Review article

Sensory Processing: Advances in Understanding Structure and Function of Pitch-Shifted Auditory Feedback in Voice Control

Charles R Larson [1], *, and Donald A Robin [2]

[1] Dept. of Communication Sciences and Disorders, Northwestern University, Evanston, IL

[2] Research Imaging Institute, University of Texas Health Science Center, San Antonio

* Correspondence: E-mail: clarson@northwestern.edu; Tel: 847-491-2424

Abstract: The pitch-shift paradigm has become a widely used method for studying the role of voice pitch auditory feedback in voice control. This paradigm introduces small, brief pitch shifts in voice auditory feedback to vocalizing subjects. The perturbations trigger a reflexive mechanism that counteracts the change in pitch. The underlying mechanisms of the vocal responses are thought to reflect a negative feedback control system that is similar to constructs developed to explain other forms of motor control. Another use of this technique requires subjects to voluntarily change the pitch of their voice when they hear a pitch shift stimulus. Under these conditions, short latency responses are produced that change voice pitch to match that of the stimulus. The pitch-shift technique has been used with magnetoencephalography (MEG) and electroencephalography (EEG) recordings, and has shown that at vocal onset there is normally a suppression of neural activity related to vocalization. However, if a pitch-shift is also presented at voice onset, there is a cancellation of this suppression, which has been interpreted to mean that one way in which a person distinguishes self-vocalization from vocalization of others is by a comparison of the intended voice and the actual voice. Studies of the pitch shift reflex in the fMRI environment show that the superior temporal gyrus (STG) plays an important role in the process of controlling voice F0 based on auditory feedback. Additional studies using fMRI for effective connectivity modeling show that the left and right STG play critical roles in correcting for an error in voice production. While both the left and right STG are involved in this process, a feedback loop develops between left and right STG during perturbations, in which the left to right connection becomes stronger, and a new negative right

to left connection emerges along with the emergence of other feedback loops within the cortical network tested.

Keywords: Voice; Neural; pitch-shift; auditory feedback; EEG; fMRI

1.   Introduction

The human voice is the bedrock of aural communication, and the foundation for one of the oldest means of expression in humans. Voice is used for many forms of communication as in speech, singing, laughter, crying and anger. Without the ability to vocalize, speech would not be possible. The human voice evolved along with the respiratory system and the presence of air-breathing animals. Thus, the evolutionary process that brought us our voice also led to a vast range of vocal abilities in non-human animals.

Despite the importance of the voice for most human activities, there is much that we do not understand about the neuromuscular mechanisms that control it. However, as with other mysteries, modern technology has allowed us to visualize and measure mechanisms of vocal control that were not possible a few decades ago. Specifically, the development of high resolution brain imaging and new analytic techniques for electrophysiology and functional magnetic resonance imaging (fMRI) have expanded dramatically in the past 10 years, allowing for a far more precise view of the brain and its role in vocal control.

This review describes current knowledge of the mechanisms of vocal control derived from the most recent technological developments in the field of neuroscience. Vocalization is understood to result from the repeated interruption of exhaled air. This process is achieved by air pressure exciting the vocal folds and their release of minute air explosions at high frequencies that are perceived as tones. The muscles involved in this process include those of the respiratory and laryngeal systems. These muscles are in turn controlled by motor neurons located in the brainstem and spinal cord. Beyond the level of the motor neurons, our understanding of how higher areas of the brain control voice pales in comparison to our understanding of how the peripheral muscles and sensory receptors control vocal expression [1].

In order to control vocalization, the nervous system relies on various forms of sensory feedback to monitor the outcome of the control process, to correct for errors in control, and to measure the effect of the voice on the environment (reactions and responses of others). Since the days of Sherrington, one of the most important tools to study the neural processes controlling externally directed movements is the perturbation technique [2]. This technique involves the perturbation of the controlled effector (limbs, lips, legs, etc.) and quantifies the relation between the timing and magnitude of the effector perturbation and the neural response to it. As Sherrington was working on

his seminal studies, Lombard [3] found that when subjects spoke in the presence of a noisy environment, they automatically increased the loudness of their voice. Additionally, approximately 60 years ago, Fairbanks [4] found that if the audio playback of a speaker's voice through earphones was delayed by about 180 ms, the speaker's ability to speak fluently was profoundly disturbed. This technique of delayed auditory feedback made it clear that normal speech production requires precise timing of voice auditory feedback on speech articulation. About 35 years ago, Elman [5] found that perturbing the pitch of the voice while vocalizing caused the speaker to make rapid adjustments to pitch. These advancements led to a series of studies beginning in about 1998 that has greatly increased our understanding of the neural mechanisms of vocal control [6]. These studies relied on advances in digital signal processing that allowed investigators to change the sound of a person's voice and then feed the acoustical signal back via headphones in such a way that the speaker responded as if they perceive that an error was made in the production of the sound (see Figure 1).
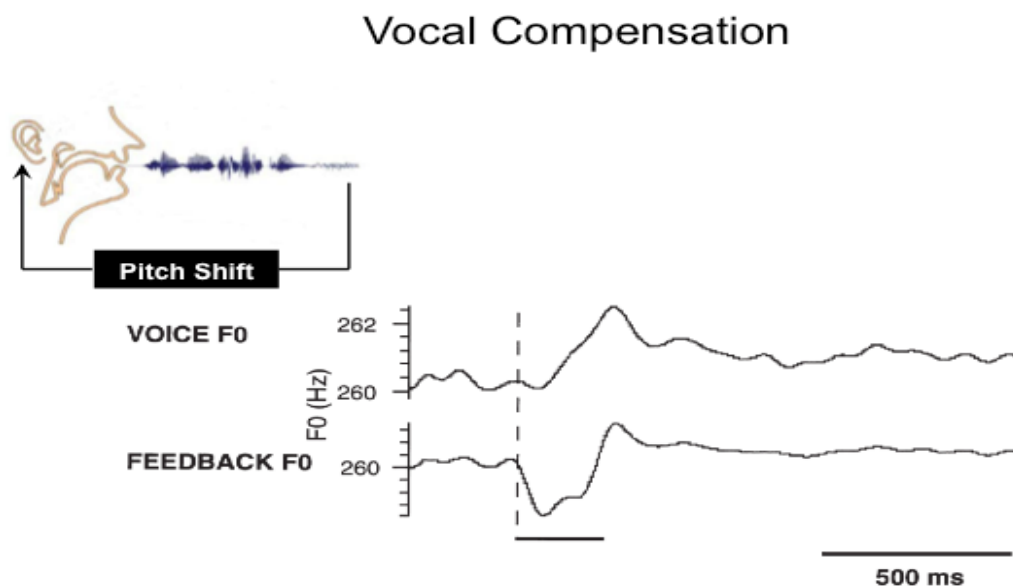
Figure 1. Schematic illustration of auditory feedback perturbation technique. The speaker's voice is altered in pitch and fed back via headphones. This alteration leads to a vocal compensatory response.

