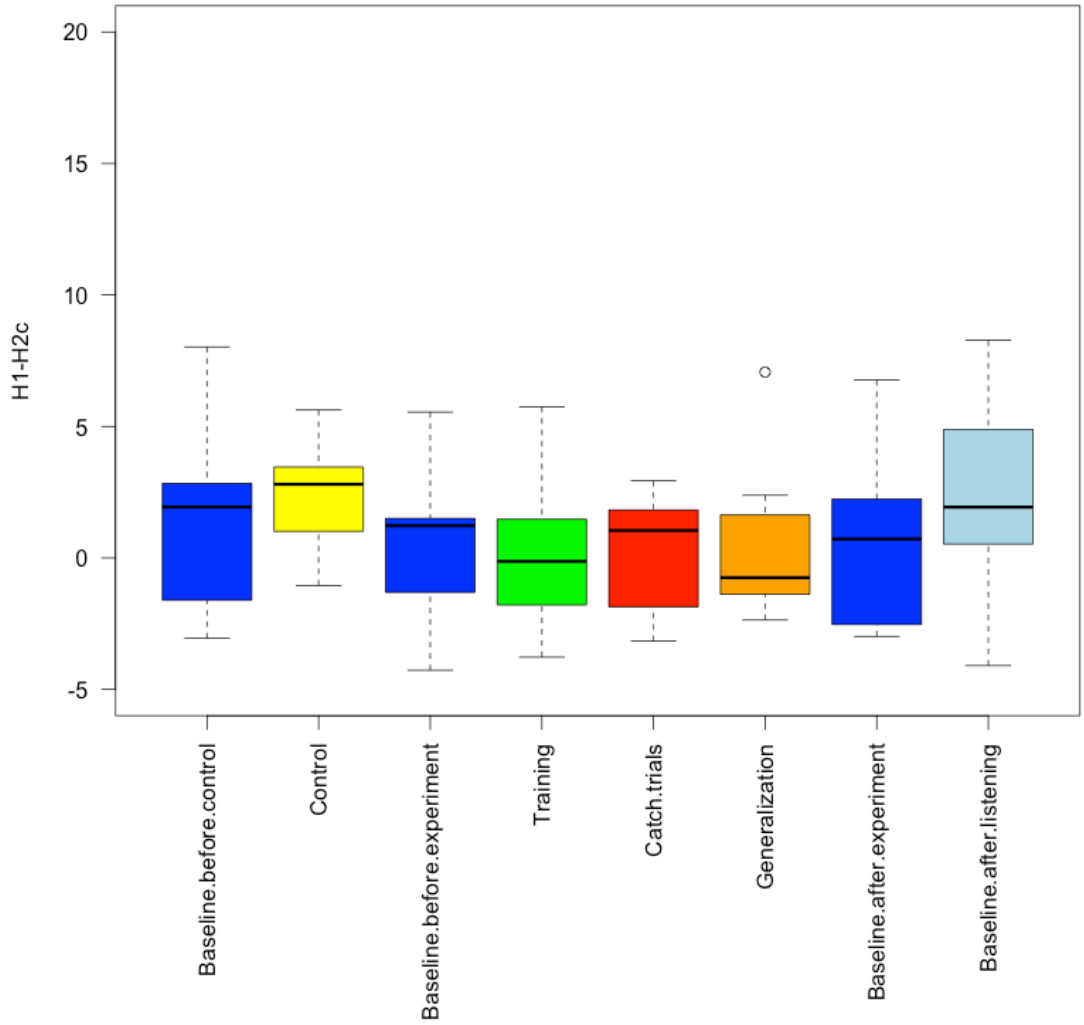


Figure 21. Boxplot of mean H1-H2c by condition for compensators (N=7).

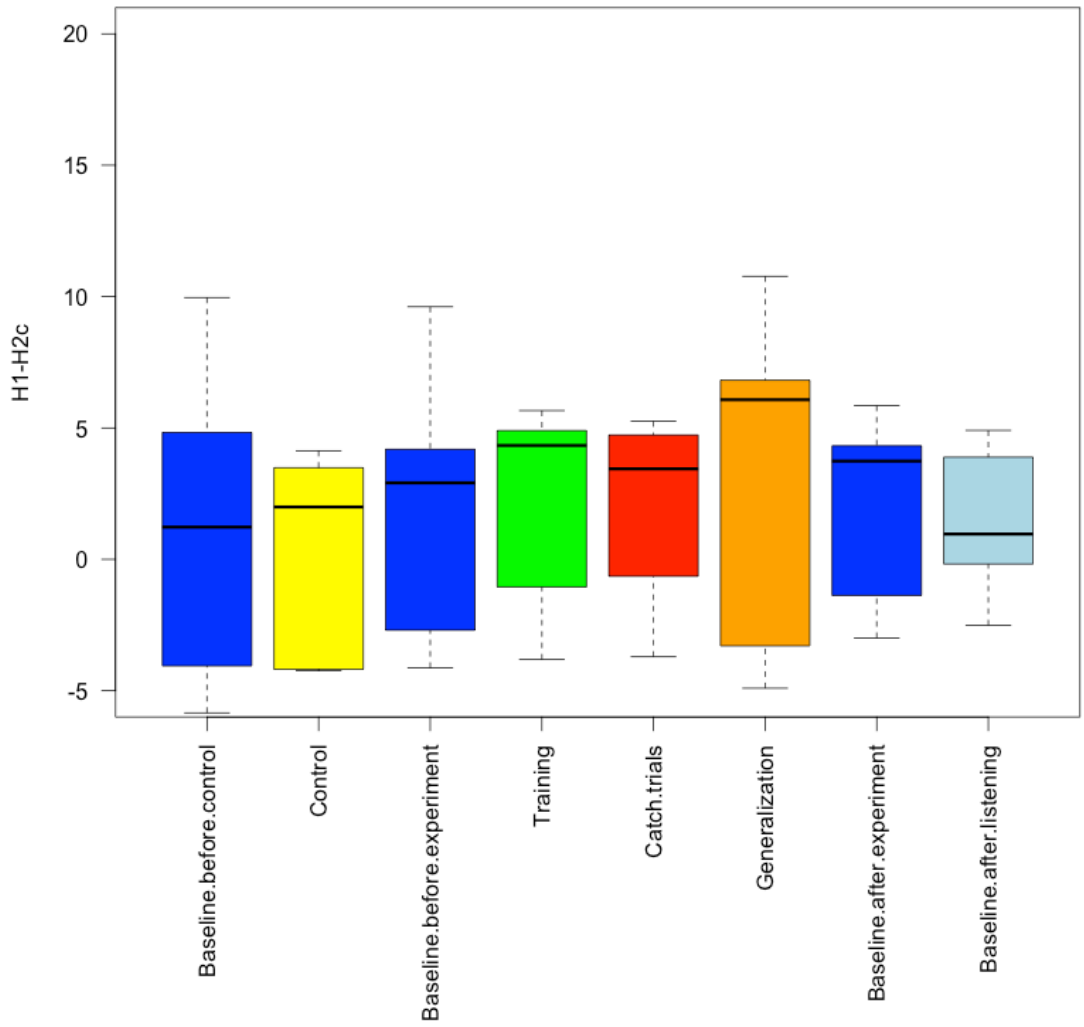


Figure 22. Boxplot of mean H1-H2c by condition for followers (N=5).

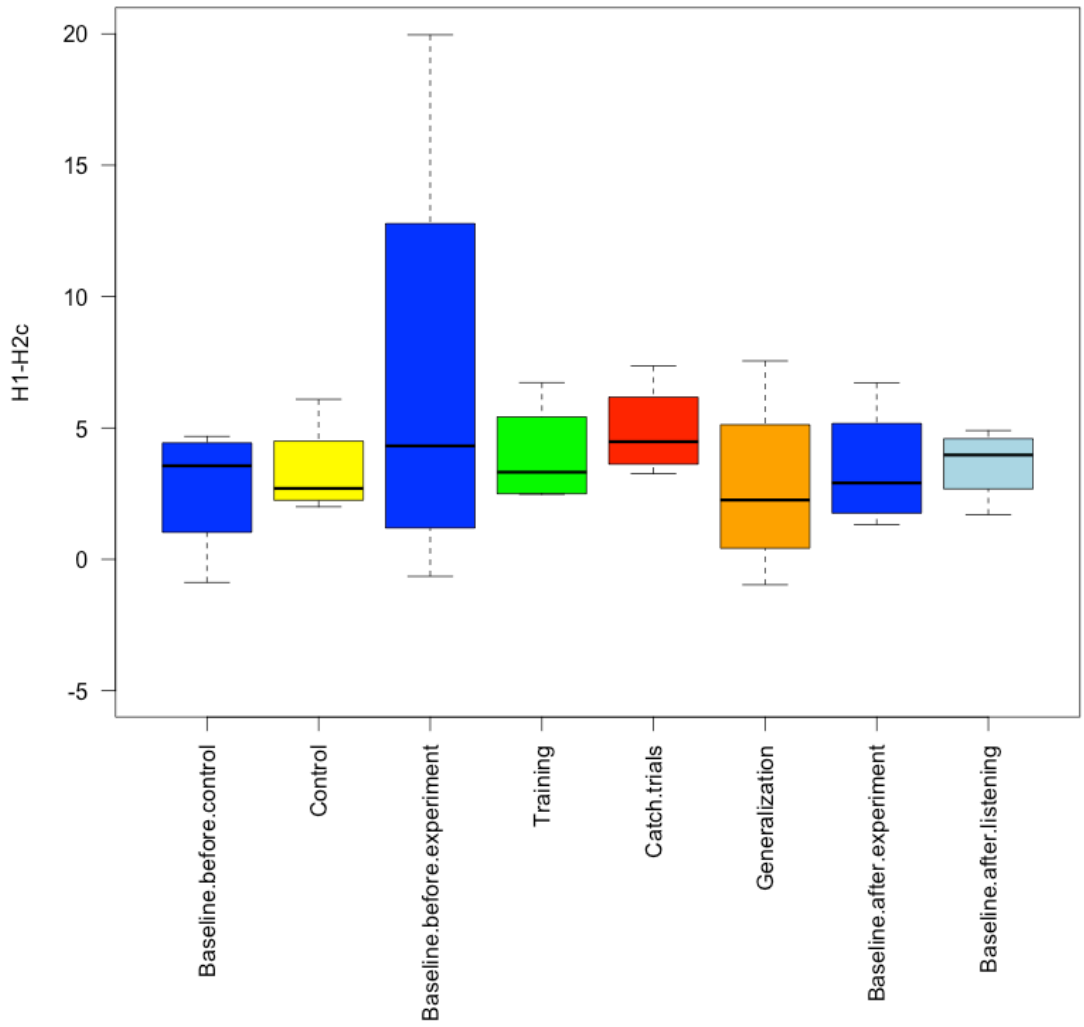


Figure 23. Boxplot of mean H1-H2c by condition for non-responders (N=4).

