ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 4pSCa: Auditory Feedback in Speech Production II

4pSCa1. Cortical mechanisms of integrating auditory feedback with vocal pitch control
Jean Mary Zarate*

*Corresponding author's address: Psychology, New York University, New York, NY 10003, jean.m.zarate@nyu.edu

Precise vocal pitch regulation is crucial for both speech and song. The pitch of a speaker's voice can indicate the intent of a sentence, set the emotional context of a conversation, or distinguish meanings in tonal languages. In singing, accurate vocal pitch is the single most important element needed to properly produce notes and melodies. Vocal pitch regulation requires the integration of auditory feedback processing with the vocal motor system, also known as audio-vocal integration; however, the neural substrates governing this integration have been elusive. Recent functional magnetic resonance imaging (fMRI) studies of singing with pitch-shifted feedback are presented here to outline the neural mechanisms of audio-vocal integration for voluntary vocal pitch regulation, and to discuss the effects of long-term vocal training on vocal performance and neural activity during vocal pitch regulation.

Published by the Acoustical Society of America through the American Institute of Physics
INTRODUCTION

Vocal pitch plays an important role in both speech and song. The pitch of a speaker’s voice can indicate the intent of a sentence (e.g., declarative statements versus questions), set the emotional context for a conversation, or distinguish words or meanings in tonal languages. In singing, vocal pitch is the single most important element needed to produce notes and melodies. In the absence of precise vocal pitch regulation, a person can still communicate intent or thoughts via speech, but a song can be rendered unrecognizable. To regulate vocal pitch, the brain must integrate information about the state of the vocal motor system with information from auditory feedback; animal research has outlined the possible neural substrates for this audio-vocal integration. Since vocal pitch regulation is more crucial for the goals of singing than speech, two functional magnetic resonance imaging (fMRI) experiments that employed singing tasks are presented below to outline the putative neural substrates of audio-vocal integration in humans.

Neural Control of Vocalization

In mammals, Jürgens outlined two neural pathways of vocal control (Figure 1) that project to a common endpoint: the reticular formation, which coordinates all phonatory muscle groups to generate complete vocal patterns. The first pathway consists of the anterior cingulate cortex (ACC) and periaqueductal gray (PAG), and is attributed with the voluntary initiation or readiness of vocalizations. The second pathway is spearheaded by the primary motor cortex and may be involved in generating learned vocalizations, such as speech and song. Vocal motor commands from the primary motor cortex are modulated by two loops—one consisting of the pontine gray and cerebellum, and the other containing the putamen and globus pallidus—that send modified motor commands back to the primary motor cortex (via the ventrolateral thalamus) for execution. Neuroimaging studies from the last two decades have verified that various speech and song tasks (e.g., syllable repetition, repeatedly singing a note, spontaneous and synchronized speaking and singing, etc.) recruit regions from both of Jürgens’ vocal control pathways: primary motor cortex, ACC, thalamus, putamen, globus pallidus, and cerebellum.

Potential Neural Substrates of Audio-vocal Integration

Within the networks discussed above, there are a few putative substrates for audio-vocal integration. In animals, the PAG not only produces vocalizations when electrically or pharmacologically stimulated, but also elicits and modulates responses to external acoustic stimuli, perhaps due to its connections with the inferior colliculus and lateral lemniscus in the ascending auditory pathway. These observations suggest that the PAG and lower-level auditory regions may be involved in audio-vocal integration. Further evidence for the PAG’s potential involvement in this process stems from investigations of the Lombard reflex, in which a decrease in the auditory feedback amplitude during vocalization (e.g., due to masking noise) results in a fast, automatic increase in vocal output amplitude. In decerebrate cats, the Lombard reflex is preserved during PAG-induced vocalizations coupled with auditory masking, suggesting that the PAG may govern audio-vocal integration during involuntary audio-vocal reflexes, without additional control from cortical regions. Finally in humans, the PAG was more active during voiced utterances than during whispered speech, signifying that producing voiced utterances require the PAG.