A quintessential aspect of most motor controlled events is the monitoring of feedback and correcting for errors in production. Depending on the nature of the behavior in question, the response to the perturbation may exist only at the level of the effector motor neurons (e.g., stretch reflex) or up to and including the highest levels of the nervous system (speech responses to unanticipated changes in the auditory feedback of the spoken word). The perturbation technique applied to voice and speech control has now been used along with common neuroimaging and physiological recording techniques such as positron emission tomography (PET), fMRI, electrocorticography (ECoG), electroencephalography (EEG) and magnetoencephalography (MEG). These approaches have

allowed us to learn the functional role of various brain structures that are involved in the correction of errors in vocalization, and by default, in vocal control itself.

Modern electrical and acoustic modulation techniques have enabled us to learn that when someone is vocalizing (e.g., saying "ah"), an unexpected presentation of altered auditory feedback (voice pitch or loudness) triggers a reflexive response that, more often than not, compensates for the perturbation. That is, the response counters the pitch-shift stimulus direction as if correcting for errors in vocal production. Subjects are unaware that they are making these adjustments, the latencies of the vocal responses are around 100 ms, while the laryngeal EMG latencies are around 50 ms, and the responses are very difficult to suppress, indicating that these responses are reflexive in nature [7,8]. Importantly, except for very small pitch-shift stimulus magnitudes (e.g., 10 cents), the response magnitudes are only a fraction of the stimulus magnitude [9]. For a pitch shift stimulus of 100 cents, most response magnitudes are on the order of 30 cents. However, when pitch-shifted stimuli are presented during running speech, the response magnitudes are larger (50–80 cents) [10,11]. This increase in response magnitudes with running speech indicates that the responses are task dependent and therefore can be modulated according to vocal task requirements.

To account for these neural processes, several investigators have modeled the vocal neural control system. Based on earlier work on the limbs or speech articulators [12–16], recent studies have depicted the vocal control mechanism as a negative feedback control system in which an efferent copy (copy of motor commands) is compared with sensory feedback [17,18]. In both vocal and limb motor control systems, a perturbation to the sensory feedback that the subject perceives results in a change in the subject's response that acts to counter the stimulus. That is, the response is in the opposite direction (opposing response) to the perturbing stimulus. However, in the vocal motor control system, it has frequently been reported that many subjects do not produce an opposing response, and instead they produce a response in the same direction as the stimulus. Such responses have been termed as "following" responses, or feedforward responses [6]. The factors determining the opposing and following responses are not understood, however, they may relate to individual differences in motor control. For example, in a recording of laryngeal EMG activity in the pitch-shift paradigm, one subject showed that one cricothyroid muscle contracted as if to oppose the stimulus direction, and the other cricothyroid muscle contracted as if to follow the stimulus direction [19]. The precise percentage of responses that "follow" the stimulus direction, compared to the opposing responses, is not known because not all studies using the pitch-shift paradigm report the relative percentages. However, in those studies that reported details of following responses, the percentage varied from 2% to 30% [9,11,17,19–25]. The large variations in these percentages are, however, misleading because the methodologies in the studies varied, which in turn affected the number of following responses. For example, Burnett et al. [6] noted that increases in the magnitude of the pitch-shift stimulus led to an increase in the percentage of following responses.

One explanation of following responses is that they may be related to random fluctuations in voice fundamental frequency (F0). That is, during normal vocalization, there are always fluctuations

(increases and decreases) in voice F0 and loudness. If the direction of a fluctuation coincides with that of the stimulus, it could lead to following responses. A second explanation is that if subjects perceive the change in feedback as if it were from an external source such as a piano, they attempt to match the note and thereby follow the direction of the pitch-shift (i.e., to sing along with it) [17].

In a recent experiment bearing on this issue, subjects were instructed to volitionally change their voice F0 either in a direction to compensate for the stimulus (oppose the direction) or to follow the direction of the pitch-shift stimulus [26]. All subjects were able to do this, however, in all subjects it was observed that the following responses had a much shorter latency (150 ms) than opposing responses (400 ms; Figure 2). Moreover, the instructions to oppose the stimulus direction led to a small, early following response that preceded the opposing response and may have delayed the onset of the volitional opposing response.
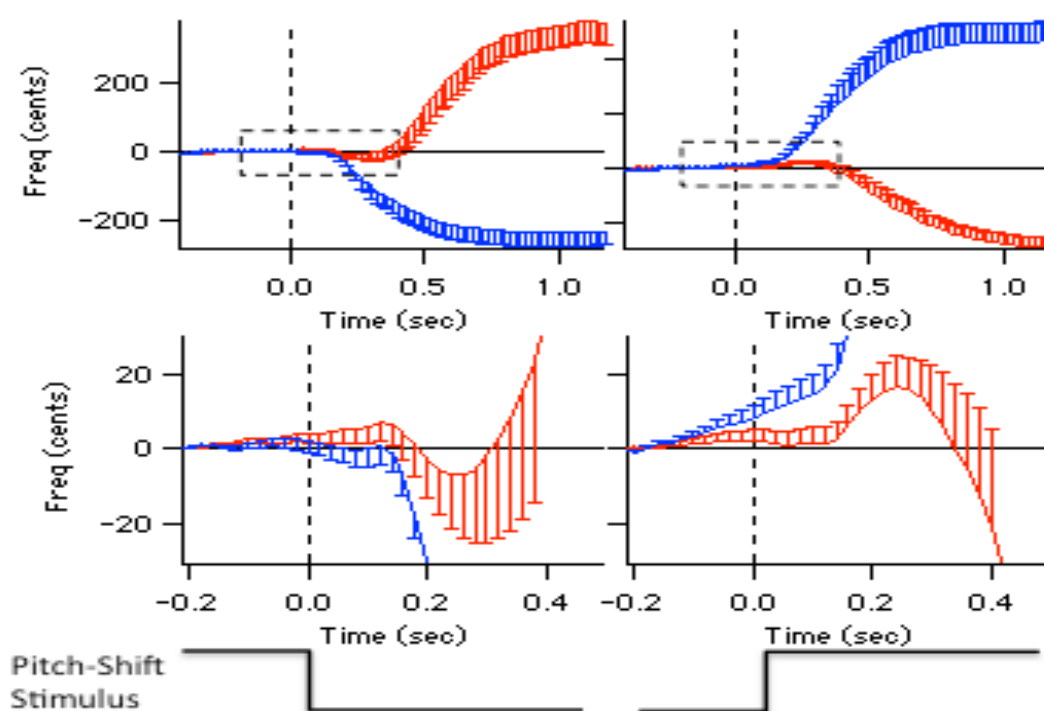

Figure 2. Volitional vocal responses to changes in voice pitch feedback. Graphs on left show responses to a decrease in voice pitch feedback while graphs on right show responses to an increase in pitch feedback. Bracket below graphs show timing and direction of pitch shifted voice feedback. Blue traces show responses that followed the stimulus direction and red traces show responses that opposed the direction of the pitch shift stimulus. Rectangular dashed boxes in upper figure show area of expanded responses in lower graphs. The lower graphs show that the subjects produced a small, short latency response (red traces) that followed the stimulus direction prior to their volitional opposing response. From Patel et al. [26].