Question 2: Do the Observed Changes Show Evidence of Adaptive Learning?

In order to demonstrate adaptive learning, it is necessary to show that behavior changes caused by a perturbation in perception persist transiently once the perturbation is removed. In the case of this experiment, we would need to show that voice quality on the catch trials, in which the breathy noise was randomly turned off, was more similar to the voice quality of the training trials, in which the breathy noise was present, than to

that of the control trials. Because none of the group effects were significant, this experiment did not provide evidence of adaptive learning of voice quality behavior.

Question 3: Do Differences in Perceptual Discrimination of Breathiness Predict Differences in how Participants Respond to Perceptual Perturbations?

We predicted that participants with better discrimination of breathiness would change their behavior more than those with worse discrimination when their auditory perception was perturbed. The discrimination task involved a paired comparison, where the participants were asked which of two stimuli was breathier. The discrimination of breathiness was measured for each participant as the percent of accurate responses when the difference in NHR was smallest (3-5 dB). This measure was chosen because there were ceiling effects when the noise difference was larger. The hypothesis was tested with a linear regression analysis to determine whether performance on the discrimination task predicted the amount of difference in spectral slant between the training and control tasks. The results showed no relationship between the two variables: $r^2 = 0.046$, $F(1,14) = 0.68$, $p = 0.42$. Also, there was no difference in the average perception score between responders and non-responders $t(34) = 0.35$, $p = 0.73$. Figure 24 shows perceptual acuity for each participant as a function of difference in H1-H2c between the control and training tasks.

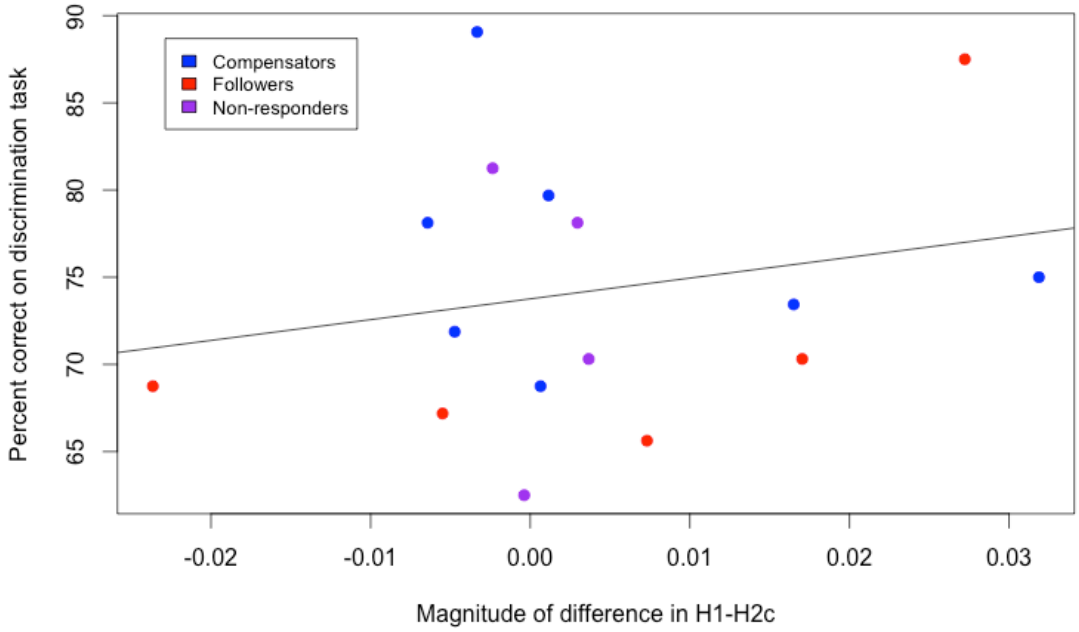


Figure 24. Relationship between perceptual acuity and response to auditory perturbation by subject. The x-axis represents the absolute value of the difference between mean H1-H2c for the control condition vs. the training condition. The y-axis represents the percent accuracy on the discrimination task when the difference between the stimuli was smallest. Pearson's correlation between the variables was $r = .03$.

Other Factors Affecting Individual Performance

Because the group effects were not significant, other factors were examined to see if they accounted for some of the variability in participant's responses, either in the direction or the magnitude of change. The small number of participants in each group (7 compensators, 5 followers and 4 non-responders) made generalizations difficult.

One question was whether participants who were better at performing the task, that is who were able to vocalize only during the intervals where the sensory

perturbation was present, would respond more than those who vocalized more in the intervals between the sensory perturbation noise. During coding, the amount of time that participants vocalized was marked in the textgrid, on a separate tier, marked "error" as described in the previous chapter. The total amount of time spent vocalizing outside of the interval (*error time*) was calculated for each participant. The amount of error time for individual participants ranged from 0.97 seconds to 14.17 seconds. The mean error time for non-responders was 5.70 seconds, and the mean for responders was 6.28 seconds. This difference was not statistically significant, $t(15) = 0.24$, $p = 0.81$.

Another possibility was that participants who were already breathy would be more likely to become pressed, and vice versa, as they were already closer to one edge of the motor space, and they have more "room" to move toward the other direction. Responders who initially had breathy voices (H1-H2c at first baseline > 1, N=8) and those that initially had pressed voices (H1-H2c at first baseline < 1, N=4) were equally likely to change in either direction, that is, they were equally likely to be compensators or followers ($\chi^2(2, N=16) = 1.15$, $p = .56$). All of non-responders started out with breathy voices ($\chi^2(2, N=16) = 8.95$, $p = .011$), however given the number of analyses performed, this is likely to be a spurious result. This finding may be explained simply by the fact that most of the participants were female speakers of North American English, and in that dialect, females tend to have breathier voices (Klatt & Klatt, 1990).

A third possibility is that participants' voice quality would be more variable during the training task than during the control task, as they tried different motor strategies to try to achieve the perceptual result they expected. Participant 19 (fig. 26) is

a particularly dramatic example of this. The scatterplot shows much more variability in H1-H2c in the training trials than in the control trials. If this pattern were true in general, then we would expect that when averaged across participants, the standard deviation of H1-H2c of the training task to be significantly higher than that of the control task. The standard deviation of the training and control trials were calculated for each participant. For the control condition, the average standard deviation across participants was 2.16, and the average for the training condition was 2.48. Although the standard deviation was slightly higher for the training condition, a paired-samples *t*-test showed this difference was not statistically significant. ($t(15) = .94$, $p = .36$).

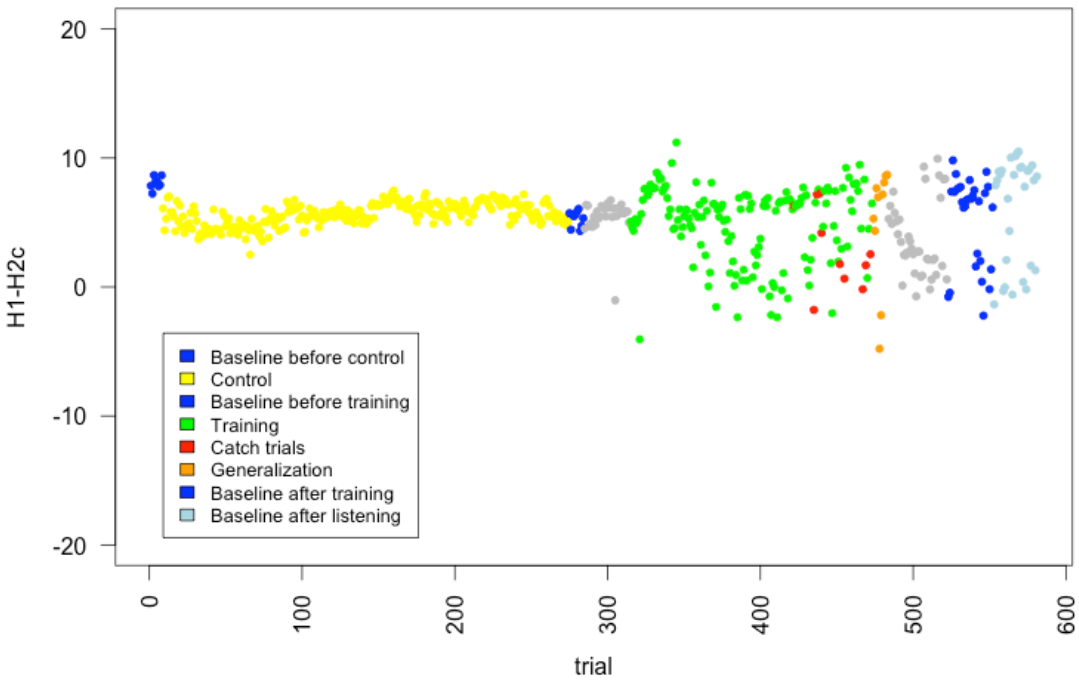


Figure 25. Value of H1-H2c for participant 19, in order of presentation. The variability in H1-H2c is much higher on the training taskthan on the control task.