Similar to the PAG, animal vocalizations are elicited when the ACC is electrically stimulated. In the echolocating bat, the ACC has been reported to contain a tonotopic representation—particular vocal frequencies are emitted after specific ACC sites are stimulated—that may be due to tonotopically organized input from auditory cortical regions that project to the ACC. The ACC also sends projections via the extreme capsule to auditory regions along the superior temporal gyrus (STG) and dorsal bank of the superior temporal sulcus (STS) and modifies their activity. In many animals, stimulation of the ACC attenuates the response of auditory neurons within STG just prior to vocalization. The reciprocal connections between ACC and auditory cortex may be sufficient for audio-vocal integration, but the insula can also contribute to this process via reciprocal connections with both regions. In fact, the anterior insula may be involved specifically during audio-vocal integration because its activity is enhanced during overt vocalization when compared with covert or internal vocalization; this suggests that voiced utterances may specifically engage the anterior insula. Furthermore, the insula’s cytoarchitecture and projections identify it as an association area that integrates auditory input with other modalities, including visual and vocal motor systems.
FIGURE 1. Network models of vocal control and audio-vocal integration based on animal work. Areas in blue represent the vocal motor network involved in producing learned vocalizations, while areas in green outline the vocal motor network associated with the voluntary initiation of vocalization. Both networks project to the reticular formation, which coordinates all phonatory motoneurons to produce vocalizations. The auditory feedback from vocalization (dashed arrows) is processed by the ascending auditory pathway and auditory cortical regions [black slanted lines indicate that only selected regions of this pathway are represented (light orange)]. Red double-headed arrows represent potential substrates for audio-vocal integration: (a) the ACC, STG/STS, and the insula (a higher-order association area shown in purple); and (b) the PAG and either the inferior colliculus or lateral lemniscus. Modified from Jürgens, 2002; 2009.

Briefly summarized, the PAG and the network between the temporal neocortex within STG/STS, the anterior insula, and ACC may form subcortical and cortical substrates, respectively, for audio-vocal integration, by virtue of their functional roles and shared connections with each other (Figure 1).

INVESTIGATING BEHAVIORAL AND NEURAL CORRELATES OF AUDIO-VOCAL INTEGRATION

Numerous behavioral studies have investigated audio-vocal integration by manipulating auditory feedback during vocalization. One of the earliest and most well-known reports is of the aforementioned Lombard reflex, in which a decrease in feedback amplitude results in increased vocal output amplitude. During speech, when the first formant frequency is shifted so that a produced vowel (e.g., /e/) sounds like a different one (e.g., /æ/), the vocal motor system quickly compensates for the formant shift. When the fundamental frequency (i.e., pitch) is shifted in auditory feedback, investigators have observed responses in which vocal pitch is adjusted quickly in the opposite direction of the feedback shift. Two response components have been isolated: 1) an early pitch-shift response that occurs approximately 100-150 ms after the pitch shift, with a magnitude of only 25-50 cents irrespective of the pitch-shift magnitude; and 2) a late pitch-shift response with a latency of 250-600 ms. The early pitch-shift response is purported to be a more automatic reaction that stabilizes vocal output by correcting small, unexpected fluctuations in vocal pitch. In contrast, the late pitch-shift response may be under more voluntary control and thus may be used to control vocal pitch during speaking and singing. For example, although singers display early pitch-shift responses, they still can maintain their intended goal for vocalization (singing steadily or gliding pitches), perhaps due to enhanced top-down voluntary vocal control that resulted from years of training. This suggests that each component may be governed by different neural mechanisms—PAG for the early, more involuntary pitch-shift response, and a cortical network of temporal neocortex, anterior insula, and ACC for the later, more voluntary component (Figure 1)—as proposed by Burnett and colleagues (1998). The experiments presented below employed variations of the pitch-shift paradigm coupled with fMRI to determine the neural correlates of voluntary (Experiments 1 and 2) and more involuntary pitch-shift responses (Experiment 2), which essentially integrate features of auditory feedback with vocal motor control. We tested non-musicians and experienced singers to investigate vocal-training effects on behavioral performance and concomitant neural activity during audio-vocal integration.
Experiment 1