Based on this recent study, as well as the earlier studies cited above, we suggest that there may be two different mechanisms of reflexive vocal control. In one type, an opposing response corrects for an error in production when subjects are speaking or sustaining a steady vowel sound. The other type of reflexive response (following) likely reflects a different mechanism that may enhance the ability of the subject to match the pitch of musical notes or another person's voice (the basis of ear training). Moreover, the fact that the volitional opposing responses were preceded by a small following response when subjects attempt to oppose the stimulus direction suggests that the tendency to follow the direction of a pitch-shift stimulus (match a note) may delay the neural mechanisms involved in correcting for an error in voice pitch production.

2. EEG/MEG Studies

Understanding the neural mechanisms underlying the vocal responses to perturbations in voice auditory feedback was sought through the recording of electrical EEG and electromagnetic MEG responses arising from brain activations. The M1 MEG and the N1 EEG auditory ERPs are thought to arise from auditory cortex and to be caused by pre-attentive processing of sound onset or changes in sound to which a person is listening [27]. In the studies discussed here, the N1 potential may also reflect a combination of auditory and vocal motor control activities. Houde and Jordan [18] used MEG to identify brain mechanisms related to the detection of self-voice compared to non-self vocalizations. MEG responses were recorded as subjects began vocalization while hearing their normal voice auditory feedback compared with altered voice auditory feedback. When hearing one's own voice, the M1 MEG response was markedly reduced in magnitude compared to when a subject heard the tape-recorded version of their vocalization, or a non-speech feedback sound while vocalizing. Follow-up studies by Behroozmand et al [28] and Heinks-Maldonado et al [29] examined self-voice identification by changing the frequency of voice feedback by amounts varying from 100 to 400 cents (100 cents = 1 semitone). In these studies, ERPs were triggered by the onset of vocalization or by the sound of the previously recorded vocalization (passive listening). The authors found that the N1 ERP components were suppressed in response to voice onset during active vocalization compared to the ERPs triggered by the sound of the voice as the subject passively listened to the previous vocalization. Moreover, they found that shifting the pitch of auditory feedback during vocalization reduced the amount of N1 suppression. If voice pitch feedback was shifted by 400 cents at voice onset, the N1 suppression was completely eliminated (Figure 3). The suppression of the N1 or M1 ERP of unaltered voice feedback at vocal onset is thought to serve as an indicator of self-vocalization as opposed to the sound of another person's voice [28,30]. Results of these studies support the theory that a precise forward model of the intended vocal output is compared with the actual output, and if there is a disparity arising from this comparison, a correction of the input signal (intended output) is made. As the sound of the voice at onset becomes more dissimilar to the intended voice, the degree of the suppression is reduced. Therefore, the N1 or M1

suppression observed in these studies during vocalization is an indicator that the audio-vocal system is similar to other motor control systems in how it discriminates self-produced actions, such as tickling oneself, from the actions of others [31].
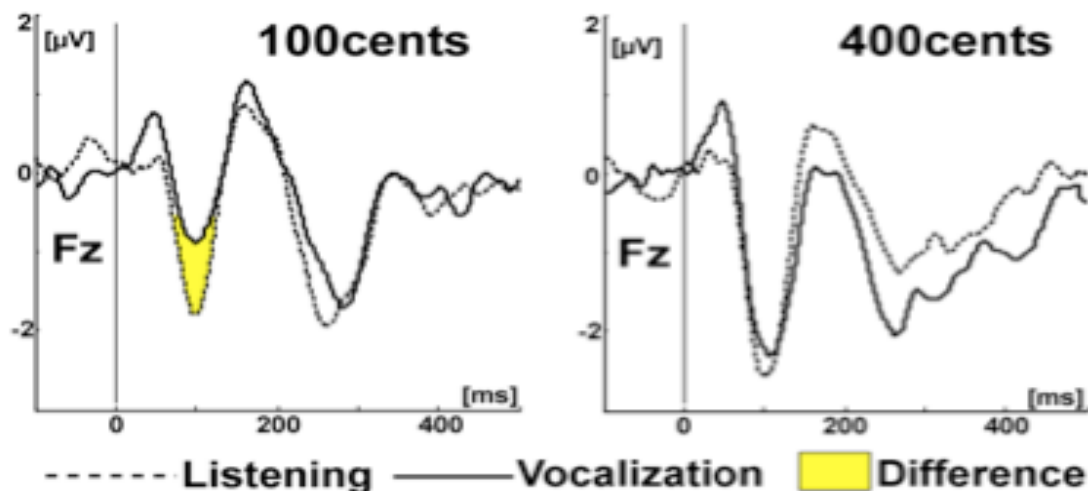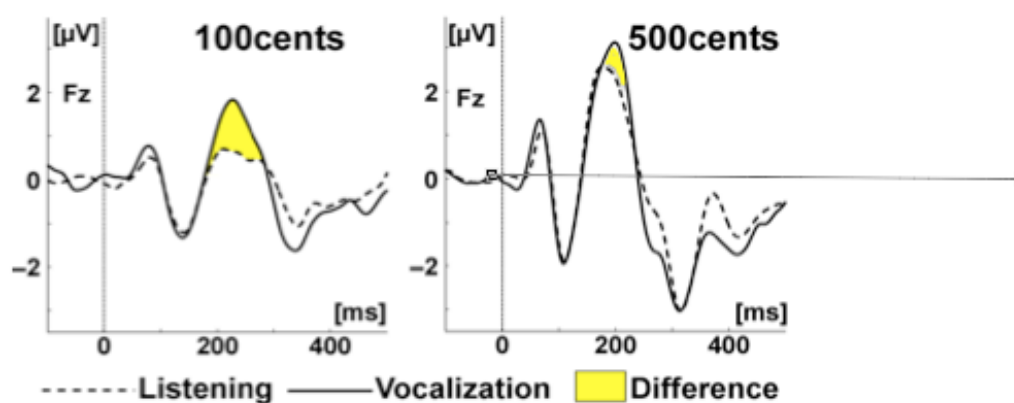

Figure 3. ERPs (averaged across subjects) in response to voice pitch-shifted feedback at voice onset with voice pitch-shifts of 100 and 400 cents magnitude. Solid lines show ERPs in response to pitch-shift stimuli during vocalization, and dashed lines show responses to the auditory feedback of the sound from the previously recorded vocalization. Difference in peak magnitudes shown in yellow. From Behroozmand et al. [28].