Reliability

To determine the reliability of the measurements, one odd-numbered and one even-numbered participant were chosen randomly (participants 17 and 14), and were re-textgridded and re-coded following the same procedure as the original coding as described in the previous chapter. The new textgrids were then analyzed by the same procedures. For each participant, the re-calculated value of H1H2c for each token was correlated with the originally calculated value. This is a measure of how sensitive the calculated value of H1H2c of each token is to the exact placement of the boundaries of the soundwave. The results varied greatly between participants. For participant 17, Pearson's $r = 0.99$, a very high reliability, whereas for subject 14 $r = 0.48$, indicating much lower reliability. A third participant (participant 12) was re-analyzed, and had an even lower correlation, $r = 0.26$. This suggests that H1H2c as calculated in this experiment is a fairly noisy measure, and the value calculated for each token will change based on exactly where the boundaries are placed. This may be due to the fact that we were eliciting dysphonia in the participants' voices. We would expect H1H2c to be less reliable for dysphonic voices than for good voice quality, as H1H2c measures the amplitude of the first two harmonics, which has more cycle-to-cycle variability in dysphonic voices than in normal voices. It may be necessary to use larger number of trials, especially of catch trials, to be able to detect a signal if it is present.

Discussion

This experiment was designed to take initial steps to show that adaptive learning occurs in the area of behaviors that affect voice quality. Specifically, it examined three

hypotheses. The first is that when participants were presented with perturbed auditory feedback consisting of an increase in noise-to-harmonic ratio, intended to simulate breathiness, they would respond by either increasing the medial compression of the vocal folds in a compensatory response, or decreasing it in a following response. The dependent measure was H1-H2c, which is correlated with the open quotient of the vocal folds. This effect was seen in 12 out of 16 participants, suggesting that most participants did change their voice quality as a result of the auditory perturbation.

The second hypothesis was that these changes in behavior would show evidence of motor learning. To do so would require a statistically significant group effect in which the catch trials (in which the perturbation was turned off in an unpredictable pattern) were more similar to the training trials (when the perturbation was present) than to the control trials (which presented speech babble instead of altered feedback). As none of the group effects were statistically significant, this was not demonstrated.

The third hypothesis was that participants with better discrimination of differences in breathiness would respond more to the perturbation than those with less acute discrimination. This was tested by comparing the response to the auditory perturbation to participants' performance on a paired-comparison discrimination task of different levels of breathiness. Linear regression showed no relationship between performance on the discrimination task and response to the auditory perturbation, so this hypothesis was not supported by the data.

Although this experiment did not provide evidence of motor learning, it did provide evidence of changes in H1-H2c in the presence of auditory perturbation of

noise-to-harmonic ratio, suggesting that participants changed their voice quality in response to the perturbation. There are a number of possible explanations for this finding.

The first possible explanation is that although feedback motor changes occur, adaptive learning does not occur in the domain of behaviors related to voice quality. Given the large body of evidence for adaptive learning in other domains of motor behavior, including speech motor behavior, this explanation is unlikely, and would require further replication to verify.

A second possible explanation is that adaptive learning occurs in the domain of behaviors related to voice quality, but this experiment did not elicit adaptive learning. This could be due to problems with the design of the experiment. For example, there might have been an insufficient number of training trials, insufficient or too much change in the level of breathiness, or a ramping-up of the perturbation that went too quickly. In most cases, experiments that have examined the effects of these methodological variables have not found them to greatly affect the outcome of the experiment. MacDonald, Goldberg, and Munhall (2010) examining formant perturbations, found that motor learning did not depend on the rate of change of the perturbation, and Purcell and Munhall (2006), also examining formant perturbations found that neither the direction of change nor the number of trials affected motor learning. MacDonald et al. did find, however, that the magnitude of change did matter. When the perturbation was too large, the compensatory response decreased, although

compensation did still occur. Larson (1998) found that increasing the magnitude of the perturbation increased the number of following responses.

A third possibility is that the changes in voice quality observed in this experiment are due to vocal fatigue, and not to the perceptual perturbation. Although the experimental conditions were counterbalanced across participants to control for vocal fatigue, since the group effects were not significant, it is possible that the effects that we did see were the results of vocal fatigue. The results of the current experiment do not distinguish between these possibilities.

A fourth explanation is that the participants responded to the auditory perturbation in ways that were not captured by measuring H1-H2c. Voice quality has more dimensions than pitch or formant frequencies. Possible ways that listeners might respond to perturbations in voice quality could include changes to respiratory pressure, medial compression of the vocal folds, elevation or lowering of the larynx, or changes in the position of the tongue or velum, and these may not be reflected in the values of H1-H2c.

Limitations of the Study

H1-H2c as a dependent measure.

Although H1-H2 is a widely used as an acoustic correlate of open quotient, it is not as strong a correlation as, for example, pitch is with fundamental frequency (Kreiman et al., 2008). A number of researchers have attempted to find instrumental measures that distinguish between patients with MTD and normal controls. Zheng (2010) examined a number of aerodynamic measures, and found that only subglottal

pressure was able to distinguish the two groups. Stepp (2010, 2011) found that relative fundamental frequency (RFF), the change in fundamental frequency at the onset/offset of vowels distinguished the two groups. Redenbaugh and Reich (1989) and Hocevar-Boltezar et al. (1998) found that surface EMG was able to distinguish the groups, although Van Houtte et. al (2011) found no difference in sEMG measures between MTD and normal groups. All of these studies looked at group differences, and none of them have shown that these measures can distinguish degree of dysphonia. Kreiman (2010) stated "Despite many years of research, we still do not know how to measure vocal quality (p. 62)." It is possible that the tools currently available to measure voice quality are not yet sensitive enough to reliably and sensitively detect the changes in voice quality that would occur with adaptive learning.

Lack of real-time manipulation.

Because all of the stimuli were pre-recorded, the design of the experiment was less interactive than experiments that used real-time manipulation of the auditory stimuli. For example, in experiments that manipulate F0, such as Larson (1998), the pitch heard by participants was $F_{0perceived} = F_{0actual} \pm \Delta F$ where ΔF represents the perturbation of pitch for that trial. Therefore, as participants changed their pitch to either compensate for or follow the perturbation, the pitch they perceive changed by an equal amount and in the same direction, even though the pitch they perceived was not the pitch they produced. The compensatory change in behavior reduced the magnitude of the error signal. In this experiment, because noise was added to the signal, there was no behavior that could reduce the amount of noise below the level of added noise.

Although participants might change their behavior to try to compensate for the noise, the feedback they perceived would not reflect those changes as directly as in the other adaptive learning experiments. Because the perceived feedback would not change as much as the behavior changed, the adaptive learning system would not guide the behavior in the same way. I would predict that the resulting motor behavior would be more variable than in these other experiments.

The acoustic feature that was perturbed was different from the acoustic feature that was measured.

The second major difference between this experiment and previous studies of adaptive learning in speech is that the other studies all manipulated the same parameter that they measured. In this case the parameter that was manipulated was noise-to-harmonic ratio, whereas the parameter that was measured was H1-H2c. This was done because increasing NHR could be done with the available speech synthesis software and gave a good percept of breathiness. Since the experiment was completed, Kreiman and Garellek (2011) showed that manipulating H1-H2 directly does produce a percept of breathiness. It seems likely that perturbing and measuring the same acoustic feature might produce a more stable result.

Suggestions for future research.

Several technical improvements are possible that might improve the reliability of the experiment. Perturbing the auditory signal in real time, either by adding noise or by perturbing spectral slant might reduce the variability in response. Collecting several dependent measures might capture more types of vocal responses to perturbation.

Potential dependent measures could include subglottal air pressure (Zheng, 2010), cepstral peak prominence (Heman-Ackah et al., 2003), and/or electroglottography (Verdolini, Chan, Titze, Hess, & Bierhals, 1998).

Clinical Implications

Because the results of the experiment were not significant, clinical advice would be premature. Nevertheless, it is possible to speculate on the clinical implication if this hypothesis is supported by future research. One implication is that voice rest during episodes of acute laryngitis should prevent the development of muscle tension dysphonia. Although voice rest is often recommended to patients with acute laryngitis, it is generally for the prevention of injury to vulnerable tissue. This model would suggest that voice rest during episodes of acute laryngitis would prevent error signals from disrupting the inverse internal model that generates motor commands. Because the probability of persistent MTD following any individual episode of laryngitis is very small, the sample size required to test this hypothesis would be prohibitively large.

Future Work

Brain imaging studies.

Future experiments will use fMRI to look for an auditory error signal caused by the added noise. This experiment will use the same vocal tract-filtered noise as the currently planned experiment. The participants will be asked to produce nonwords that will randomly have the voice quality perturbed or not. Guenther (2006) observed increased activation in the temporal lobe during trials where auditory perturbation was present. It is expected that if that activation is due to an auditory error signal, then the

same activation should be observed in this experiment on trials with perturbed voice quality compared to trials with no perturbation.

Individuals with cerebellar damage should not develop MTD.

If MTD is a disorder of adaptive learning, and individuals with cerebellar damage have impaired adaptive learning, then MTD should not develop in individuals with cerebellar damage who have impaired adaptive learning. Impaired adaptation is associated with damage to the parts of the cerebellum fed by the posterior inferior cerebellar artery, while ataxia is associated with damage to the parts fed by the superior cerebellar artery (Martin, Keating, Goodkin, Bastian, & Thatch, 1992), so ataxic dysarthria will likely not be a perfect predictor of impaired adaptive learning in the speech system. Nevertheless, since many people with cerebellar damage have both ataxia and impaired adaptation, this theory implies that although MTD is more common in people with laryngeal pathology, it should occur at a frequency closer to that of the general population in patients with ataxic dysarthria. This hypothesis is supported by Duffy (2005, p. 173) who reports that

> Although harshness [a sign of MTD] was among the most deviant characteristics [of ataxic dysarthria] noted by [Darley, Aronson, and Brown], in the author's experience it does not seem to occur frequently or be more than mildly evident in people with isolated ataxic dysarthria. In general, the presence of significant harshness in someone with ataxic dysarthria should raise questions about an accompanying spastic component.

One way to test this hypothesis would be to measure the prevalence of MTD among patients with cerebellar ataxia. However, since epidemiological studies such as Van Houtte et al. (2011) and Cohen, Kim, Roy, Asche, and Courey (2012) have only looked at populations seeking medical assistance for hoarseness, the prevalence of MTD in the general population is unknown. Another way would be to replicate the above experiment of with two groups of participants: cerebellar patients and age-matched, neurologically-normal controls. In addition to the speech task, both groups would complete a motor adaptation task such as throwing to a target with prism glasses (as in Martin et al., 1992) or force adaptation with a manipulandum (as in Maschke et al. 2004). This model would predict that performance on the motor adaptation task would strongly predict performance on the speech task, and that the normal controls would show adaptation with aftereffects on both the motor task and the speech task, while the participants with cerebellar damage would show reduced adaptation on both tasks. A clinical consequence of this hypothesis is that I would not expect those patients with ataxic dysarthria who do present with dysphonia to benefit from behavioral treatment, and that they should instead be taught compensatory strategies such as reducing background noise or using amplification.