Our first experiment was designed to identify brain regions where audio-vocal integration may occur during voluntary vocal pitch regulation. In a “simple singing” task, we asked subjects to match five target notes while singing the syllable /a/ for four seconds. In two other singing tasks, auditory feedback was pitch-shifted by ±200 cents approximately one second after the trial began, and the shifted feedback was held constant throughout each trial. In the “ignore” task, subjects had to ignore the pitch shift and stay on their original vocal pitch, while in the “compensate” task, subjects were required to adjust their vocal output to correct fully for the shift. We believed that maintaining these responses across four seconds would both probe the degree of cognitive control over the incoming feedback signal and increase the likelihood of capturing neural activity during voluntary responses to feedback shifts in fMRI. Moreover, since subjects had to monitor auditory feedback continuously while adjusting their vocal output to ensure that they completely canceled out the feedback shift, we believed that the compensate task specifically would target the cortical substrates of audio-vocal integration, namely the auditory cortex, anterior insula, and ACC. Additionally, we tested non-musicians and experienced singers to assess whether vocal experience would alter the functional network for audio-vocal integration during singing. Singers should excel predictably at all singing tasks compared to non-musicians. We hypothesized that varying degrees of vocal pitch regulation are accompanied by different patterns of functional adaptations in the brain, and as such, singers would exhibit experience-dependent modulations of activity within the potential neural substrates for audio-vocal integration.

Despite expected differences in pitch matching, both groups recruited the same brain regions for simple singing, which included auditory cortex and many regions outlined in both of Jürgens’ networks for vocalization (Figure 2a): primary motor cortex, ACC, thalamus, and cerebellum. During the ignore task, non-musicians could not suppress involuntary pitch-shift responses and were approximately 75 cents away from the target notes, while singers successfully ignored the 200-cent pitch shift and maintained the original target notes as expected (Figure 2b, left). Singers’ enhanced performance was accompanied by increased activity in bilateral auditory cortex, relative to non-musicians (Figure 2b, right); this might be attributed to some singers monitoring the pitch stability of the altered feedback. Surprisingly, both groups showed similar performance profiles during the compensate task (Figure 2c, left), which engaged the intraparietal sulcus (IPS) and the dorsal premotor cortex (dPMC) in both groups relative to simple singing (Figure 2c, right)—regions involved in spatial transformations of sensory input for motor preparation and selecting motor programs involved in sensory-motor associations, respectively. Since these regions were recruited in both groups also during the ignore task (compared to simple singing; Figure 2b, right), we suggested that the IPS is recruited during transformations of auditory input into spatial information within the frequency domain (i.e., up or down), which is then used by the dPMC to prepare a vocal response (e.g., maintain steady vocal output or correct for the pitch shift). Since non-musicians recruited the dPMC more than singers (Figure 2c, right), we proposed that this region was a basic, experience-independent substrate for audio-vocal integration. All subjects, regardless of vocal expertise, may recruit the dPMC when hearing a pitch shift to which they need to generate a specific response. However, the singers also engaged parts of our hypothesized cortical network for audio-vocal integration more than non-musicians: auditory cortex within the posterior STS (pSTS), and ACC (Figure 2c, right). The lack of significant insular activity may be due to the common recruitment of this region in both groups, albeit in a possibly subthreshold manner among the non-musicians (see (38)). We concluded that while non-musicians possess this hypothesized functional network—auditory cortex, insula, and ACC are all functionally connected to each other in both groups—it may take extensive vocal training and practice to recruit these regions specifically for audio-vocal integration.

Experiment 2

In the previous study, non-musicians exhibited automatic pitch-shift responses when they ignored pitch-shifted feedback; Experiment 2 was designed to determine the neural correlates that govern this involuntary pitch-shift response. Since singers exhibited better vocal performance in Experiment 1, we recruited another group of experienced singers to optimize the induction of reliable pitch-shift responses and to investigate its related neural activity. To distinguish between voluntary and involuntary forms of vocal pitch regulation, singers performed compensate and ignore tasks with both large and small pitch shifts, as well as a simple singing task with no altered feedback as a control condition. We believed that compensating for a 200-cent pitch shift (“COMP200” task) would represent voluntary vocal pitch regulation, and would therefore recruit an audio-vocal integration network as observed in singers from Experiment 1: auditory cortex, anterior insula, and ACC. In contrast, we expected that ignoring a 25-cent shift (IGN25 task) would elicit involuntary pitch-shift responses, since trained singers exhibit
FIGURE 2. Behavioral (left) and fMRI (right) results from Experiment 1. (a) During simple singing, non-musicians (green circles) were less accurate (error was greater or less than 0 cents) across five target notes than experienced singers (blue squares), but they recruited a similar functional network for this task (statistical conjunction between groups shown). (b) Non-musicians consistently displayed a pitch-shift response in the ignore task (response magnitude > 0 cents), while singers were more successful at this task. Both groups engaged the dPMC and IPS specifically for this task, but singers displayed more activity in planum temporale (PT) and other auditory areas (STG/STS) than non-musicians. (c) There were no significant group differences in compensating for the pitch-shifted feedback, and accordingly, both groups recruited dPMC and IPS for this task. However, non-musicians recruited more dPMC activity than singers, while singers relied more on pSTS and ACC for this task.