In a subsequent study by Behroozmand et al [32] the identification of self-voice was shown to be registered as a reduction in magnitude of the N1 ERP (equivalent to the M1 potential) when pitch-shifted voice auditory feedback occurs within a time window of approximately 200 ms after vocal onset. With a longer delay, the degree of suppression is reduced and is completely absent 1000 ms after vocal onset. Thus, in addition to matching the sound of the voice (e.g. voice pitch), identification of one's own voice requires that the sound of the feedback must occur within 200 ms of the vocalization. Delays in the feedback indicate to the speaker that the voice is not self-produced. This research suggests that differentiation of one's own voice from that of others may be related to the constellation of hallucinogenic symptoms in some patients with schizophrenia [33]. Such patients may misattribute the source of external sounds to objects or other persons.

After the onset of vocalization, such as in speech or singing, and if it has been determined that the voice is self-produced, voice auditory feedback becomes important for vocal control (i.e., the ability of the subject to sustain a steady note with minimal variation in pitch or loudness). In order to investigate neural mechanisms related to voice control after the onset of vocalization, P2 ERPs triggered by pitch-shift stimuli were recorded during vocalization, and then again following the pre-recorded sound of the voice as the subject listened to the previous pitch-shifted vocalization [34].

The P2 auditory potential most likely arises from multiple sources in or near Heschl's gyrus, as well as other cortical areas. In this study, the normalized difference index (comparison of the ERP magnitude in response to auditory feedback during active vocalization compared to the ERP magnitude in response to the sound of the previously recorded vocal signal) was larger for pitch shifts of 100 cents compared to shifts of 500 cents (Figure 4). Greater neural sensitivity to voice auditory feedback during vocalization compared to auditory feedback in the absence of self-vocalization indicates that during vocalization, efference copies of the intended vocalization are compared with auditory feedback and correct for errors if there is a discrepancy between the intended and actual auditory characteristics of the voice [34]. Moreover, greater sensitivity to smaller shifts (100 cents) rather than the larger shifts (200 or 500 cents) suggests a greater sensitivity to self-vocalization than to an abnormal sound, such as someone else's voice [34–36]. Thus, the P2 ERP may reflect mechanisms involved in the comparison of voice auditory feedback with intended output and the subsequent corrective modulation of vocal output.

Figure 4. ERPs (averaged across subjects) recorded from Fz electrode in response to voice pitch-shifted feedback of 100 and 500 cents magnitude presented after voice onset. Solid lines show ERPs during active vocalization. Dashed lines show ERPs in response to playback of the previously shifted voice feedback without active vocalization. From Behroozmand et al. [34].

Complementing the pitch-shifted auditory feedback voice studies in humans, Eliades and Wang [37] recorded activity from neurons in the auditory cortex in Marmoset monkeys during self-initiated vocalizations. Neurons that were usually suppressed during normal voice auditory feedback showed enhanced responses to pitch-shifted vocalizations. These observations from neuronal recordings in primates, using a similar pitch-shift paradigm that has been used with humans, support the idea that the increased amplitude of the P2 ERPs recorded in humans may result from increased

responsiveness of auditory cortex neurons to the sound of their own voice that is shifted in pitch and fed back to the subjects as they are vocalizing.

An important issue regarding the role of neural processing of auditory feedback in vocal control relates to the harmonic complexity of the feedback signal. That is, is the vocal control system sensitive to only the F0 of voice auditory feedback, or does the acoustical complexity of the feedback signal affect responsiveness of the system to alterations in pitch of auditory feedback? To address this issue, Behroozmand et al. [38] compared vocal and ERP responses from subjects who vocalized and heard either the F0 (only) of their voice auditory feedback, the F0 and first harmonic, the F0 and first two harmonics or the F0 and the first three harmonics of their voice. With the increased complexity of the auditory feedback, both the vocal responses and the N1 and P2 ERPs increased in magnitude. The acoustical structure of vocalizations during speech or singing are highly complex, and results from this study suggest that neurons in the more lateral areas of auditory cortex are very sensitive to acoustical signals such as the human voice. That is to say, normal vocalizations are rich in harmonic partials and have a certain F0, while non-vocal sounds may have a diminished harmonic content in comparison to the voice. Furthermore, Behroozmand et al's [38] results indicate that the auditory cortex may respond to inaccuracies in vocal quality and help the speaker (or singer) adjust the voice towards a desired sound structure.

Additional evidence provides more precise details on the cortical areas that are involved in vocal control. Several investigations have shown that there are distinct regions of auditory cortex that are sensitive to the human voice [39–42]. Moreover, direct recording from the cortical surface using ECoG techniques have shown that discrete areas of the STG are sensitive to changes in voice pitch auditory feedback [43,44]. It is therefore highly likely that these areas of auditory cortex contributed to the vocal and ERP results [18,28,32,34,35,45,46] from studies of pitch-shifted feedback.

