Journal of Experimental Psychology:
Human Perception and Performance
2015, Vol. 41, No. 1, 152–166

0096-1523/15/$12.00 http://dx.doi.org/10.1037/a0038285

The Role of Auditory Feedback in Speech and Song

Tim A. Pruitt and Peter Q. Pfordresher
University at Buffalo, State University of New York

When singing a melody or producing sentences, we take for granted the fact that the sounds we create (auditory feedback) match the intended consequences of our actions. The importance of these perception/action matches to production is illustrated by the detrimental effects of altered auditory feedback (AAF). Previous research in the domain of music has shown that when AAF leads to asynchronies between perception and action, timing of production is disrupted but accuracy of sequencing is not. On the other hand, AAF manipulations of pitch disrupt sequencing but not timing. Such dissociative effects, as well as other findings, suggest that sensitivity to AAF may be based on hierarchical organization of sequences. In the current research we examined whether similar effects are found for the production of speech, for which syllables rather than pitches may constitute content units. In the first experiment, participants either sang melodies or spoke sequences of nonsense syllables. In the second experiment, the tasks were combined such that participants sang syllable sequences. Production in both experiments was accompanied by either normal, asynchronous, or content altered auditory feedback. Across experiments, effects of AAF on the accuracy of sequencing were similar in speaking and singing tasks, and in all cases reflected the dissociative effects described earlier. For timing of production, however, previous results were only found when participants sang sequences that did not have varying syllabic content. These results suggest that sensitivity to timing exists at multiple hierarchical levels, particularly at the syllable and phonetic levels.

Keywords: auditory feedback, sequencing, timing, music and language

Anecdotes from drive-thru workers, cellular phone users, and online video game players tell of communication difficulties due to signal delays between speaking and hearing their own voice. This postponement in hearing self-produced auditory information results in speech disfluencies such as stuttering and repeating words. The novelty mobile phone application Speech Jammer (Hou, 2014) operates similarly by allowing users to implement a delay between the input to the device's microphone and its audio output. Speech Jammer has gained popularity on the Internet as users have posted YouTube videos documenting production disturbances during their attempts to read selections, give consumer reviews, or perform songs. Likewise, the speech jammer gun (Kurihara & Tsukada, 2012) technologically elaborates on this principle to create practical applications for crowd control or maintaining silent environments by disrupting speech without physically distressing its targets.

All of the earlier cases illustrate that even the slightest desynchronization between producing and subsequently hearing auditory information can have profound effects on speech. But given the lack of control in such real-world examples it is difficult to pinpoint the origin of this disruption. One possibility involves feedback *synchrony*, which refers to whether the onsets and offsets of speech sounds line up in time with each other. Asynchronies between actions (spoken syllables) and auditory feedback have been the focus of accounts for such disruptive effects. However, another possibility emerges from cases in which the resulting content of auditory feedback has been altered such that the categorical event (a syllable) of feedback does not match the intended event. If, for instance, a feedback delay is as long as a spoken syllable, then the speaker would hear the previous syllable when generating the current syllable and any resulting disruption would reflect a deviation in content rather than asynchronous timing.

The distinction between feedback content and timing is critical here because it bears on the nature of mapping between perception and action. Previous research in the domain of music, reviewed in the following, has suggested that these alterations have distinct effects on production thus suggesting that perception and action associations are constrained by the temporal hierarchy used to represent the structure of a sequence. However, no research to date has addressed whether comparable effects may occur for speech, thus leaving open the question of whether perception/action association in speech relate to those of music.

In light of this, we report on two experiments that address critical questions involving how sensory information relates to motor information. First, do people use feedback to guide speech in the same way that they use feedback to produce melodies? This question reflects a critical debate in the current literature regarding representations used to process music versus language (e.g., Patel, 2008). Second (and related), to what degree is the use of feedback

This article was published Online First November 10, 2014.

Tim A. Pruitt and Peter Q. Pfordresher, University at Buffalo, State University of New York.

This research was supported in part by NSF Grants BCS–0642592 and BCS–1256964. We are grateful to Anastasiya Kobrina and Esther Song for assistance with data collection, and to Pauline Larrouy-Maestri, James Mantell, Kathleen Jocoy, Ken Steele, James Sawusch, and Eduardo Mercado III for helpful comments on an earlier version of this paper.

Correspondence concerning this article should be addressed to Tim A. Pruitt or Peter Q. Pfordresher, Department of Psychology, 204 Park Hall, University at Buffalo, State University of New York, Buffalo, NY 14260. E-mail: tapruitt@buffalo.edu or pqp@buffalo.edu

in either domain constrained by the hierarchical control of actions? Hierarchical control is considered to be a critical part of action planning for most sequential behaviors (e.g., MacKay, 1987; Rosenbaum, 2010) yet it is not clear whether the nature of this hierarchy is comparable (or to what degree) across the domains of music and language.

The Role of Auditory Feedback in Sequential Behaviors

Fluent production of sequential behaviors is, in part, reliant on the coordination between perception and action. This relationship is especially important during the production of speech and music. Whereas both tasks require actions to be arranged into complex sequences designed to produce a series of auditory events in which the timing and content are critical to communication. These tasks also involve coordination among cognitive planning and motor execution while concurrently processing the auditory consequences (here referred to as *auditory feedback*) of actions. Under normal circumstances auditory feedback complements production with respect to both the synchronization of an action and its feedback as well as the actual content (linguistic units, musical pitches) of feedback events themselves. The importance of this compatibility is demonstrated by the disruptive effects of altered auditory feedback (AAF). When auditory feedback is delayed with respect to onset timing or altered with respect to feedback content, production suffers. Such disruptive effects of AAF have been observed across many tasks, including tapping (Finney & Warren, 2002; Ruhm & Cooper, 1963), clapping and whistling (Kalmus, Denes, & Fry, 1955), speech (Black, 1951; Fairbanks, 1955; Lee, 1950; MacKay, 1968; Yates, 1963), and music (Finney, 1997; Gates & Bradshaw, 1974; Gates, Bradshaw, & Nettleton, 1974; Havlicek, 1968; Pfordresher, 2003). Speech and singing are particularly interesting as both tasks involve using the same peripheral motor and perceptual systems, thereby providing an opportunity to determine whether auditory feedback is used to coordinate with actions similarly across domains.