these responses to shifts smaller than 100 cents. We hypothesized that the PAG would be recruited during this involuntary form of vocal pitch regulation, due to its connections with the lateral lemniscus and inferior colliculus in the ascending auditory pathway and its previously implicated role in the Lombard reflex, another automatic audio-vocal response.

Singers displayed involuntary pitch-shift responses while ignoring both the 25- and 200-cent feedback shifts, but the responses observed in the IGN25 task were significantly larger than those in the IGN200 task (Figure 3a). This reiterates that the mechanisms governing involuntary pitch-shift responses are better suited to correct for minor perturbations and suggests that these mechanisms are less amenable to voluntary control during the IGN25 task, relative to the IGN200 task. Unfortunately, we did not observe significant changes in neural activity specifically associated with the IGN25 task. Given that the early, more involuntary pitch-shift response usually occurs at 100-150 ms after the pitch shift, the poor temporal resolution of fMRI techniques may have failed to capture neural activity attributed to this response. Thus, our imaging results may be associated with late pitch-shift responses; methods with better temporal resolution, such as electroencephalography (EEG) or magnetoencephalography (MEG), may be better suited to verify the role of the PAG in audio-vocal integration during involuntary vocal pitch regulation. Nevertheless, irrespective of the feedback-shift magnitude, the shared process of maintaining vocal pitch in the presence of pitch-shifted feedback recruited bilateral planum temporale and right Brodmann areas 6/44 (ventral premotor cortex and inferior frontal gyrus, pars opercularis; Figure 3b).
FIGURE 3. Behavioral and fMRI results for Experiment 2. (a) Response magnitudes as percentages of the pitch-shift magnitude. The horizontal dashed line at 100% represents both full voluntary correction for the COMP200 and COMP25 tasks and complete involuntary correction for the IGN200 and IGN25 tasks. (b) Both IGN tasks recruited bilateral planum temporale (PT) and right Brodmann areas (BA) 6/44 (statistical conjunction between tasks shown). (c) COMP200 and COMP25 both engaged the IPS and the hypothesized network for audio-vocal integration: ACC, anterior insula, and auditory cortex. (d) The larger shift in IGN and COMP tasks specifically enhanced functional connectivity between a pSTS seed voxel and right IPS, relative to simple singing.

During the compensate tasks, singers did not fully correct for the 200-cent shift and overcorrected for the 25-cent shift (Figure 3a). Yet, the audio-vocal integration that presumably occurred during these tasks recruited the IPS and the hypothesized regions for audio-vocal integration: pSTS, anterior insula, and ACC (Figure 3c). Interestingly, the 200-cent pitch shift in both ignore and compensate tasks recruited more activity within posterior auditory cortex (compared to 25-cent shifts) 50, and specifically enhanced functional connectivity between pSTS and IPS (relative to simple singing; Figure 3d); a similar modulation in BOLD signal and functional connectivity was not observed with the 25-cent shift in either task. This magnitude-specific modulation in neural activity complements previous reports of posterior auditory cortex and IPS exhibiting greater sensitivity to larger pitch changes 51, 52. We proposed that the larger 200-cent pitch shift is processed by both the auditory cortex and IPS to encode the direction and magnitude of the feedback shift, which are then passed along to the rest of the audio-vocal integration network (anterior insula and ACC) for preparation and initiation of the appropriate task-dependent vocal response.