3.   fMRI Studies

The advent of fMRI has allowed researchers to gain important information about the regions of the brain involved in responses to perturbations in voice as well as some notion of the how those regions support voice control. However, such studies have been challenging because the production of the magnetic field during fMRI data collection results in extremely loud background noise. Since the primary goal of these studies is to gain insight into auditory feedback and the voice, such noise cannot be present during vocalization because the microphone would amplify scanner noise and send it back to the speaker, which would mask the sound of the speaker's voice. In order to overcome this limitation, researchers use an fMRI paradigm called "sparse sampling." Sparse sampling refers to a paradigm whereby the subject vocalizes and then when vocalization has stopped the scanner is turned on. Our paradigm is shown in Figure 5 [47]. Subjects lie still and the scanner is turned off. A subject vocalizes a prolonged "ah" after being cued by a written instruction on a monitor. During vocalization (which can vary in length but is 3 s in the example) there may be no perturbation or a

perturbation at onset or mid vocalization (a similar paradigm is used for voluntary responses to perturbations). After 3 s, subjects are instructed to stop vocalizing and there is a 2 s rest period. The scanner is then turned on for 3 s while the subject continues to rest, followed by an additional 2 s rest period before the next trial. We note that it takes approximately 5 seconds for the hemodynamic response to the pitch-shift stimulus to reach a peak. Hence, we turn the scanner on 5 s after vocal onset.
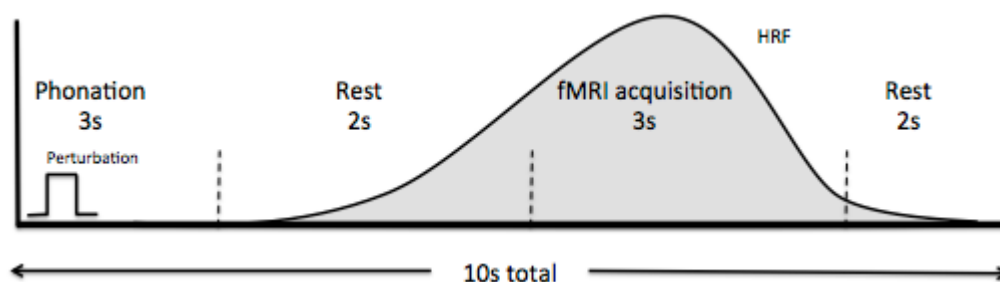
Figure 5. Schematic illustration of the sparse sampling fMRI technique for analysis of vocal responses to pitch-shifted voice feedback. Since the scanner noise would interfere with the microphone recording of the voice, the scanner is turned on after vocalization when the hemodynamic response to the pitch-shift event reaches its maximum.

While fMRI studies show many regions of the brain that are involved in human vocalization [48–50], we first discuss the STG because it has emerged as perhaps the key region involved in the role of auditory feedback in vocal control. In our first study of reflexive responses to auditory feedback perturbations in vocal pitch [47], we studied subjects' vocal and BOLD (blood oxygen level dependent) responses to a 100 cent pitch shift during mid-vocalization. We found that the only brain regions that survived statistical correction were the left and right STG. As noted above, responses to pitch shifted stimuli likely involve an efference copy mechanism. Thus, we argue that the STG is critical, first in the determination of self vs. non-self voice, and second, in generating opposing or following responses associated with auditory feedback perturbations. Hence, changes in ERP activity associated with predicted versus unpredicted changes in voice auditory feedback likely rely on the STG.

We followed this study with one in which the auditory feedback perturbation was 600 rather than 100 cents. The increased pitch shift perturbation was used to (1) explore brain responses to large degrees of error, (2) use a passive listening condition as a comparison to the perturbed events rather than rest only (as had been done in Parkinson et al. above), (3) include correlations between vocal responses to pitch shifted feedback and BOLD responses (performance correlations) in addition to contrast analyses and (4) improve signal to noise ratio (greater error signal) in order to be able to include brain regions in addition to STG that are involved in vocal control [51]. In this study, contrast analysis of vocalization minus rest revealed a complex set of regions that included STG,

primary auditory cortex, precentral gyrus, supplementary motor area (SMA), rolandic operculum, postecentral gyrus and right inferior frontal gyrus (IFG). Contrast of vocalizing versus self-voice playback revealed activity in bilateral precentral gyrus, SMA, IFG, post central gyrus and insula. Performance correlations revealed that vocal responses to pitch shift perturbations were related to increases in the BOLD response in bilateral STG and left precentral gyrus.

Other groups have further delineated the brain regions involved in the neural control of the voice. Toyomura et al. [49] randomly altered auditory feedback in either direction while participants sustained a vowel sound "ah" for 5 seconds. Rather than following or opposing the changes in pitch, participants were instructed to hold the pitch of the feedback voice constant. When compared to those in a non-shift condition, participants in the shift condition displayed right hemisphere BOLD activations in the supramarginal gyrus, premotor cortex (PMC), anterior insula, STG, and intraparietal sulcus [49]. In the left hemisphere, significant BOLD activations were observed only in the PMC, indicating right hemispheric dominance when voluntarily responding to transformed auditory feedback [49]. In contrast with Zarate and Zatorre [7,48] and Zarate et al. [52], in which volitional changes in voice pitch were not made, significant BOLD activations were not observed in the anterior cingulate cortex (ACC), rostral cingulate zone, putamen, and superior temporal sulcus (STS), supporting the notion that these brain regions might be specifically related to voluntary compensation responses.

There is also a body of work on singing that has led to greater understanding of the neural control of the voice using fMRI techniques. For example, Zarate and Zatorre [7,48] reported that differences in BOLD activation between singers and non-singers involved bilateral primary auditory cortices, bilateral primary motor cortex (M1), supplementary motor cortex, ACC, thalamus, insula and cerebellum. In this study, subjects were asked to either ignore a pitch-shift or to compensate for it. In non-musicians, increased BOLD activations were found only in ACC and inferior parietal lobe during both voluntary conditions. By contrast, singers showed numerous regional activations including ACC, inferior parietal lobe (IPL), pre-SMA, STS, insula and putamen. Zarate and Zatorre [7,48] also reported that compensatory vocal responses involve a network of connected regions that include rostral cingulate, ACC, putamen and primary auditory cortex.

To summarize, fMRI data have shown that bilateral STG is key in neural control of vocalization and that it appears to be a hub for comparison between predicted and actual production of the voice. Other regions critical to the neural control of the voice include the IFG, medial superior temporal plane, primary auditory cortices, dorsal PMC, insula, cerebellum and basal ganglia structures.

## 4.   Connectivity Modeling with fMRI Data or Combined fMRI-ERP Signals

Our group recently innovated the use of effective connectivity modeling to understand network coupling properties associated with vocal control and responses to feedback perturbations. Effective connectivity modeling refers to the analysis of fMRI signals that define causal functional relations