In the music literature, it has been shown that performers are primarily sensitive to temporal relationships between auditory feedback and actions that give rise to melodies. However, such temporal coordination is not limited to onset synchrony between actions and feedback. When producing a melody, planning and execution operate along higher order time-scales that regulate the serial ordering of event sequences. It has long been known that the production of instrumental music can be disrupted by delaying auditory feedback (Gates & Bradshaw, 1974; Gates et al., 1974; Havlicek, 1968). However, altering feedback content (i.e., changing the pitch of a feedback events while preserving synchrony with actions) to produce a random-like pitch sequence fails to significantly disrupt performance (Finney, 1997; Pfordresher, 2005). A similar content manipulation that results in a coherent melody that is structurally distinct from the intended melody also fails to be disruptive (Pfordresher, 2005). On the other hand, hearing pitches from the current melody at *serially shifted* positions (e.g., every piano key press generates the pitch associated with two events previous) is significantly disruptive. Such serially shifted feedback creates a mismatch between action and perception with respect to the serial pattern. Further evidence suggests that in music the basis of this disruption may be primarily due to mismatches in the melodic contour (Pfordresher, 2008).

More important, AAF effects in music suggest that the coordination of perception and action may be based on distinct mechanisms for sequencing and timing. Short-latency asynchronies (i.e., 25–400 ms) of feedback slow production and increase timing variability, but do not interfere with the performer's ability to accurately sequence pitches of a melody. By contrast, altered content, such as serially shifted feedback, does disrupt the accuracy of sequencing, but does not affect performance timing (Pfordresher, 2003). Thus, the maintenance of timing in production depends on synchronization between perception and action. Retrieval for serial order however, depends on the content alignment between produced actions and the expected perceptual consequences. This sequencing/timing dissociation has been observed in behavioral tasks (Couchman, Beasley, & Pfordresher, 2012; Pfordresher, 2003; Pfordresher & Kulpa, 2011; Pfordresher & Mantell, 2012) and recently has been supported at the neural level as well (Pfordresher, Mantell, Brown, Zivadinov, & Cox, 2014). Moreover, the observed independence begs the question of how such representations interact and how are they cognitively organized? One account suggests that auditory feedback and action planning may share resources within a common representation of the temporal hierarchy used to organize event sequences (Pfordresher, 2006; cf. Hommel, Müsseler, Aschersleben, & Prinz, 2001; MacKay, 1987; Shin, Proctor, & Capaldi, 2010). The observed disruptive effects of AAF have been interpreted as supporting the notion of domain-general action-perception codes, but the evidence favoring such representations being employed outside the context of music—with speech being of particular interest—is lacking.

Music and Language: Separate or Integrated Perception/Action Systems?

Given these aforementioned conclusions, one may wonder how auditory feedback is used in other production domains. As the distinction between sequencing and timing has been applied to domains beyond music (Krampe, Mayr, & Kliegl, 2005; MacKay, 1987), it is plausible to expect that the use of auditory feedback in speech production would reflect a similar temporal organization. Some support for this assumption derives from a recent study that demonstrated the sequencing/timing dissociation in keyboard and vocal music production (Pfordresher & Mantell, 2012). Although singing can be characterized as being more closely related to speech production, it is possible that shifting domains from music to speech may bring about a more profound change than simply changing the effector (i.e., hand-digit or vocal motor) system. Evidence from neuropsychological dissociations has been taken as supporting a modular architecture in which speech and music are separate (Peretz & Coltheart, 2003), including the control of timing. Even moderate views of music and language typically stop short of promoting full integration, such as Patel's (2008) suggestion that domains share neural/cognitive resources while operating on distinct representations. Based on such claims, one might expect effects of AAF to reflect a different kind of temporal organization for the domains of music and speech.

As for the specific differences across music and language that may arise, drawing on previous research we propose two alternatives. One possibility is that the temporal hierarchy used to coordinate perception and action for speech will include a broader range of levels than for music, reflecting the greater temporal complexity in the speech signal. This hypothesis derives in part on the inherent differences in spectra-temporal variability across the speech and music auditory signals. In particular, speech contains important acoustic characteristics extrapolated from time scales corresponding to phonetic information. As a result, the speech action-perception system may require more sensitivity to information commensurate to these finer grained time scales and therefore may use auditory feedback differently. Indeed, previous research has repeatedly observed hemispheric lateralization of speech versus music processing, in which activity is dominant in left hemisphere areas with linguistic signals while there is right hemisphere dominance for musical signals (Peretz & Zatorre, 2005; Saito, Ishii, Yagi, Tatsumi, & Mizusawa, 2006; Wong, Parsons, Martinez & Diehl, 2004). Findings from other studies using both covert and overt speaking and singing tasks further align with these hemispheric asymmetries (Ackermann & Riecker, 2004; Riecker, Ackermann, Wildgruber, Dogil, & Grodd, 2000). One account for hemispheric lateralization posits that particular brain areas, such as the temporal lobes, have become specialized for processing domain-specific signals (Zatorre & Belin, 2001; Zatorre, Belin & Penhune, 2002). Given the need for high spectral resolution in the processing of music, the right temporal lobes have been specialized for fine-grained spectral processing. An alternative, speech perception requires a high degree of temporal resolution due to the signal's fine-grained spectra-temporal variability (corresponding to phonetic and feature components).

Additional support for broader temporal sensitivity derives from cognitive models of speech production. Typically these models conform to a hierarchical structure with higher order grammatical and lexical representations and, more relevant to the discussion at hand, lower order phonological and feature level representations (Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Although it is commonly agreed that syllables are rhythmic units of speech, it is possible that AAF effects will reflect sensitivity to temporal coordination at both the phonemic and syllabic levels.

Another type of difference between music and speech domains may occur at the sequencing level. It is important to note the previously observed disruptive effects of altered content have only been found in the domain of music. We know of only two examples of roughly analogous manipulations in speech (Kaspar & Rübeling, 2011; Müller, Aschersleben, Esser, & Müsseler, 2000), but neither of these led to effects comparable to the sequencing/timing dissociation.[1] In addition, Howell (2004a; Howell & Archer, 1984; Howell, Powell, & Khan, 1983) proposed that, at least for speech, the role of auditory feedback may be limited to onset timing based on the amplitude contour. Disruption results from perturbations to an internal timekeeper rather than a more elaborate cognitive representation of sequence structure. A good deal of support has emerged for this view from within the domain of speech (Howell, 2004a, 2004b, 2007; Howell & Archer, 1984; Howell & Sackin, 2002; Kaspar & Rübeling, 2011). However, to date we know of no study that has compared speech and music production directly with respect to both feedback manipulations and the structure of sequences being produced.