Summary and Conclusions

Adaptive learning has been demonstrated in many areas of motor behavior, including speech motor behavior. It is a process that allows animals and humans to change their learned behaviors in response to changes in their bodies or environments. This dissertation presented the hypothesis that adaptive learning may cause the changes

in vocal motor behavior that result in muscle tension dysphonia in cases where the disorder is due to compensation for an organic voice disorder.

The first step in gathering evidence to support this hypothesis is to demonstrate that adaptive learning does occur when voice quality changes. This experiment attempted to find evidence of adaptive learning in the presence of auditory perturbation simulating breathiness. Demonstrating adaptive learning would have required showing both change in voice quality in the presence of auditory perturbation, and the persistence of that change once the auditory perturbation was removed. As none of the differences between experimental conditions were significant at the group level, this experiment failed demonstrate adaptive learning. Conclusive evidence for the presence or absence of adaptive learning in the presence of auditory perturbation may need to await improvements in the tools available to measure voice quality.

Bibliography

Aasland, W.A., Baum, S.R., & McFarland, D.H. (2006). Electropalatographic, acoustic, and perceptual data on adaptation to a palatal perturbation. *J Acoust Soc Am, 119*(4), 2372-2381.

Andruski, J., & Ratliff, M. (2000). Phonation types in production of phonological tone: the case of Green Mong. *J. International Phonetic Association 30*(1-2), 37-61.

Aronson A.E. (1990). *Clinical Voice Disorders, an Interdisciplinary Approach.* New York: Brian C. Decker.

Belafsky, P. C., Postma, G. N., Reulbach, T. R., Holland, B. W., & Koufman, J. A. (2002). Muscle tension dysphonia as a sign of underlying glottal insufficiency. *Otolaryngol Head Neck Surg, 127*(5), 448-451.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.2.19, retrieved 3 March, 2011 from http://www.praat.org/

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *J Acoust Soc Am, 103*(6), 3153-3161.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *J Speech Lang Hear Res, 43*(3), 721-736.

CDex (2007). [Computer program]. Free Software Foundation.

Cohen, S. M., Kim, J., Roy, N., Asche, C., & Courey, M. (2012). Prevalence and causes of dysphonia in a large treatment-seeking population. *Laryngoscope, 122*(2), 343-348.

Donath, T. M., Natke, U., & Kalveram, K. T. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *J Acoust Soc Am, 111*(1 Pt 1), 357-366.

Duffy J.R., (2005). *Motor Speech Disorders.* 2nd Ed. St. Louis: Elsevier Mosby.

Dworkin, J.P., Meleca, R.J., & Abkarian, G.G. (2000). Muscle tension dysphonia. *Curr Opin Otolaryngol Head Neck Surg, 8,* 169–173.

Ferrand, C. T. (2007). *Speech Science* (2nd ed.). Boston: Allyn and Bacon.

E-Prime Version 1.2. [Computer software] (Psychology Software Tools, Pittsburgh, PA)

Garellek, M., Keating, P., Esposito, C.M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *J. Acoust. Soc. Am., 133*(2), 1078-1089.

Ghosh, S. S., Tourville, J. A., & Guenther, F. H. (2008). A neuroimaging study of premotor lateralization and cerebellar involvement in the production of phonemes and syllables. *J Speech Lang Hear Res, 51*(5), 1183-1202.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *J Commun Disord, 39*(5), 350-365.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychol Rev, 105*(4), 611-633.

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., et al. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol, 112*(4), 324-333.

Hillenbrand, J. & Houde, R.A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *J Speech Lang Hear Res, 39*, 311-321.

Hocevar-Boltezar, I., Janko, M., & Zargi, M. (1998). Role of surface EMG in diagnostics and treatment of muscle tension dysphonia. *Acta Otolaryngol, 118*(5), 739-743.

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

Javkin, H. R., Antonanzas-Barroso, N., & Maddieson, I. (1987). Digital inverse filtering for linguistic research. *J Speech Hear Res, 30*(1), 122-129.

Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Curr Biol, 15*(19), 1768-1772.

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: studies of Mandarin tone production. *Journal of Phonetics, 30*, 303–320.

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

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *J Acoust Soc Am, 87*(2), 820-857.

Koufman, J. A., Amin, M. R., & Panetti, M. (2000). Prevalence of reflux in 113 consecutive patients with laryngeal and voice disorders. *Otolaryngol Head Neck Surg, 123*(4), 385-388.

Koufman, J. A., & Blalock, P. D. (1982). Classification and approach to patients with functional voice disorders. *Ann Otol Rhinol Laryngol, 91*(4 Pt 1), 372-377.

Kreiman, J., & Garellek, M. (2011). Perceptual importance of the voice source spectrum from H2 to 2 kHz. *J Acoust Soc Am, 130*(4). 2570.

Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *J Acoust Soc Am, 117*(4 Pt 1), 2201-2211.

Kreiman, J., & Gerratt, B. R. (2010). Perceptual assessment of voice quality: past, present, and future. *Perspectives on Voice Disorders, 20*(2), 62-67.

Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2006). Analysis and Synthesis of Pathological Voice Quality. Retrieved from http://www.surgery.medsch.ucla.edu/glottalaffairs/files/GASoftwareManual2006.pdf

Kreiman, J., Iseli, M., Neubauer, J., Shue, Y-L., Gerratt, B.R., & Alwan, A. (2008). The relationship between open quotient and H1*-H2*. *J Acoust Soc Am, 124*(4), 2495.

Lane, H. & Tranel, B., (1971). The Lombard sign and the role of hearing in speech. *J Speech Hear Res, 14,* 677 - 709.

Larson, C. R. (1998). Cross-modality influences in speech motor control: the use of pitch shifting for the study of F0 control. *J Commun Disord, 31*(6), 489-502.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. *Exp Brain Res, 187(*4), 613-621.

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *J Acoust Soc Am, 107*(1), 559-564.

MacDonald, E. N., Goldberg, R., & Munhall, K. G. (2010). Compensations in response to real-time formant perturbations of different magnitudes. *J Acoust Soc Am, 127*(2), 1059-1068.

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.J.Hardcastle & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 131-149). Boston: Kluwer Academic Publishers.

Martin, T.A., Keating, J.G., Goodkin, H.P., Bastian, A.J. & Thach, W.T. (1996). Throwing while looking through prisms I: Focal olivocerebellar lesions impair adaptation. *Brain, 119,* 1183-1198.

Mattar, A. A., Darainy, M., & Ostry, D. J. (2013). Motor learning and its sensory effects: time course of perceptual change and its presence with gradual introduction of load. *J Neurophysiol, 109*(3), 782-791.

Maryn, Y., Roy, N., De Bodt, M., Van Cauwenberge, P., & Corthals, P. (2009). Acoustic measurement of overall voice quality: a meta-analysis. *J Acoust Soc Am, 126*(5), 2619-2634.

Maschke M., Gomez C., Ebner T., & Konczak J. (2004). Hereditary cerebellar ataxia progressively impairs force adaptation during goal-directed arm movements. *J Neurophysiol, 91*. 230-238.

Metta G, Sandini G, Konczak J. (1999). A developmental approach to visually-guided reaching in artificial systems. *Neural Netw,. 12*(10),1413-1427.

Nasir, S. M., & Ostry, D. J. (2009). Auditory plasticity and speech motor learning. *Proc Natl Acad Sci U S A, 106*(48), 20470-20475.

Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nat Neurosci, 11*(10), 1217-1222.

Natke, U., & Kalveram, K. T. (2001). Effects of frequency-shifted auditory feedback on fundamental frequency of long stressed and unstressed syllables. *J Speech Lang Hear Res, 44*(3), 577-584.

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

Redenbaugh, M. A., & Reich, A. R. (1989). Surface EMG and related measures in normal and vocally hyperfunctional speakers. *J Speech Hear Disord, 54*(1), 68-73.

Roy, N. (2003). Functional dysphonia. *Curr Opin Otolaryngol Head Neck Surg, 11*(3), 144-148.

Roy, N., & Bless, D. M. (2000). Personality traits and psychological factors in voice pathology: a foundation for future research. *J Speech Lang Hear Res, 43*(3), 737-748.

Roy. N., Bless, D.M., & Heisey, D. (2000). Personality and voice disorders: a multitrait-multidisorder analysis. *J Voice, 14,* 521–548.

Shiller D.M., Sato M., Gracco V.L., & Baum S.R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *J Acoust Soc Am, 125*(2),1103-1113.

Shue, Y-L. (2010). VoiceSauce - A program for voice analysis. [Computer software] http://www.ee.ucla.edu/~spapl/voicesauce/

Stepp, C. E., Hillman, R. E., & Heaton, J. T. (2010). The impact of vocal hyperfunction on relative fundamental frequency during voicing offset and onset. *J Speech Lang Hear Res, 53*(5), 1220-1226.

Stepp, C. E., Merchant, G. R., Heaton, J. T., & Hillman, R. E. (2011). Effects of voice therapy on relative fundamental frequency during voicing offset and onset in patients with vocal hyperfunction. *J Speech Lang Hear Res, 54*(5), 1260-1266.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage, 39*(3), 1429-1443.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature, 423*(6942), 866-869.

Van Houtte, E., Claeys, S., D'Haeseleer, E., Wuyts, F., & Van Lierde, K. (2011). An examination of surface EMG for the assessment of muscle tension dysphonia. *J Voice, 27*(2), 177-186.

Van Houtte, E., Van Lierde, K., & Claeys, S. (2011). Pathophysiology and treatment of muscle tension dysphonia: a review of the current knowledge. *J Voice, 25*(2), 202-207.

Van Houtte, E., Van Lierde, K., D'Haeseleer, E., & Claeys, S. (2009). The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia. *Laryngoscope, 120*(2), 306-312.

Verdolini, K. (2005). *Lessac-Madsen Resonant Voice Therapy.* Seminar presented at Park Nicollet Clinic, St. Louis Park, MN.