between parts of the brain. Structural equation modeling (SEM) is rooted in Bayesian prediction and allows for statements about the strength and sign of neural connections. Results of this analysis allow one to make hypotheses about how one region of the brain modulates another [53]. We used dynamic causal modeling (DCM), also based on Bayesian prediction techniques, on ERP signals to understand smaller sub-network connections between different neural areas [54]. We have used both structural equation modeling (SEM) of fMRI BOLD signals and dynamic causal modeling (DCM) of ERP data. Flagmeier et al. [55] used SEM to study effective connectivity of a cortical network during vocalization with and without a pitch shift perturbation (See Figure 6). We modeled left and right STG, PMC, IFG and M1 because these regions had the greatest activity during vocalization. We determined the best-fit connectivity model for vocalization with no perturbation and when there was an auditory feedback perturbation of 100 cents mid-vocalization. Results showed that left and right STG connectivity was critical for compensating during a perturbation. Specifically, with no shift there was a positive unidirectional connection from left to right STG during vocalization. When a pitch-shift occurred, a feedback loop emerged in which the left to right connection became stronger and a right to left negative connection was present. Other feedback loops that emerged involved (1) right STG to right IFG, in which there was a strong positive connection from STG to IFG and a strong negative connection from IFG to STG, and (2) on the left side only during the pitch shift there was a strong positive IFG to PMC connection and a strong negative PMC to IFG connection that was not present with no-shift vocalization.
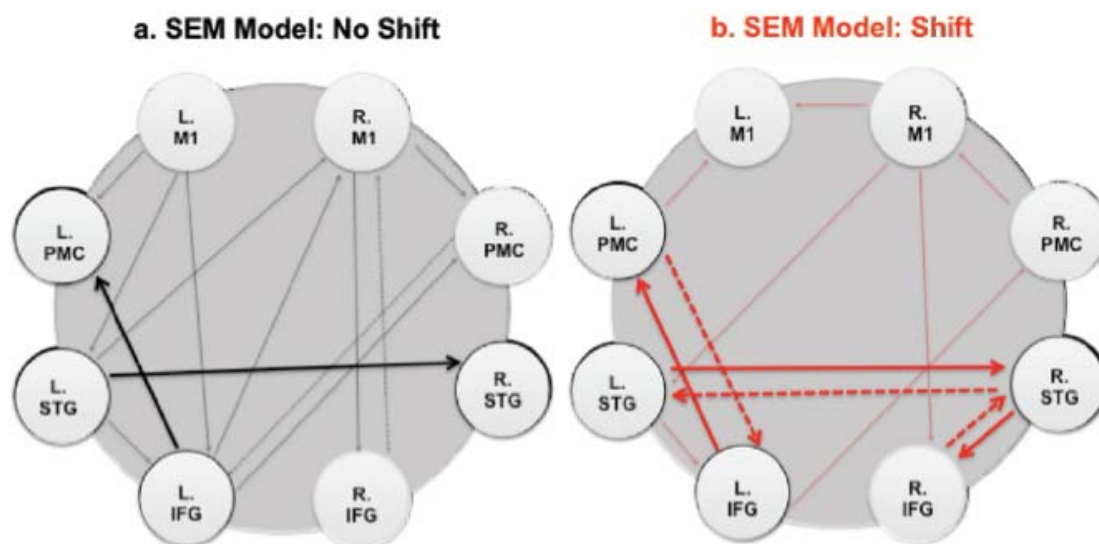
Figure 6. Structural equation modeling results of fMRI responses to non-shifted (a. black) and pitch-shifted (b. red) feedback during vocalization from 8 healthy subjects. Positive path coefficients = solid lines; negative path coefficients = dashed lines. The pitch-shifted feedback condition resulted in negative feedback loops between left and right STG and right STG and right IFG, which were not present in the non-shifted feedback condition. From Flagmeier et al. [55].

We then conducted two experiments using DCM in which we used fMRI to localize the regions of the brain, and uniquely used ERP data to model the electrophysiological signals in the regions identified by fMRI [56,57]. We studied three conditions: passive listening, and pitch-shifts of 100 and 400 cents. We modeled left and right STG, IFG and PMC (see Figure 7). We used Bayesian model selection to first determine that connectivity between left and right STG accounted for responses to a pitch shift of 400 Cents. The next step was to determine the pattern of connectivity between left and right STG that was associated with a perturbation of 400 cents. Thus, we tested right to left, left to right and bilateral connectivity of STG indicated by blue in the Figure. The first finding was that intrinsic connectivity of left and right STG was associated with both the 100 and 400 cent pitch-shift conditions. We also found that both the 100 and 400 cent pitch shift conditions were associated with left to right STG connectivity. In sum, this study also points to STG as the hub of error detection correction mechanisms during perturbed auditory feedback.
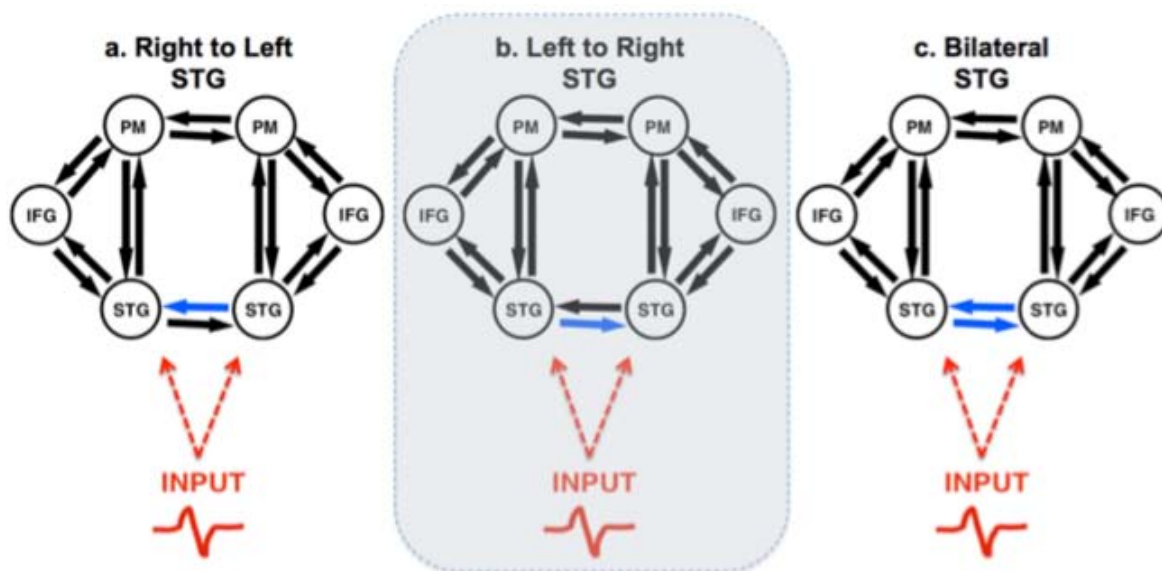
Figure 7. The blue connections are those that were tested in the DCM modeling of STG (a, right to left STG only; b, left to right connectivity of STG and c, bilateral connectivity of STG). Results of DCM shows that the model that best fits the results is left to right connectivity of STG, Figure 7b (the middle model).