The Present Experiments

The current experiment elaborates on results by Pfordresher and Mantell (2012) which suggested that shared perception-action associations deployed during music performance were not effector specific. This effector independence further implies that shared representations operate at an abstract level. The current work logically follows their findings and seeks to explore whether action-perception associations are shared across production domains. In addition, it is of further interest to determine whether sequencing and timing representations are similarly separated in speech as in music production.

As such, the present experiments sought to determine whether a separation of sequencing and timing, with sequencing being established at the level of syllables, exists in the use of auditory feedback for speech. We adopted a procedure similar to that reported by Pfordresher and Mantell (2012), which involves manipulations of feedback delays for production of sequences at a fixed rate. Produced sequences could involve event variability based on different syllabic content and/or pitch content. Finding a qualitatively and quantitatively similar sequencing/timing dissociation in both the production of syllable and pitch sequences would provide strong support for a common system used to coordinate perception and action. Differences in the magnitude of the dissociation, by contrast, could suggest mediating effects of the auditory signal's features (e.g., use of fine-grained temporal resolution) on a potentially shared system. Finally, an absent dissociation (or a qualitatively different dissociation) in speech would suggest that a fundamentally different type of perception/action coordination exists across domains.

Experiment 1: Singing and Speaking Tasks

Experiment 1 involved two separate production tasks: a sung pitch sequence task and a spoken syllable sequence task. For the singing task, participants repeatedly sung an eight-note melody on a single syllable ("La"). The speaking task required participants to produce an eight-syllable sequence while keeping pitch constant. A metronome was used to control timing at 600 ms per pitch/syllable, and feedback manipulations were designed to disrupt feedback timing (a delay of 300 ms), or sequencing (a delay of 600 ms). Feedback conditions in which the auditory signal was delayed by 300 ms is henceforth referred to as *asynchronous feedback*, while conditions with a sequencing delay of 600 ms is known as *serially shifted feedback*. The latter manipulation is an approximation of a lag–1 serial shift, a feedback alteration similarly used in studies with spoken syllables (Kaspar & Rübeling, 2011; Müller et al., 2000) and found to significantly disrupt music production (Pfordresher, 2003, 2005, 2012; Pfordresher & Mantell, 2012).

Given the previously observed dissociative effects of sequencing and timing manipulations, we hypothesized that in both singing and speaking tasks asynchronous feedback should disrupt performance timing, while sparing sequencing accuracy. Furthermore, serially shifted feedback should selectively disrupt sequencing while not significantly affecting production timing across tasks.

[1] We should point out that production rate was not fully controlled in either of these studies, so comparability with previous serial shift manipulations must be treated tentatively.

Taken together, observing such a double dissociation would be evidence for speech and music using auditory feedback similarly during production. On the contrary, finding that either altered feedback condition does not selectively disrupt performance timing or selectively disrupt sequencing suggests domain differences in how auditory feedback is used to coordinate action with perception.

Method

Participants. Sixteen undergraduate students were recruited from the University at Buffalo introductory psychology subject pool. Eight participants were women and eight were men with a mean age of 18.5 years (range 18–21 years). We assessed the presence of speech and hearing disorders, specialized musical abilities (e.g., absolute pitch), and demographic information through self-report questionnaires. No participants reported any hearing problems or absolute pitch and all participants reported right-hand dominance.

Participants were sampled without regard to musical training and four participants reported no training or musical experience with any instrument. Six participants reported 1 or 2 years of formal training on various instruments and were classified as having minimal training. The remaining five participants reported 5 or more years of musical training and were classified as musicians. Two of these musicians indicated they had received extensive vocal training (one reported 6 years of vocal lessons and the other 8 years of chorus). Of these two, one reported training on three additional musical instruments: 9 years of piano lessons, 4 years of clarinet training, and 3 years of guitar training.

We conducted several analyses to determine whether effects of AAF reported later were modulated by years of musical experience. In line with previous research (Pfordresher, 2005; Pfordresher & Mantell, 2012), none of these analyses were significant. Thus we report only analyses based on averages across all participants.

Materials. Two eight-note melodic sequences comprising five diatonic pitch classes from the C-major scale were constructed for the singing task. These melodies have been previously used in keyboard and other singing tasks (Pfordresher, 2005; Pfordresher & Mantell, 2012). One melody began on C and proceeded through a scalar contour (C D E G F E D E) whereas the other began on G and featured an alternating contour (G E F D C E D F). The auditory stimuli used in singing trials were generated in Yamaha's Vocaloid Leon software package (Zero-G Limited, Okehampton, England). We synthesized gender-specific sequences whose pitch and formant structure were similar to those typical to male or female voices with the most notable difference being that female sequences were an octave higher than the male counterparts. The temporal spacing of each note was set to 600 ms inter-onset-intervals (IOI) to match the prescribed production rate of 100 beats per minute (BPM). We also perceptually matched each note's loudness so that no single note's loudness was different from others in the sequences. More important, the phonetic composition of these stimuli was held constant across all sung notes such that each note was sung on the syllable "La." Thus the only variations in the melodic sequences were the changes in pitch.

We then mapped a consonant vowel (CV) nonsense syllable onto each of the five selected diatonic pitches that composed the melodic sequences. For example, the pitch C was associated with the nonsense syllable "Loo" and anytime the C pitch appeared in a melodic sequence the "Loo" syllable would appear in the same serial position of its syllable sequence counterpart. This was done to create novel syllable sequences that were structurally isomorphic to their melodic counterparts. As a result, there were two eight-syllable speech sequences, each directly associated with one of the melodic sequences (see Figure 1.). The melody beginning on the pitch C was associated with the syllable sequence beginning with "Loo" (Loo Bah Gee Tay Poh Gee Bah Gee), while the melody beginning with the pitch G was associated with the other syllable sequence (Tay Gee Boh Pah Loo Gee Bah Poh).

When constructing the auditory stimuli for the spoken syllable sequences we elected to record and modify natural speech because we found that synthesized files, though ideal with respect to the accuracy and precision of pitch (for speech), sounded unnatural as speech targets.[2] This was done by digitally recording a male and female model repeating a single target syllable (e.g., Loo) 15 to 20 times. They were instructed to maintain a monotone pitch and to clearly articulate the syllables to the best of their abilities. On obtaining a recorded set of each nonsense syllable, we then selected the single best iteration of each target syllable along the dimensions of syllabic articulation clarity, minimal pitch glide, and appropriate syllable duration that was no longer than 500 ms. We then constructed the two eight-syllable sequences by concatenating the target syllables in the appropriate serial order and temporally separating syllable onsets by 600 ms. Syllable onsets were determined by identifying the amplitude rise, or its deviation from the zero crossing, of the syllables first phonetic segment. To further control for variations in the auditory signal each syllable was normalized for perceptual loudness and pitch. These alterations were conducted to minimize the perception of a pitch contour in the speech sequences, thus the only variations in the sequence could be attributed to the changes in the syllabic information.