CORTICAL INTEGRATION OF AUDITORY FEEDBACK WITH VOCAL CONTROL

Across both experiments, we discovered two cortical networks that integrate features in auditory feedback with voluntary vocal motor control. The basic, experience-independent network—comprised of the IPS and dPMC (Figure 4a)—is active in both non-musicians and experienced singers, but is more heavily relied upon by non-musicians. We propose that the IPS encodes information about the pitch shift (e.g., magnitude and direction) and interacts with the dPMC via insular projections 20, such that this pitch-shift information is used to select the learned or associated motor program appropriate for the task (i.e., ignore or correct for the pitch shift); the selected vocal motor output is then generated via the primary motor cortex as per Jürgens’ (2009) model (i.e., generation of learned vocalizations). The experience-dependent network containing the IPS, posterior auditory cortex, anterior insula, and ACC (Figure 4b)—while functionally connected in both non-musicians and experienced singers 47, 57, 58—appears to be engaged only in experienced singers. Perhaps it is the intensive vocal training and practice, in which singers may
FIGURE 4. Updated models of vocal control and audio-vocal integration based on our fMRI results with the pitch shift paradigm. The shifted feedback (dashed blue arrow) is processed by the ascending auditory pathway, and the pSTS and IPS work jointly to process features of the pitch shift (solid double-headed arrow). (a) In this experience-independent network (green solid arrows), pitch-shift information is transmitted via the anterior insula (aINS) to dPMC for selection of the learned vocal response and eventual execution through M1 [and its modulatory loops (dimmed blue boxes)]. (b) The experience-dependent network (red solid arrows) is engaged after extensive vocal training and practice. The pitch-shift information is sent via the aINS to the ACC, which initiates a voluntary response to the pitch shift.

monitor auditory feedback closely to assess pitch accuracy and stability, which specifically recruits and strengthens this network. This network may then become the preferred network for voluntary vocal pitch regulation, within which posterior auditory cortex and IPS extract and encode the pitch-shift magnitude and direction from auditory feedback, and then send that information via the anterior insula to ACC vocal motor areas to voluntarily initiate [as per Jürgens’ (2009) model] a vocal response to the pitch-shifted feedback.

A LARGER FRAMEWORK

The two audio-vocal integration networks outlined above encompass the dorsal auditory stream, and functionally resemble two dorsal-stream models focused on auditory-motor transformations. Hickok and Poeppel [53-55] put forth a speech-processing model in which the dorsal stream transforms auditory representations of speech into motor representations of speech gestures. An alternative model from Warren, Wise, and Warren [56] expanded this auditory-motor transformation to include both speech-related and auditory spatial information—relevant sound features are extracted from auditory stimuli and compared with stored templates of responses, and then the proper response representation is sent via the dorsal stream to prefrontal and premotor areas for preparation of motor output. Importantly, both models argue that the planum temporale (i.e., posterior, higher-order auditory cortex) is the most important region for auditory-motor transformations. Based on our findings, we suggest that posterior auditory cortex constitutes only one part of a functional network for general auditory-motor transformations (including audio-vocal integration). Extractions and transformations of relevant auditory features may be governed by posterior auditory cortex and the IPS. The transformed auditory info is then sent via the anterior insula to premotor areas—in our studies, only dPMC in non-musicians, and dPMC and/or ACC in singers—for the matching and preparation of task-appropriate motor responses.

It is plausible that our functional networks may be specific to vocalization, since posterior auditory regions and dPMC were engaged also when the first formant frequency was shifted in auditory feedback [57]. However, a recent fMRI study employing auditory perception tasks has demonstrated that similar regions (e.g., posterior auditory cortex, IPS, anterior insula, premotor cortex) are also involved in auditory-motor transformations that are not related to speech or song production [58]. Future research will elucidate whether these networks serve a more general purpose of transforming auditory input into motor programs, regardless of whether these transformations are vocalization-related.
ACKNOWLEDGMENTS

The author thanks Robert J. Zatorre, Ph.D. and David Poeppel, Ph.D. for their mentorship and support. This work was supported in part by grants from the GRAMMY Foundation®, the Eileen Peters McGill Majors Fellowship, and the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT).

REFERENCES