Verdolini, K., Chan, R., Titze, I.R., Hess, M., & Bierhals, W. (1998). Correspondence of electroglottographic closed quotient to vocal fold impact stress in excised canine larynges. *J. Voice, 12*(4), 415–423.

Verdolini, K., Rosen, C., & Branski, R., (Eds). (2006). *Classification Manual for Voice Disorders-I.* Mahwah, NJ: Lawrence Erlbaum Associates.

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *J Acoust Soc Am, 122*(4), 2306-2319.

Widrow, B., & Walach, E. (1986). *Adaptive Inverse Control: A Signal Processing Approach.* New Jersey: Prentice Hall.

Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *J Acoust Soc Am, 116*(2), 1168-1178.

Zheng, Y. Q., Zhang, B. R., Su, W. Y., Gong, J., Yuan, M. Q., Ding, Y. L., et al. (2010). Laryngeal aerodynamic analysis in assisting with the diagnosis of muscle tension dysphonia. *J Voice, 26*(2), 177-181.

Appendix: Scatterplots of H1-H2c by Trial for Each Participant

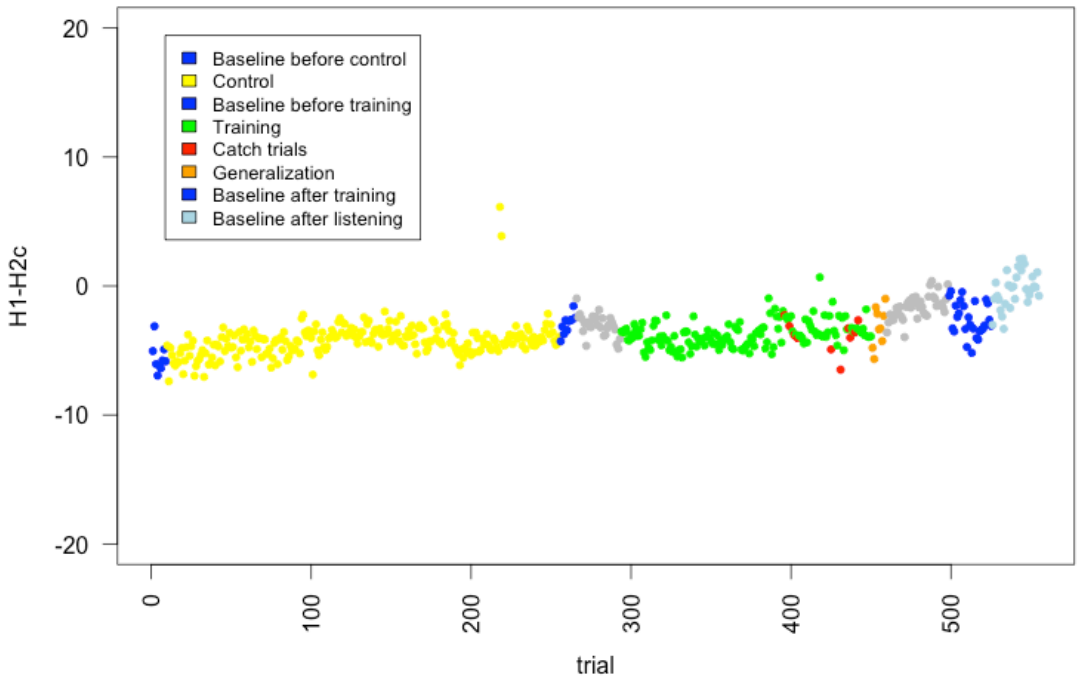


Figure 26. Scatterplot of H1-H2c by trial for participant 1 (follower).

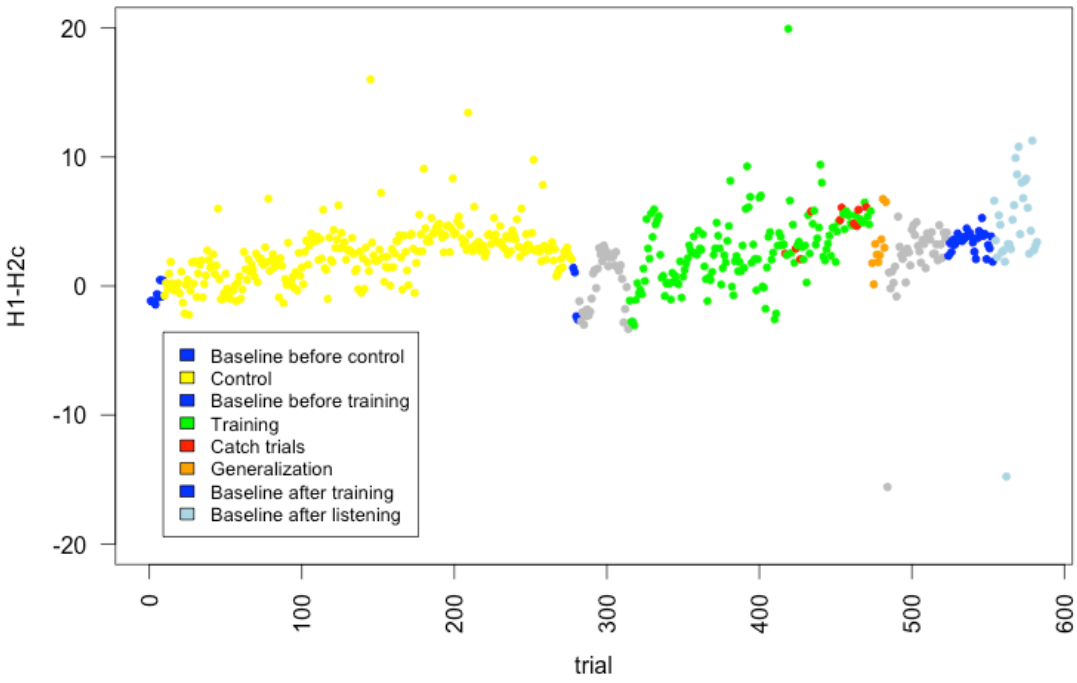


Figure 27. Scatterplot of H1-H2c by trial for participant 3 (non-responder).

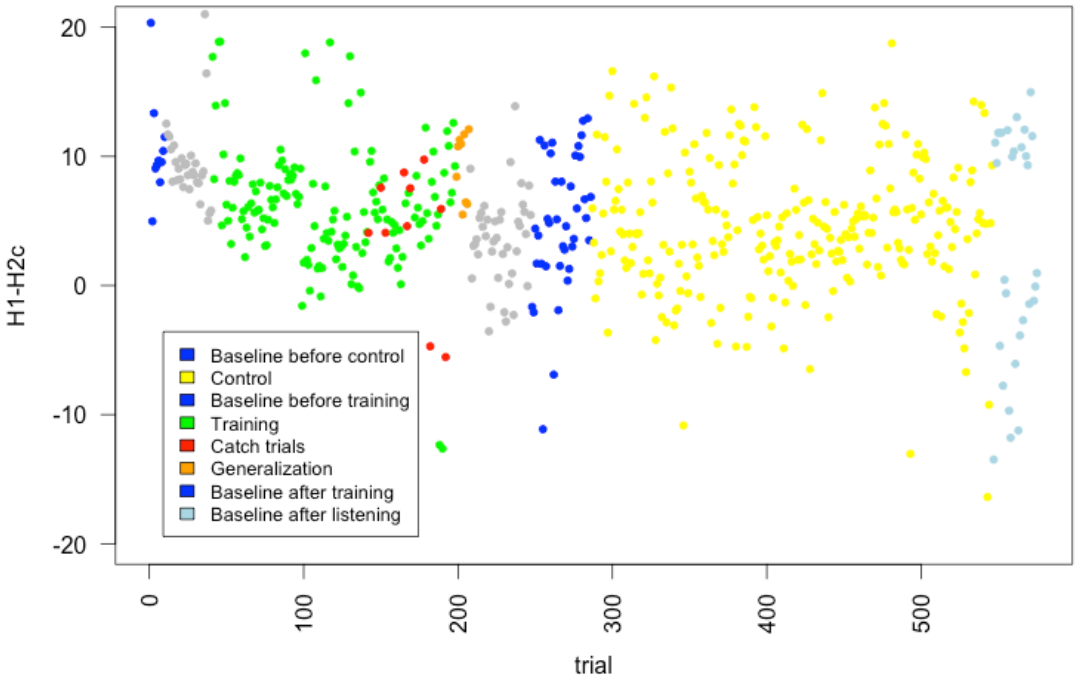


Figure 28. Scatterplot of H1-H2c by trial for participant 4 (follower).

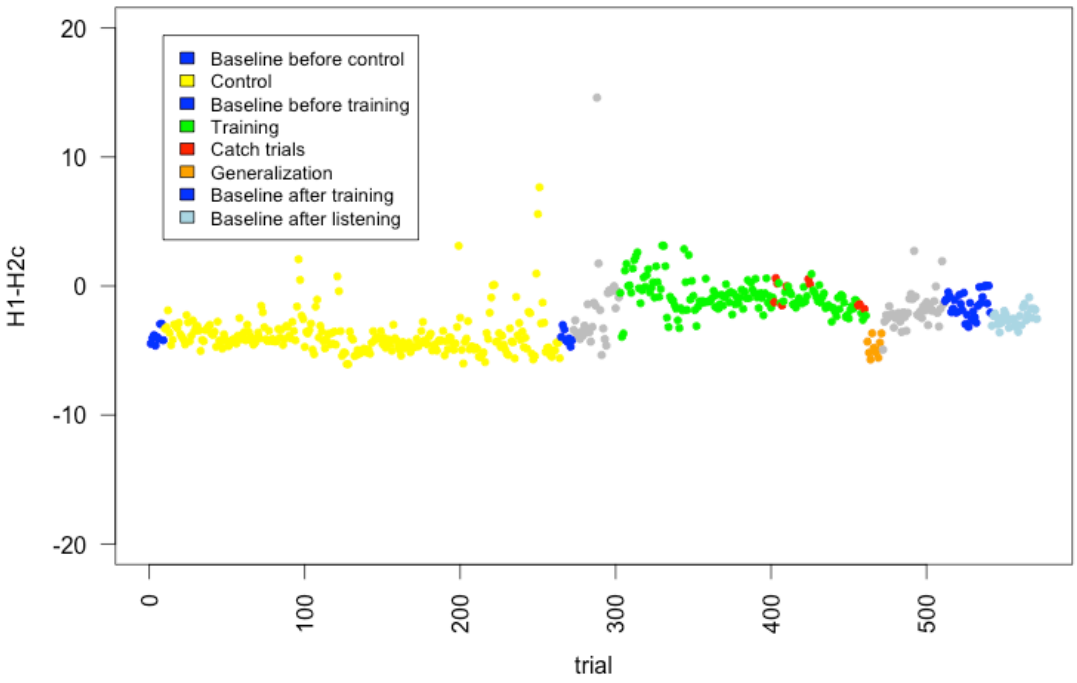


Figure 29. Scatterplot of H1-H2c by trial for participant 5 (follower).