Conditions. During experimental trials, auditory feedback associated with the experimental manipulation was presented over noise attenuating headphones. In addition, we presented pink noise to further mask air-conducted feedback. Participants would experience one of three different auditory feedback conditions on each individual trial. These conditions included normal unaltered feedback, which served as a control, and two types of AAF: asynchronous and serially shifted feedback. AAF was manipulated by altering the auditory signal via the software package Cubase (Steinberg Media Technologies, Hamburg, Germany). In particular, auditory feedback was delayed by 300 ms to produce the asynchronous effect. Likewise, a delay equal to the 600 ms prescribed IOI was used in the serially shifted condition. Such a delay condition approximates the experience of hearing lag-1 serially shifted feedback used for the production of musical sequences on a musical instrument digital interface (MIDI) keyboard. This method to simulate both the asynchronous and serially shifted AAF was the same used in Pfordresher and Mantell (2012).

[2] It is important to note that slight differences between the recorded targets for speech and song are not critically important for this design. The role of target stimuli was simply to cue production for the participant. Auditory feedback, the critical acoustic stimulus for this study, was always based directly on the participant's voice.

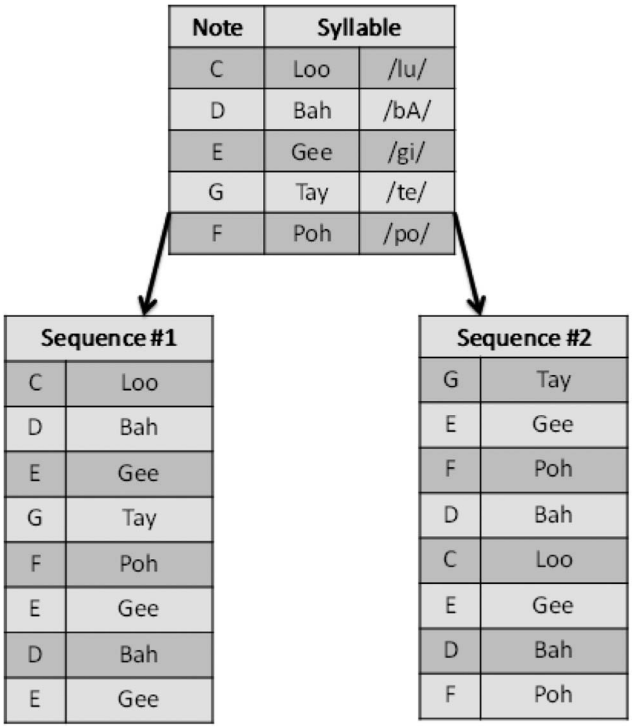

Figure 1. Note to syllable mapping.

Apparatus. Melodic stimuli were synthesized using the software package Vocaloid whereas speech stimuli were digitally recorded using Praat (Boersma & Weenink, 2014). Both singing and speaking tasks were conducted inside a sound attenuated chamber (Whisper Room, SE 2000 Series, Morristown, TN) and stimuli were presented through Sennheiser HD 280 Pro headphones (Sennheiser, Wedemark, Germany). A Shure PG58 microphone (Shure, Niles, Illinois) channeled through a Lexicon Omega recording interface (Harman, Stamford, Connecticut) was used to capture participant recordings. The microphone was mounted on a stand and the experimenter adjusted the microphone such that there was a 2-in. distance between it and the participant's mouth. The digital sequencing program Cubase was employed to deliver stimuli and feedback alternations as well as record participant vocalizations. Cakewalk Delay (Cakewalk, Boston, Massachusetts), a virtual studio technology plug-in, was used within the Cubase interface to alter the audio signal in real time.

Procedure. The experiment was divided into separate sections in which participants either sung melodies or spoke nonsense syllable sequences. These sections were presented in different blocks with the order of the blocks counterbalanced across participants. Before the speaking task, participants were instructed to engage in a warm-up exercise where they read the Rainbow Passage (Fairbanks, 1960). Participants then proceeded to a learning phase where they memorized the nonsense syllable sequence in which they would be producing during the experimental trials. A single learning trial contained six repetitions of the model sequence. Participants were instructed to only listen to the first learning trial to help commit the sequence to memory. On subsequent learning trials participants were encouraged to speak along with the stimulus sequence. Learning trials could be repeated as many times as needed until both the experimenter and participant were confident that the sequence was memorized. Following the learning phase, participants experienced a single practice trial and then the nine experimental trials. On each trial, participants first heard the model sequence once and were instructed to begin producing the sequence themselves immediately afterward. A metronome set at 100 BPM (600 ms IOIs) concurrently played during the initial participant production to facilitate the entrainment of the prescribed production tempo. The metronome stopped after eight clicks and participants continued repeated production of the sequence for 37 s until a burst of white noise signaled the end of the trial. AAF manipulations were initiated after the participant's initial production of the sequence and continued for the remainder of the trial.

The singing procedure was nearly identical to the speaking procedure but with minor task-specific variations. During the warm-up phase, participants were asked to sustain a single pitch in which they were comfortable singing as well as singing a rendition of "Happy Birthday." The structure of the learning, practice, and experimental trials only differed with respect to the type of sequence the participant practiced. There were a total of 18 experimental trials with nine trials for each of the speaking and singing tasks. The three feedback conditions were each presented three times and participants experienced all three conditions before any one condition was repeated.

Data processing and analysis. Sung and spoken syllable events were analyzed by first demarcating the event onsets with Praat Spectral, amplitude, and pitch tracking information was used to determine both sung and spoken syllable onsets. Acoustic–phonetic information, such as formant transitions, was additionally used in the identification of spoken syllable onsets. Once boundaries were identified, IOIs were calculated by measuring the time between the current onset and the proceeding event's onset.

We measured disruption of timing using the mean IOI within a trial. Generally speaking, mean IOIs approximated the prescribed IOI during normal feedback and increased when disrupted by AAF. The aforementioned sequencing/timing dissociation predicts that mean IOI should increase for asynchronous but not for serially shifted AAF. Thus, mean IOIs in asynchronous AAF conditions should be significantly higher than in the other two conditions, which should not differ.