In order to study differences in neural responses to individuals that have different musical skill levels, Parkinson et al. [56] studied causal brain connectivity of ERP signals, seeding the same regions used in the ERP study described immediately above: bilateral STG, IFG and PMC, with a 100 cent pitch shift only. We compared musicians with absolute pitch, musicians with relative pitch and non-musicians. Our critical finding was that STG connectivity best separated groups within this network. Specifically, vocal responses to pitch shifted stimuli in subjects with absolute pitch were driven by connectivity of left to right STG. Interestingly, in musicians with relative pitch and

non-musicians, the opposite pattern emerged; connectivity of right to left STG was associated with vocal responses to feedback perturbations.

Finally, we have begun to translate our network modeling work to patients with voice disorders. Our first experiment [58] studied functional connectivity of fMRI data while subjects with Parkinson's disease were at rest in the scanner. Resting state fMRI is task-free and requires subjects to simply lie in the scanner with eyes open for approximately 10 minutes. We developed a model using the brain regions reported by Brown et al. [50] in a meta-analysis of fMRI brain activations associated with vocalization (not speech). The model included SMA, left and right rolandic operculum, left and right PMC, left and right STG, left and right putamen, left and right thalamus and left and right cerebellum. Functional connectivity was determined by performing correlations of BOLD signals between each region of the model. The critical finding was that in healthy subjects there was rich connectivity among all regions of the model, but patients with Parkinson's disease had hypo-connectivity between subcortical and cortical regions as well as between left and right STG and other cortical regions. Parkinson's disease generally affects many areas of the body and is related to deterioration in neural connections between the striatum and cerebral cortex [59]. Our work shows that changes in vocalizations in patients with Parkinson's disease also seems to be related to a disconnection between sub-cortical and cortical regions and between STG and other cortical regions in Parkinson's disease.

In summary, our studies of EEG, fMRI, and connectivity modeling all support our claim that STG is the hub of error detection and correction mechanisms during vocalization. This region of the brain is the only one in which intrinsic connectivity is associated with feedback perturbations and is involved as a central region in modulating activity in other brain regions during pitch shifts. It is also the case that our modeling has uncovered important differences in connectivity patterns in musicians versus non-musicians Finally, our work shows that resting state connectivity is also important in understanding vocal control, and of great interest is the fact that the vocalization network in Parkinson's disease is hypo-connected compared to healthy populations.

5.   Conclusions

We have reviewed studies that have used the pitch-shift paradigm in order to improve our understanding of how voice auditory feedback is used to control vocalization. Studies have shown that one mechanism for this control process involves a reflexive response to variations in voice pitch that counteracts the change in auditory feedback. A second mechanism facilitates the ability to match a change in voice pitch auditory feedback that may possibly be related to neural processes underlying speech or vocal learning. The studies of neural mechanisms underlying these vocal control processes have utilized MEG, EEG, and fMRI techniques. Results of electophysiology studies indicate that at the onset of vocalization, there is a suppression of the M1 or N1 ERP that seems to reflect a process of discrimination of self from non-self vocalization. Later potentials such as the P2 ERP seem to be

involved in registering the magnitude of a pitch shift stimulus during the process of compensating for the detected voice error. Review of studies using fMRI techniques have shown several cortical and subcortical regions involved in the processes described above. These areas include the STG, IPS, PMC, IFG, insula and cerebellum. Finally, structural equation and dynamic causal modeling techniques have shown that the STG plays a critical role in the process of generating vocal responses to changes in voice pitch feedback. Functional connectivity modeling has revealed differences in the intrinsic neural connections involved in voice control between healthy control subjects and people with Parkinson's disease.

Acknowledgments

References

1. Jurgens U (2009) The neural control of vocalization in mammals: A review. *J Voice Found* 23: 1-10.
2. Sherrington CS (1910) Flexion-reflex of the limb, crossed extension-reflex, and reflex stepping and standing. *J Physiol* 40: 28-121, PMC1533734.
3. Lombard E (1911) Le signe de l'évélation de la voix. *Ann Mal Oreille Larynx* 37: 101-119.
4. Fairbanks G (1955) Selective vocal effects on delayed auditory feedback. *J Speech Hear Dis* 20: 333-346.
5. Elman JL (1981) Effects of frequency-shifted feedback on the pitch of vocal productions. *J Acoust Soc Am* 70: 45-50.
6. Burnett TA, Freedland MB, Larson CR, et al. (1998) Voice F0 Responses to Manipulations in Pitch Feedback. *J Acoust Soc Am* 103: 3153-3161.
7. Zarate JM, Zatorre RJ (2008) Experience-dependent neural substrates involved in vocal pitch regulation during singing. *NeuroImage* 40: 1871-1887.
8. Liu H, Behroozmand R, Bove M, et al. (2011) Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback. J Acoust Soc Am 129: 3946-3954, 3135150.
9. Liu H, Larson CR (2007) Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *J Acoust Soc Am* 122: 3671-3677.
10. Xu Y, Larson C, Bauer J, et al. (2004) Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *J Acoust Soc Am* 116: 1168-1178.
11. Chen SH, Liu H, Xu Y, et al. (2007) Voice F0 responses to pitch-shifted voice feedback during English speech. *J Acoust Soc Am* 121: 1157-1163.
12. Sanes JN, Evarts EE (1983) Effects of perturbation on accuracy of arm movements. *J Neurosci* 3: 977-986.

13. Abbs JH, Gracco VL (1984) Control of complex motor gestures: orofacial muscle responses to load perturbations of lip during speech. *J Neurophysiology* 51: 705-723.

14. Kelso JAS, Tuller B, Vatikiotis-Bateson E, et al. (1984) Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures. *J Expe Psy-Hum Percept Perform* 10: 812-832.

15. Cole KJ, Abbs JH (1988) Grip force adjustments evoked by load force perturbations of a grasped object. *J Neurophysiology* 60: 1513-1522.

16. Baum SR, McFarland DH, Diab M (1996) Compensation to articulatory perturbation: Perceptual data. *J Acoust Soc Am* 99: 3791-3794.

17. Hain TC, Burnett TA, Kiran S, et al. (2000) Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Expe Brain Res* 130: 133-141.

18. Houde JF, Nagarajan SS, Sekihara K, et al. (2002) Modulation of the auditory cortex during speech: An MEG study. *J Cog Neurosci* 14: 1125-1138.

19. Liu H, Behroozmand R, Bove M, et al. (2011) Laryngeal electromyographic responses to perturbations in voice pitch auditory feedback. *J Acoust Soc Am* 129: 3946-354, 3135150.

20. Hain TC, Burnett TA, Larson CR, et al. (2001) Effects of delayed auditory feedback (DAF) on the pitch-shift reflex. *J Acoust Soc Am* 109: 2146-2152.