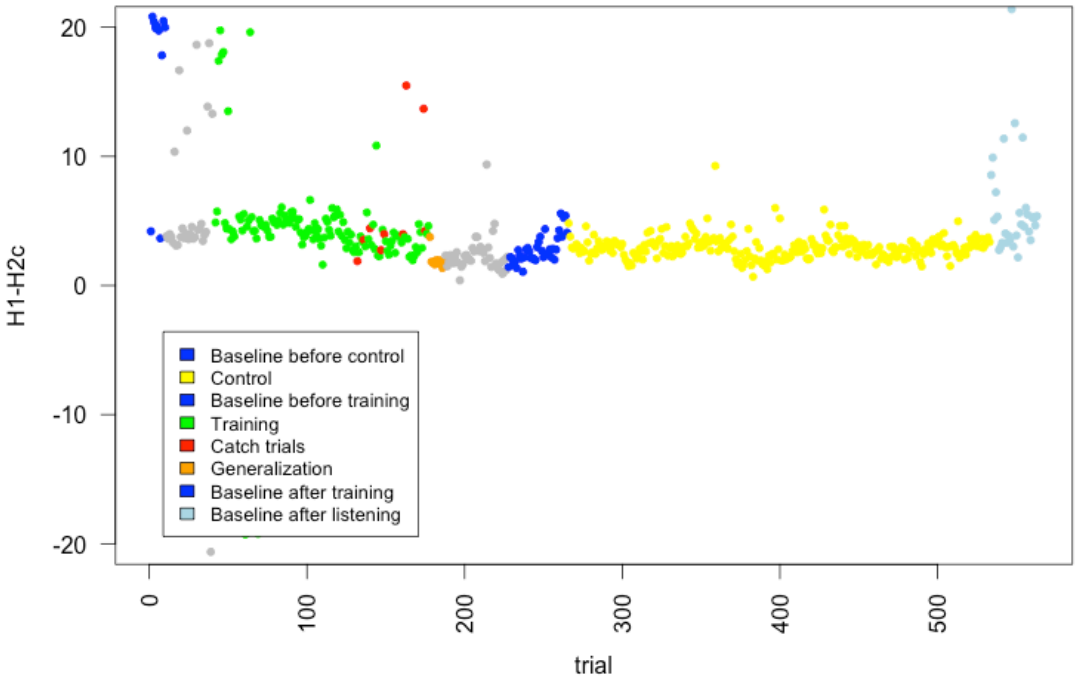


Figure 30. Scatterplot of H1-H2c by trial for participant 6 (non-responder).

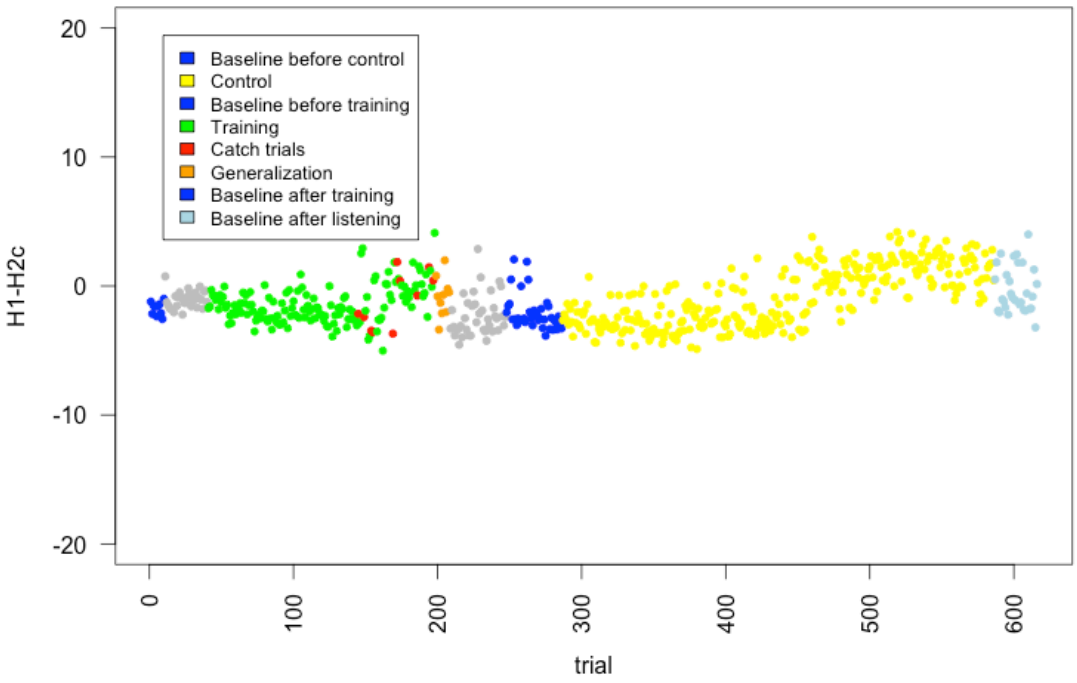


Figure 31. Scatterplot of H1-H2c by trial for participant 8 (compensator).

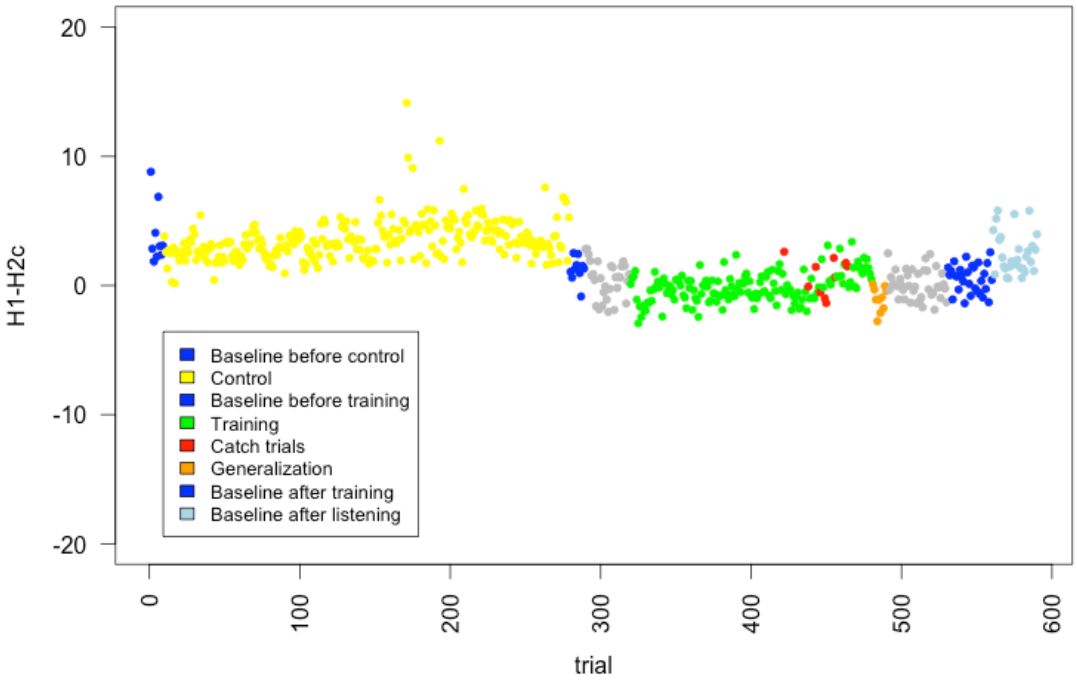


Figure 32. Scatterplot of H1-H2c by trial for participant 9 (compensator).

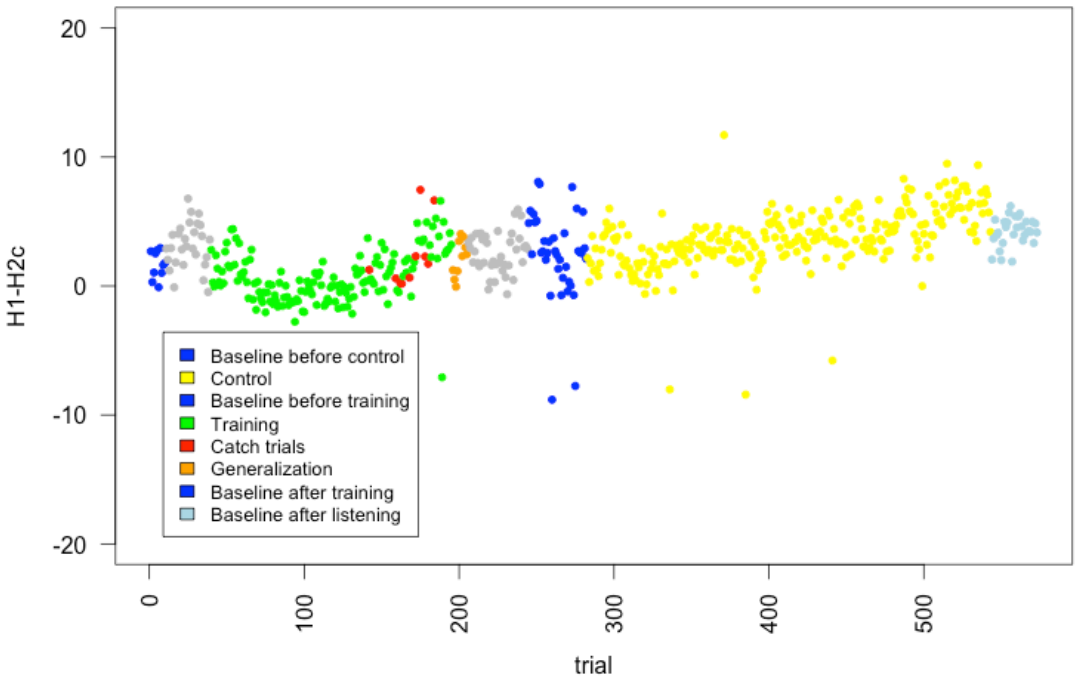


Figure 33. Scatterplot of H1-H2c by trial for participant 10 (compensator).