Accuracy in production was based on the proportion of errors in pitch direction (contour) for singing, and the proportion of errors in the serial ordering of syllables for speech. In both cases we predict that disruption from AAF will increase errors. Moreover, the sequencing/timing dissociation predicts that the increase should only be found in conditions with serially shifted AAF. Thus, error rates in serially shifted conditions should be higher than error rates in the other conditions, which should not differ from each other. Errors rates in both tasks (singing and speaking) were derived from an algorithm that determined the minimum number of changes to the produced sequence necessary to match the target sequence (Large, 1993; Palmer & van de Sande, 1993, 1995). Prior to analyzing errors via this algorithm we coded sung and spoken productions in discrete quantitative units as follows.

Singing accuracy was measured by examining the produced melodic contour, the shape of the successive rising and falling pitches that compose a melody. This measure was used, rather than accuracy of sung pitches or intervals because it is a characteristic of pitch sequencing that most novice singers produce accurately under normal conditions (e.g., Pfordresher & Brown, 2007). A sung event's pitch was determined by extracting the middle 50% within the demarcated IOI boundaries and calculating the signal's median F_0 (median smoothing procedure; Gold, Morgan, & Ellis, 2011). This method of pitch estimation serves to reduce spurious F_0 measurements that result from pitch tracking artifacts or idiosyncratic voice qualities (e.g., "voice breaking") and has been used in previous studies (Pfordresher & Mantell, 2012). In-house Matlab scripts (Mathworks, Natick, MA) were used to extract these F_0 estimates and verified through visual inspection. Changes between pitches were then coded as $+1$ if they formed an upward pitch contour or -1 if they formed a downward pitch contour. We used only these two codes because no unisons were present in the stimuli, and performers did not seem to produce unisons.[3]

Spoken syllable sequences were transcribed by listening and annotating the produced utterances. Target syllables were assigned a specific MIDI code, thus an ideal sequence of uttered syllables has a corresponding MIDI code sequence (Loo = 60, Bah = 62, Gee = 64, etc.). The transcribed annotations for each produced event were then converted to a MIDI code. Correct utterances were converted to the assigned MIDI code, while annotations that deviated from target were assigned an error code (66). Doing so allowed the use of the same error detection algorithms for speech as those used for the musical sequence production analyses. These in house computer programs compare the target MIDI sequence to MIDI converted annotations.

We initially computed a two-way within-subject analysis of variance (ANOVA) with factors of feedback condition (normal, asynchronous AAF, and serially shifted AAF) and task (singing, speaking). Separate ANOVAs were computed for each dependent measures associated with testing the sequencing/timing hypothesis: IOI for timing and the proportion of sequencing errors within a trial as a measure of accuracy. The sequencing/timing dissociation is supported by main effects of feedback for each measure, reflecting increases of disruption for different conditions in each measure (increased IOI for asynchronous feedback, increased error rates for serially shifted feedback).

Given that a primary motivation of this research was to determine if the sequencing/timing dissociation is significant across speaking and singing tasks, we performed further analyses separately within song and speech production trials. In the first of these analyses, we computed one-way ANOVAs for both singing and speech to further examine the effects of feedback independent of the variance associated with potential interactions within selected trials. This also allowed us to incorporate post hoc measures to determine differences in means across feedback conditions.

Finally, we assessed the sequencing/timing dissociation in each domain using multiple regression analyses designed to precisely focus on how well patterns of data fit the predicted dissociation. Two predictor variables were derived by assigning effect codes for feedback conditions in two ways. The first effect code was based on the prediction that asynchronous feedback, but not serially shifted feedback, causes disruption of produced timing. In this effect code every asynchronous feedback condition was assigned the value 1 and all other conditions were assigned the value 0. The second predictor variable was based on effects coding from the prediction that serially shifted feedback, but not asynchronous feedback, disrupts production of content, with similar coefficients. We regressed each dependent variable on both predictor variables, and analyzed partial regression coefficients. The sequencing/timing dissociation is supported if the predictor variable based on disruption from asynchronous feedback predicts mean IOI, whereas the predictor variable based on disruption from serially shifted feedback predicts error rates. However, finding that both variables significantly predict a dependent measure suggests that the disruptive effects of AAF are not fully dissociable.

Results

Disruptive effects of AAF on timing. Figure 2A displays the effect of sequencing task and feedback condition on mean IOIs. Production rate slowed considerably during asynchronous feedback, as compared to performance with normal feedback, during both singing and speech. By contrast, serially shifted feedback led to IOIs that were negligibly higher than those seen in performances with normal feedback. The two-way omnibus ANOVA on mean IOI yielded a significant main effect of feedback condition, $F(2, 30) = 18.77$, $p < .001$, $\eta_p^2 = .56$. However, no significant main effect for task, $F(1, 15) = 0.33$, $p = .57$, $\eta_p^2 = .02$, and no significant interaction, $F(2, 30) = 2.65$, $p = .088$, $\eta_p^2 = .15$, was observed.

To examine whether the sequencing/timing dissociation was present within each production task we computed one-way ANOVAs on mean IOI within singing and speaking trials. There was a significant effect of feedback during the singing task, $F(2, 30) = 18.90$, $p < .001$, $\eta^2 = .56$. Mean differences in IOI across feedback conditions were ascertained by using Tukey's honestly significant difference (HSD) tests ($\alpha = .05$). The results of this post hoc test were in-line with the dissociation hypothesis, where mean IOIs were highest during asynchronous feedback ($M = 717.5$ ms, $SD = 128.8$) and significantly different from both normal ($M = 596.4$ ms, $SD = 37.5$) and serially shifted ($M = 643.1$ ms, $SD = 70.7$) feedback conditions. Furthermore, the mean difference in IOIs between normal and serially shifted feedback did not reach statistical significance.

As described in the Data Analysis section, we used effects coding in regression to determine whether the type of effects on IOI predicted by the sequencing/timing dissociation (slower IOIs than the prescribed tempo for asynchronous AAF, with faster and more accurate IOIs for other conditions) was a better predictor of timing than the alternate prediction (slowed IOIs for serially shifted AAF). The multiple regression equation with both predictors accounted for a significant proportion of the variance, $R^2 = .22$, $F(2, 141) = 20.69$, $p < .001$, and each predictor variable accounted for a significant proportion of the variance when variance from the other predictor was partialed out, for effect coding of disruption from asynchronous AAF, $\beta = .55$, $t(141) = 6.38$, $p < .001$; for effect coding of disruption from serially shifted AAF, $\beta = .21$, $t(141) = 2.47$, $p < .05$. However, a test of

[3] To verify this intuition, we examined the frequency of pitch interval production within bins of 25 cents for all participants and trials. This analysis indicated that only 1% of all sung intervals fell within the bin surrounding zero cents (no change); when the criterion was expanded to be ± 50 cents only 4% of all sung intervals. Thus, intervals coded as unisons were rare as well as being contrary to the putative intentions of the performer.