21. Larson CR, Burnett TA, Bauer JJ, et al. (2001) Comparisons of voice F_0 responses to pitch-shift onset and offset conditions. *J Acoust Soc Am* 110: 2845-2848.

22. Larson CR, Burnett TA, Kiran S, et al. (2000) *E*ffects of pitch-shift onset velocity on voice F0 responses. *J Acoust Soc Am* 107: 559-564.

23. Larson CR, Liu H, Behroozmand R, et al. (2008) *Laryngeal muscle responses to voice auditory feedback perturbations*, in *International Conference on Voice Physiology and Biomechanics* 2008: Tampere, Finland.

24. Larson CR, Sun J, Hain TC (2007) Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *J Acoust Soc Am* 121: 2862-2872.

25. Liu H, Xu Y, Larson CR, et al. (2009) Attenuation of vocal responses to pitch perturbations during Mandarin speech. *J Acoust Soc Am* 125: 2299-306, 2677266.

26. Patel S, Nishimura C, Lodhavia A, et al. (2014) Voice control during voluntary responses to pitch-shifted auditory feedback. *J Acoust Soc Am* 135: 3036-3044.

27. Burkard RF, Eggermont JJ, Don M (2007) *Auditory Evoked Potentials*. Baltimore: Williams and Wilkins. 731.

28. Behroozmand R, Liu H, Larson CR, et al. (2011) Time-dependent neural processing of auditory feedback during voice pitch error detection. *J Cogn Neurosci* 23: 1205-1217, 3268676.

29. Heinks-Maldonado TH, Nagarajan SS, Houde JF, et al. (2006) Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 17: 1375-1379.

30. Houde JF, Jordan MI (2002) Sensorimotor adaptation of speech I: Compensation and adaptation. *J Speech Lan Hearing Res* 45: 295-310.

31. Wolpert DM, Ghahramani Z, Jordan MI (2014) An internal model for sensorimotor integration. *Science* 269: 1880-1882.

32. Behroozmand R, Liu H, Larson CR (2011) Time-dependent neural processing of auditory feedback during voice pitch error detection. *J Cogn Neurosci* 23: 1205-1217, 3268676.

33. Heinks-Maldonado TH, Mathalon DH, Houde JF, et al. (2007) Relationship of imprecise corollary discharge in schizophrenia to auditory hallucinations. *Arch General Psychiatry* 64: 286-296.

34. Behroozmand R, Karvelis L, Liu H, et al. (2009) Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin Neurophysiol* 120: 1303-1312, 2710429.

35. Hawco CS, Jones JA, Ferretti TR, et al. (2009) ERP correlates of online monitoring of auditory feedback during vocalization. *Psychophysiology*.

36. Scheerer NE, Behich J, Liu H, et al. (2013) ERP correlates of the magnitude of pitch errors detected in the human voice. *Neuroscience* 240: 176-185.

37. Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453: 1102-1106.

38. Behroozmand R, Korzyukov O, Larson CR (2011) Effects of voice harmonic complexity on ERP responses to pitch-shifted auditory feedback. *Clin Neurophysiol* 122: 2408-2417, 3189443.

39. Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport* 14: 2105-2109.

40. Belin P, Zatorre RJ, Ahad P (2002) Human temporal-lobe response to vocal sounds. Brain Research. *Cog Brain Res* 13: 17-26.

41. Fecteau S, Armony JL, Joanette Y, et al. (2004) Is voice processing species-specific in human auditory cortex? An fMRI study. *NeuroImage* 23: p. 840-848.

42. Fecteau S, Armony JL, Joanette Y, et al. (2005) Sensitivity to voice in human prefrontal cortex. *J Neurophysiology* 94: 2251-2254.

43. Greenlee J, Jackson AW, Chen F, et al. (2011) Human auditory cortical activation during self-vocalization. *PLOS One* 6: 1-15, PMC3135150.

44. Greenlee JD, Behroozmand R, Larson CR, et al. (2013) Sensory-motor interactions for vocal pitch monitoring in non-primary human auditory cortex. *PLoS One* 8: e60783, 3620048.

45. Jones SJ (2003) Sensitivity of human auditory evoked potentials to the harmonicity of complex tones: evidence for dissociated cortical processes of spectral and periodicity analysis. *Expe Brain Res* 150: 506-514.

46. Liu H, Behroozmand R, Larson CR (2010) Enhanced neural responses to self-triggered voice pitch feedback perturbations. *NeuroReport* 21: 527-531.

47. Parkinson AL, Flagmeier SG, Manes JL, et al. (2012) Understanding the neural mechanisms involved in sensory control of voice production. *Neuroimage* 61: p. 314-322, 3342468.

48. Zarate JM, Zatorre RJ (2005) Neural substrates governing audiovocal integration for vocal pitch regulation in singing. *An New York Aca Sci* 1060: 404-408.

49. Toyomura A, Koyama S, Miyamaoto T, et al. (2007) Neural correlates of auditory feedback control in human. *Neuroscience* 146: 499-503.

50. Brown S, Ngan E, Liotti M (2008) A larynx area in the human motor cortex. *Cerebral Cortex* 18: 837-845.

51. Behroozmand R, Shebek R, Hansen DR, et al. (2015) Sensory-motor networks involved in speech production and motor control: an fMRI study. *Neuroimage* 109: 418-428, 4339397.

52. Zarate JM, Wood S, Zatorre RJ (2010) Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia* 48: p. 607-618.

53. Friston KJ (1994) Functional and Effective Connectivity in Neuroimaging: A Synthesis. *Hum Brain Map* 2: p. 56-78.

54. Kiebel SJ, David O, Friston KJ (2006) Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. *Neuroimage* 30: 1273-1284.

55. Flagmeier SG, Ray KL, Parkinson AL, et al. (2014) The neural changes in connectivity of the voice network during voice pitch perturbation. *Brain Lang* 132C: 7-13.

56. Parkinson AL, Behroozmand R, Ibrahim N, et al. (2014) Effective connectivity associated with auditory error detection in musicians with absolute pitch. *Front Neurosci* 8: 1-9, PMC3942878.

57. Parkinson AL, Korzyukov O, Larson CR, et al. (2013) Modulation of effective connectivity during vocalization with perturbed auditory feedback. *Neuropsychologia* 51: 1471-1480, 3704150.

58. New AB, Robin DA, Parkinson AL, et al. (2015) The intrinsic resting state voice network in Parkinson's disease. *Hum Brain Mapp* 36(5): 1951-1962.

59. Duffy JR (1995) *Motor Speech Disorders*. St. Louis: Mosby. 467.