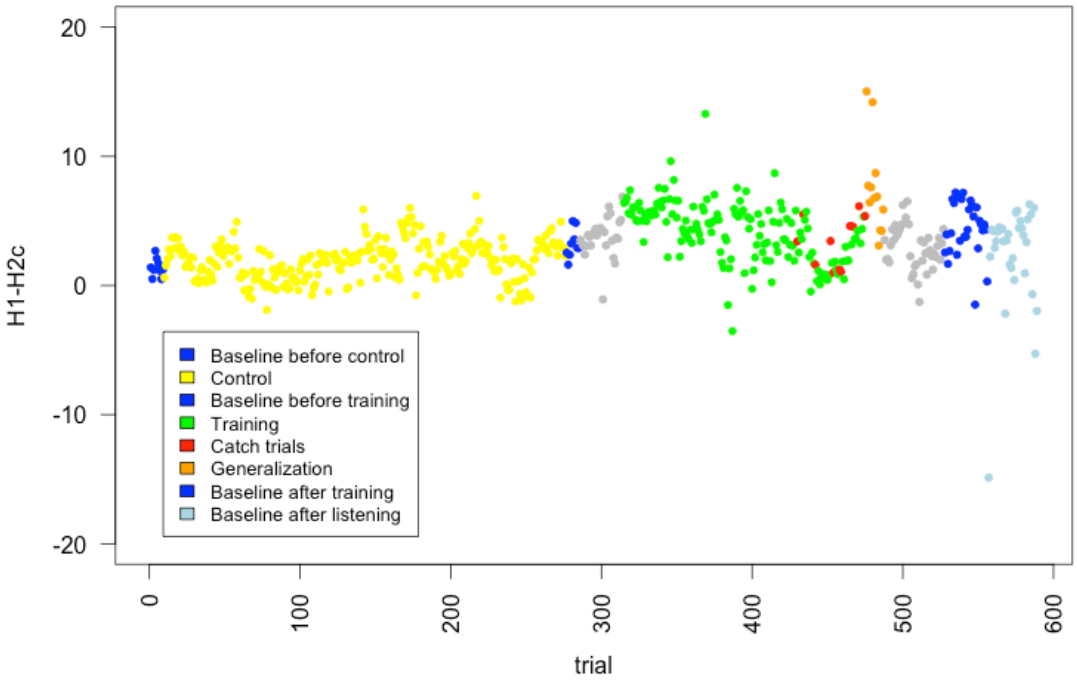


Figure 34. Scatterplot of H1-H2c by trial for participant 11 (follower).

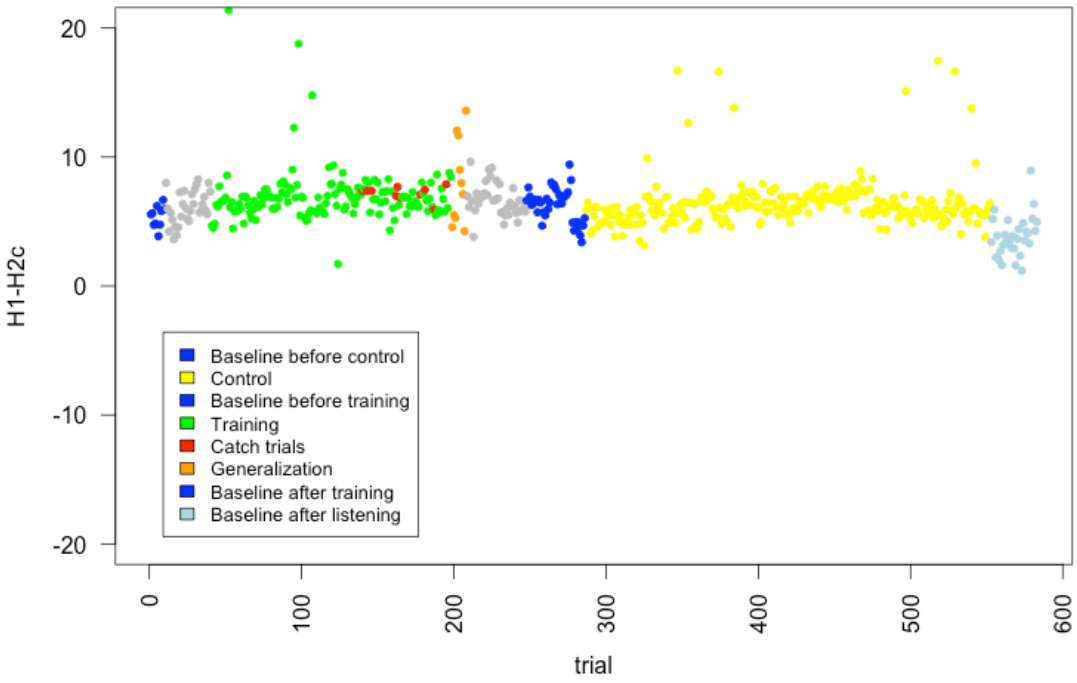


Figure 35. Scatterplot of H1-H2c by trial for participant 12 (non-responder).

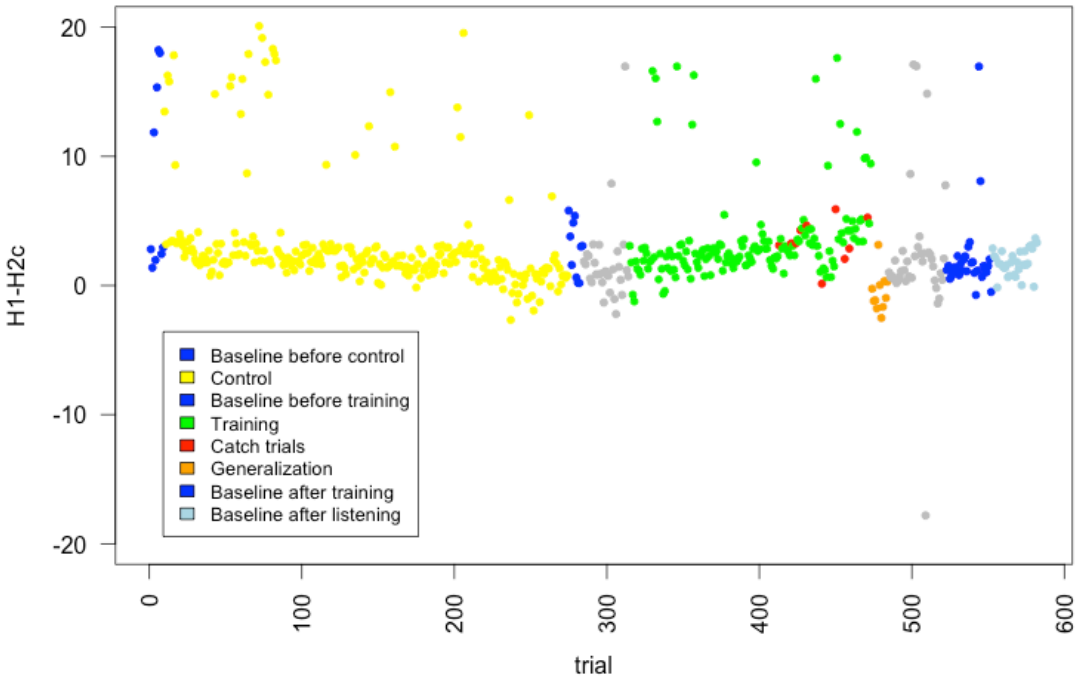


Figure 36. Scatterplot of H1-H2c by trial for participant 13 (non-responder).

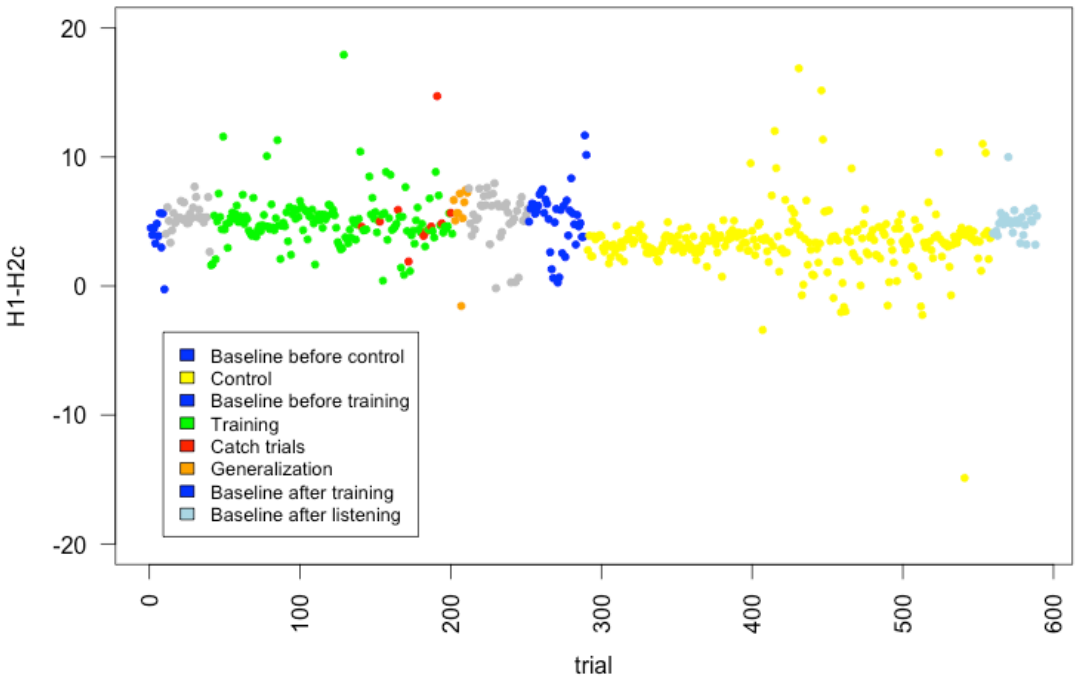


Figure 37. Scatterplot of H1-H2c by trial for participant 14 (follower).