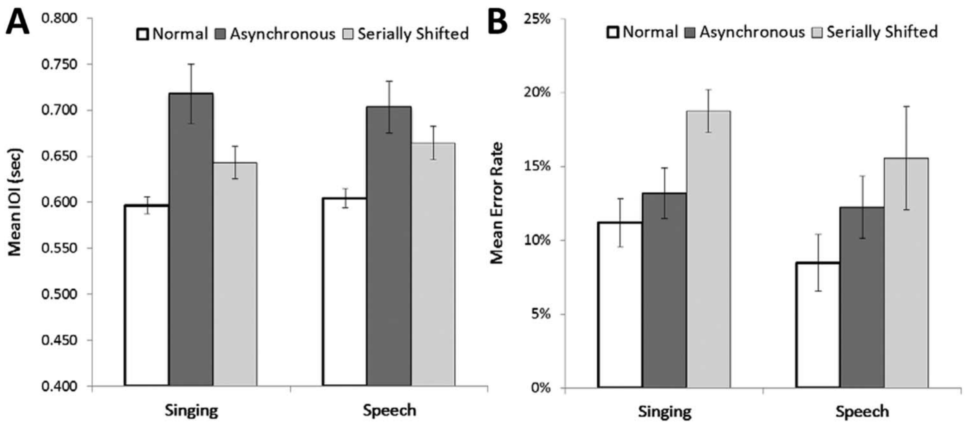

Figure 2. The effects of auditory feedback condition and task on production timing (A) and sequencing (B) for Experiment 1. Error bars represent standard errors. IOI = inter-onset-intervals.

magnitude for dependent β weights (Cohen, Cohen, West, & Aiken, 2003), suggested that the appropriate predictor variable (disruption from asynchronous AAF) accounted for more variance than the other variable, $t(140) = 2.00$, $p < .05$.

The evidence in support of the dissociation hypothesis for mean IOIs was less robust in speech task. The results of the one-way ANOVA showed a significant effect of feedback, $F(2, 30) = 13.65$, $p < .001$, $\eta^2 = .47$. However, Tukey's HSD tests ($\alpha = .05$) revealed significant production rate slowing in both asynchronous ($M = 703.5$ ms, $SD = 112.7$) and serially shifted ($M = 664.3$ ms, $SD = 72.7$) conditions when individually compared to the normal ($M = 604.3$ ms, $SD = 41.3$) feedback condition, and the two AAF conditions did not differ from each other. Furthermore, the regression analysis, $R^2 = .18$, $F(2, 141) = 15.6$, $p < .001$, suggested that both predictors accounted for a significant proportion of the variance in mean IOI, disruption from asynchronous AAF, $\beta = .49$, $t(141) = 5.54$, $p < .001$; disruption from serially shifted AAF, $\beta = .30$, $t(141) = 3.35$, $p < .01$; and the test for magnitude of dependent beta values was not significant, $t(140) = 1.58$, $p = .12$. Thus, both AAF conditions lead to an increase in mean IOIs for the speaking task.

Disruptive effects of AAF on sequencing. The mean percentage of errors for both syllabic and melodic sequence production tasks are shown in Figure 2B. The results here are consistent with the dissociation hypothesis in that more production errors were committed during serially shifted feedback exposure compared to normal feedback performance. An ANOVA on mean percentage of errors yielded significant main effect for feedback condition, $F(1, 15) = 16.65$, $p < .001$, $\eta_p^2 = .53$. However, the main effect of task and the feedback by task interaction failed to reach significance, $F(1, 15) = 1.02$, $p = .33$, $\eta_p^2 = .06$, and no significant interaction, $F(2, 30) = 0.29$, $p = .75$, $\eta_p^2 = .02$, respectively.

Separate one-way ANOVAs on mean percentage of errors were computed to examine the significant effect of feedback within each production task. For singing, there was again a significant effect of feedback, $F(2, 30) = 11.41$, $p < .001$, $\eta^2 = .43$. Post hoc tests indicated that mean percentage of errors were highest in the serially shifted ($M = 18.7\%$, $SD = 5.7\%$) feedback condition and was significantly different compared to both normal ($M = 11.2\%$, $SD = 6.5\%$) and asynchronous feedback ($M = 13.7\%$, $SD = 6.8\%$) conditions, which did not differ from one another. In the regression analysis, $R^2 = .17$, $F(2, 141) = 14.52$, $p < .001$; only the predictor associated with disruption from serially shifted feedback accounted for a significant proportion of variance when the other predictor was partialed out, disruption from serially shifted AAF, $\beta = .46$, $t(141) = 5.20$, $p < .001$, disruption from asynchronous AAF, $\beta = .12$, $t(141) = 1.39$, $p = .17$.

The results of the one-way ANOVA on mean percentage of errors during syllable production offered some support for the dissociation hypothesis in that there was a significant effect of feedback, $F(2, 30) = 4.65$, $p = .02$, $\eta^2 = .24$, and post hoc tests revealed higher errors during serially shifted feedback ($M = 15.6\%$, $SD = 13.9\%$) than normal feedback ($M = 8.5\%$, $SD = 7.7\%$). In addition, error rates did not differ between normal and asynchronous feedback ($M = 12.2\%$, $SD = 8.4\%$) conditions. However, contrary to the sequencing/timing dissociation hypothesis, errors during serially shifted feedback were not significantly higher than errors during asynchronous feedback. Similar to the error data in the singing task, the regression analysis of the syllable errors, $R^2 = .22$, $F(2, 141) = 3.51$, $p < .05$, supported the dissociation hypothesis. Again, the only predictor variable accounting for a significant independent proportion of the variance was based on effects coding of disruption from serially shifted feedback, disruption from serially shifted AAF, $\beta = .25$, $t(141) = 2.65$, $p < .01$, disruption from asynchronous AAF, $\beta = .13$, $t(141) = 1.41$, $p = .16$. Thus, in general the speech error data support the dissociation hypothesis, with the exception of one post hoc contrast.

Discussion

The results of the first experiment replicate Pfordresher and Mantell (2012) in that they demonstrate the sequencing/timing dissociation within vocal music production. Figure 3 shows a comparison between the results of the first experiment and the relevant conditions from Pfordresher and Mantell's (2012) singing data. With respect to production timing (Figure 3A), the pattern of AAF's disruptive effects is highly similar across experiments and asynchronous feedback leads to the most disruption. Likewise, disruption to pitch sequencing (Figure 3B) is comparable across

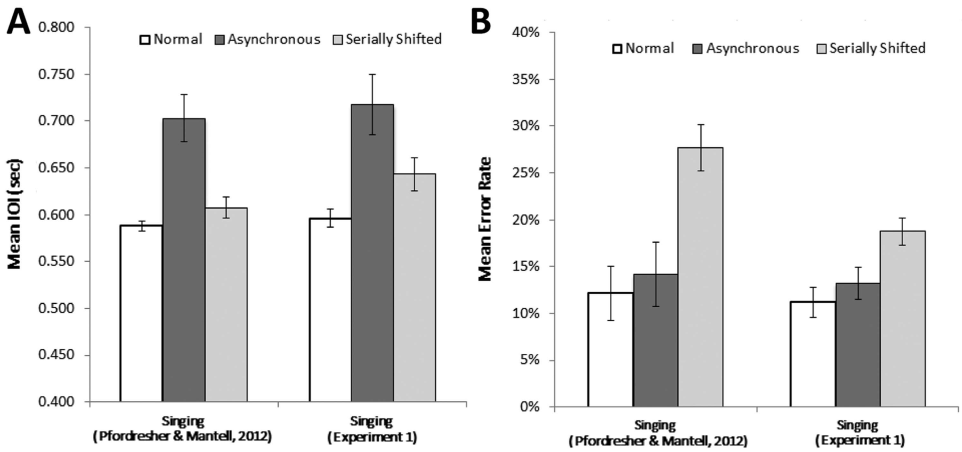

Figure 3. A comparison of the altered auditory feedback disruption on production timing (A) and note sequencing for sequencing data between Experiment 1 and the relevant conditions selected from Pfordresher and Mantell (2012). IOI = inter-onset-intervals. From "Effects of Altered Auditory Feedback Across Effector Systems: Production of Melodies by Keyboard and Singing," by P. Q. Pfordresher & J. T. Mantell, 2012, *Acta Psychologica, 139,* p. 172. Copyright 2012 by Elsevier.

experiments and in both cases only serially shifted feedback led to significantly more errors.

The most notable difference between the current results and that of Pfordresher and Mantell (2012) is the magnitude of disruption to sequencing due to serially shifted feedback. Whereas normal and asynchronous feedback led to comparable rates of production errors across experiments, there was a 127% change in errors during serially shifted feedback in Pfordresher and Mantell whereas we observed only a 67% change in the current data. The most likely reason we observed this difference is due to the fact that the current research required participants to only perform within the vocal effector system.

Despite the consistencies across singing tasks of Experiment 1 and Pfordresher and Mantell (2012), we found that the sequencing/timing dissociation did not fully generalize to the production of syllable sequences. The failure of this generalization was most apparent in the mean IOI analysis, for which both asynchronous and serially shifted feedback led to significant slowing of timing. For error rates, both tasks led to a pattern of results that matches the sequencing/timing dissociation, though there was evidence that this dissociation may have been slightly less pronounced for the speaking task.

If, as suggested by Experiment 1, the complexity of speech timing diminishes the sequencing/timing dissociation effect (particularly with respect to its effect on timing), an important question emerges that pertains to the integration of speech and song. For instance, when one sings a song with lyrics, as is typically the case in common practice, does the dissociation effect diminish as seen in the speech data of Experiment 1? Or does the prominent rhythmic structure of song, which has been shown to promote entrainment more so than speech timing (Dalla Bella, Białuńska, & Sowiński, 2013) cause the dissociation to return, even when analyzing production characteristics that are specific to speech (e.g., speech errors as opposed to pitch errors)? We addressed these possibilities in Experiment 2.

Experiment 2: Sung Syllable Sequences

Experiment 1 demonstrated some similarities in the effect of AAF on speech versus song, as well as some differences. Whereas the qualitative effects of AAF on mean IOI versus error rates were similar across domains, the robustness of the dissociation varied subtly. We next examined the effect of combining the two production tasks, such that participants sung nonsense syllables. By including phonetic content into the melodic sequence structure, we were interested in how the pattern of results would change with respect to the sequencing/timing dissociation hypothesis.

Despite the fact that both production tasks in Experiment 1 were set to isochronous production rates, it has been found that the regular temporal structure of music facilitates coupling between action and sound (Dalla Bella et al., 2013). Thus, one hypothesis is that making the syllable sequences more "song-like" will lead to a pattern of results that better fit the predictions outlined by the dissociation hypothesis. Moreover, previous research has demonstrated that sung speech is processed in an integrated, holistic fashion and results in processing advantages that facilitate performance when domain information is structurally redundant (Kolinsky, Lidji, Peretz, Besson, & Morais, 2009; Schön et al., 2008). Following this, it would be further predicted that domain differences should be reduced.

On the other hand, modular accounts (cf. Peretz & Coltheart, 2003) draw clear divisions between speech and music processing. Where, pitch information is processed by a dedicated module for tonal encoding, speech information (syllables, phonemes) is processed by its own domain-specific module. AAF that affects both pitch and speech information simultaneously will disrupt each respective module independently. As a result, it would be predicted that these two processing systems will need to coordinate with one another thereby leading to potentially additive AAF disruptive effects beyond the levels observed in the first experiment's separate tasks.

Method

Participants. Fifteen undergraduate students were recruited from the University of Buffalo introductory psychology subject pool and their mean age was 22.5 years (range 18–32 years). Nine participants were women and six were men. No participants reported hearing problems and all but one participant was right hand dominant.

As in in the first experiment, participants were sampled without regard to musical training. Five participants reported no musical training or experience, six reported 1 to 4 years of training, and the remaining four were classified as musicians as they reported 5 or more years of training and experience. Three of these musicians also reported having absolute pitch. Only one of these musicians indicated that they had received extensive vocal training (8 years of private lessons) while the other three musicians reported receiving no vocal training.

Materials. The stimulus materials from Experiment 1 were essentially combined to create sung nonsense syllable sequences. This was accomplished by modifying the pitch of the nonsense syllable sequences to match the contour of its melodic counterpart (see Figure 1 for reference). Using Praat's pitch bending tools, instances of the "Loo" syllable's F_0 were reassigned to 261.262 Hz, the frequency of the C_4-note pitch class. This process of F_0 reassignment was then repeated for each of the other seven syllables in the sequence. After these pitch modifications were conducted, the sequence was then perceptually normalized with respect to loudness to control for any intensity contours that might have emerged due to the alterations in pitch.