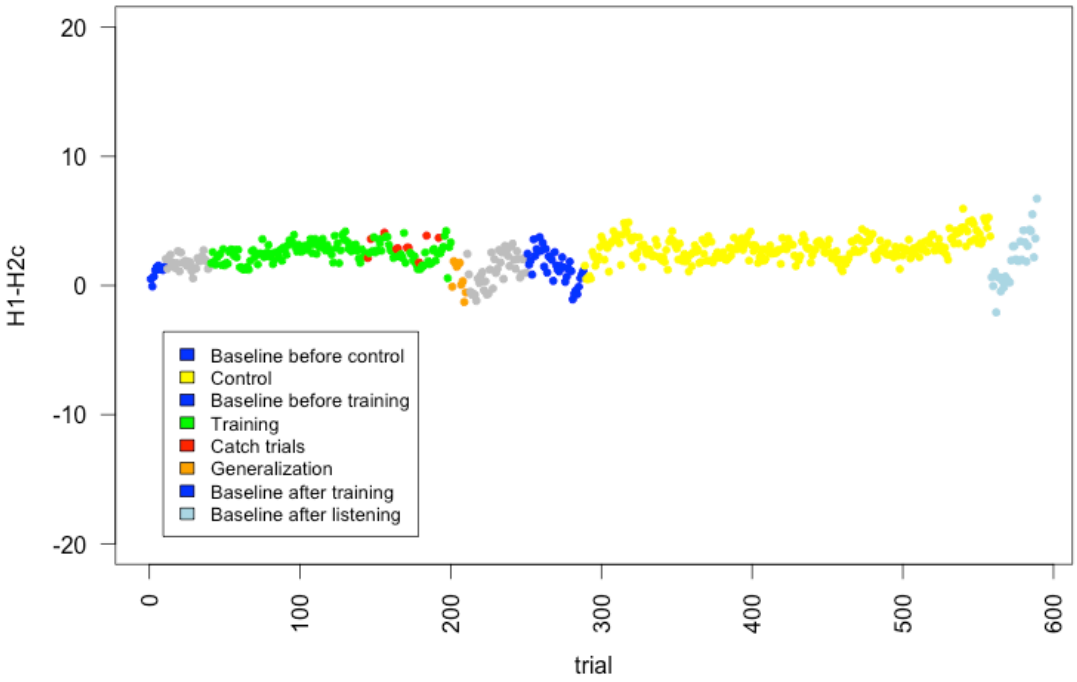


Figure 38. Scatterplot of H1-H2c by trial for participant 16 (compensator).

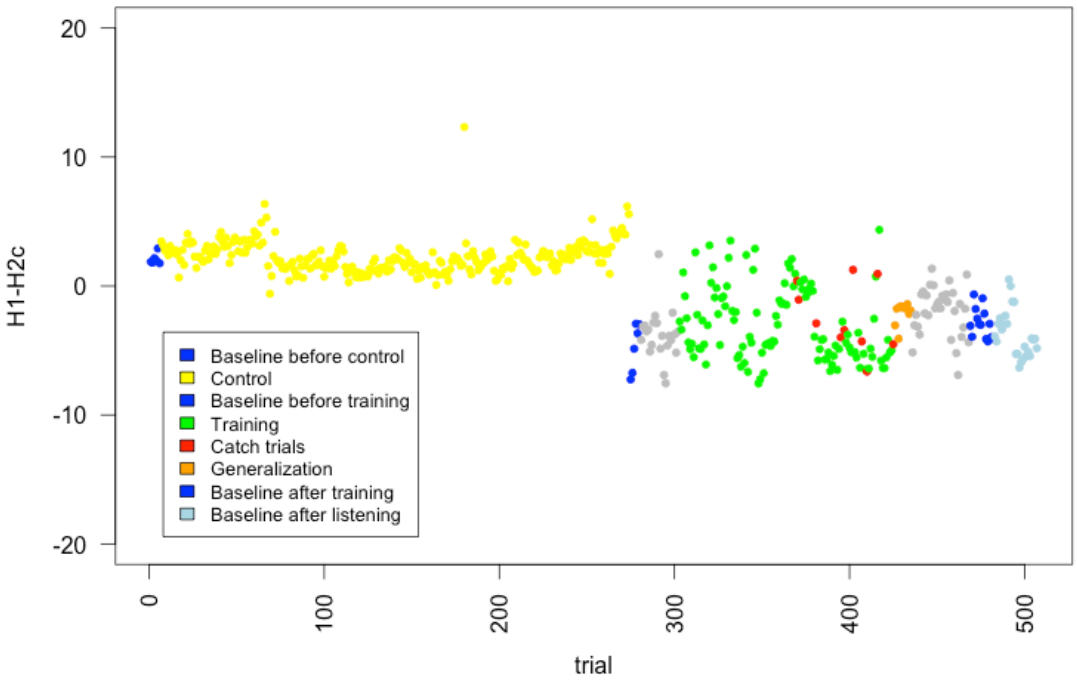


Figure 39. Scatterplot of H1-H2c by trial for participant 17 (compensator).

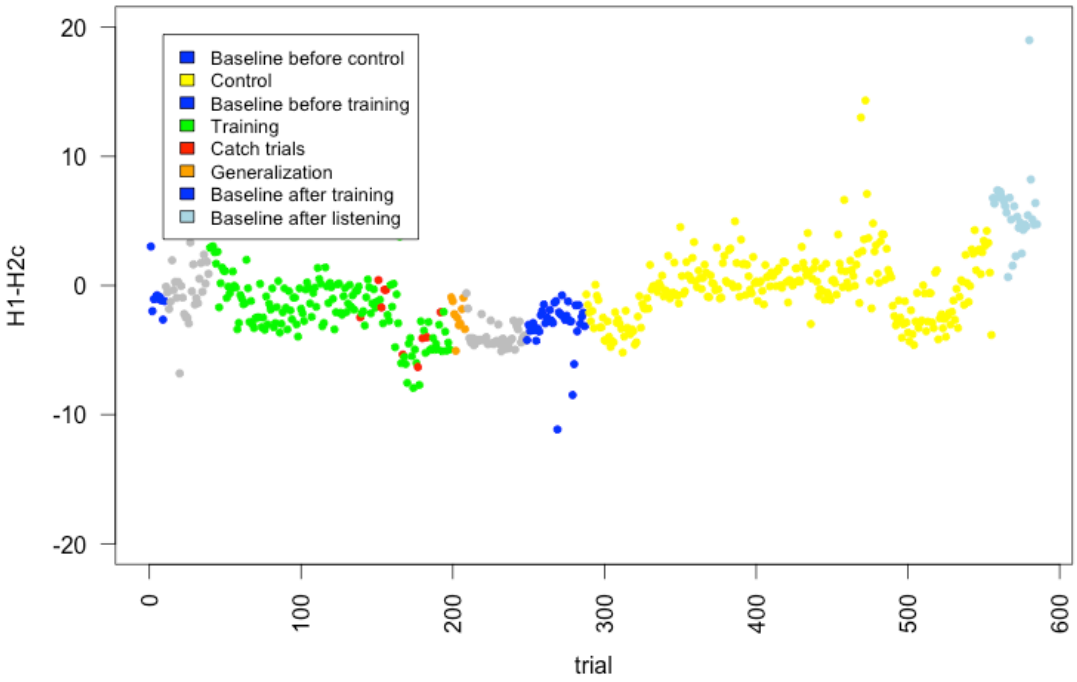


Figure 40. Scatterplot of H1-H2c by trial for participant 18 (compensator).

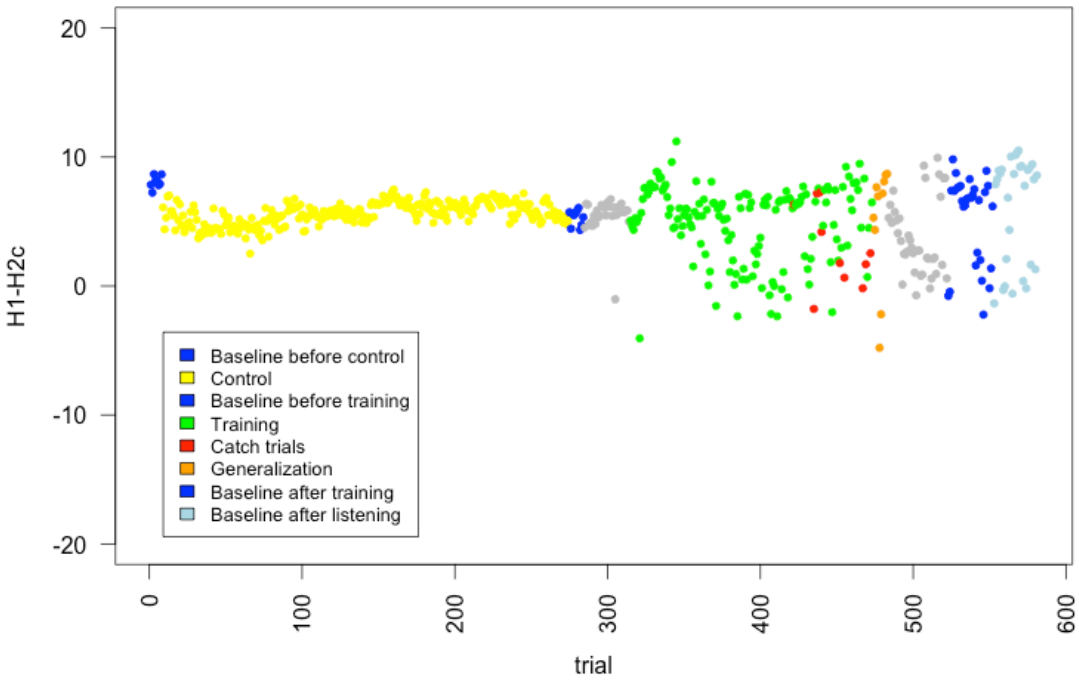


Figure 41. Scatterplot of H1-H2c by trial for participant 19 (compensator).