Conditions, apparatus, and procedure. The conditions and apparatus were identical to those of Experiment 1. The procedure was highly similar as well, but combining the production tasks required slight modifications. Before the experimental task, participants were directed to engage in all three of the aforementioned warm-up exercises. Again, participants then proceeded to the learning phase where they had the opportunity to memorize the sung syllable sequence. Here, participants required slightly more learning trials than during both the singing and speech learning trials of Experiment 1 to sufficiently memorize the sung nonsense syllable sequences (M = 6.87, SD = 2.72). After the learning phase participants were given a single practice trial followed by 12 total experimental trials. Learning, practice, and experimental trials' structures were identical to Experiment 1. During experimental trials the three feedback conditions were each presented four times and participants experienced all three conditions before any one condition was repeated.

To examine the disruptive effects of AAF on production timing we used the same procedure as Experiment 1. However, we only computed a single-factor ANOVA on mean IOIs due to the concurrent sung syllable task. Likewise, the combined task of syllable and pitch production generates the opportunity to analyze errors from three perspectives: syllabic production errors, melodic contour errors, and simultaneously occurring errors.

Syllabic and melodic contour errors were defined as two separate features of a single produced event, but we analyzed both error types independently to test the sequencing/timing hypothesis. The error analysis procedures were the same described earlier in the first experiment and two-way ANOVAs were computed to examine the disruptive effects of AAF on production sequencing.

Results

Disruptive effects of AAF on timing. Figure 4A shows the effects of AAF on mean IOIs. Because speech and song were integrated in this experiment, we did not separate timing data across domains, as in Experiment 1. The trend in mean IOI across feedback conditions conforms to the predicted pattern, where asynchronous feedback led to the most slowing. However, the serially shifted feedback resulted in some slowing as well, albeit to a lesser degree. A one-way ANOVA on mean IOI verified the effect of feedback condition, $F(2, 28) = 9.19$, $p < .01$, $\eta^2 = .40$. Although the Tukey's HSD test ($\alpha = .05$) showed that asynchronous feedback (M = 642.1 ms, SD = 72.8) led to significant slowing, serially shifted feedback (M = 620.4 ms, SD = 40.6) did

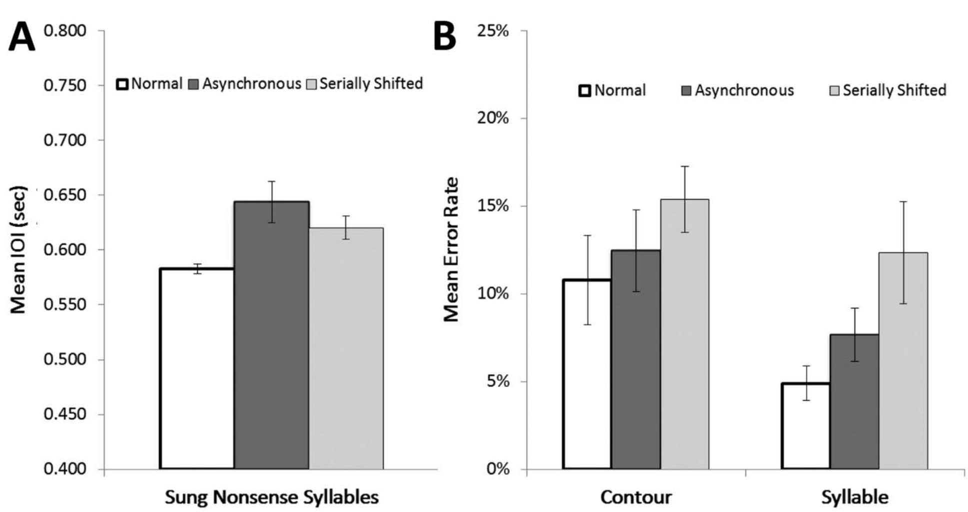

Figure 4. The effects of auditory feedback condition on the sung nonsense syllable task of Experiment 2. Graphs depict disruptive effects of production timing (A) and sequencing error type (B). Error bars represent standard errors. IOI = inter-onset-intervals.

as well. This pattern of mean differences from the normal feedback (M = 583.2 ms, SD = 15.8) condition was similar to the results of Experiment 1 for the speaking task, but not the singing task. Likewise, the regression analysis, R^2 = .18, $F(2, 177) = 18.86$, $p < .001$, suggested that both predictors accounted for a significant proportion of the variance in mean IOI, disruption from asynchronous AAF, β = .48, $t(177) = 6.09$, $p < .001$, disruption from serially shifted AAF, β = .30, $t(177) = 3.76$, $p < .001$; and the test for magnitude of dependent beta values was not significant, $t(176) = 1.23$, p = .11. Thus, as in the speech production task in Experiment 1, both AAF conditions slowed timing in Experiment 2.

Disruptive effects of AAF on sequencing. Figure 4B shows the mean percentage of syllabic and melodic contour errors resulting from the effects of AAF while singing nonsense syllable sequences. As a reminder, the data were analyzed by jointly categorizing each produced event in terms of its pitch and syllabic content, and then separately analyzing accuracy along each dimension. The trend in mean percentage of errors do conform to the expected pattern outlined by the dissociation hypothesis, where serially shifted feedback leads to more frequent production errors compared to normal auditory feedback. A two-way ANOVA on mean percentage of errors resulted in no significant main effect of task error type, $F(1, 14) = 2.47$, p = .14, η_p^2 = .14, and no significant interaction, $F(2, 28) = 2.48$, p = .10, η_p^2 = .15. However, there was a significant main effect of feedback condition, $F(2, 28) = 7.82$, p = .002, η_p^2 = .36. To further understand whether this main effect is representative of each error type considered separately, we computed a one-way ANOVA for each task error type.

As can be seen in Figure 4B, mean percentage of melodic contour errors differed across feedback conditions. This effect of feedback condition was verified with the one-way ANOVA, $F(2, 28) = 6.43$, $p < .01$, η_p^2 = .31. Tukey's HSD (α = .05) post hoc analyses confirmed that serially shifted feedback (M = 15.4%, SD = 7.3%) led to the significantly more errors compared to normal feedback conditions (M = 10.8%, SD = 9.9%), whereas asynchronous feedback did not significantly increase errors (M = 12.5%, SD = 9.0%). However, there was no significant difference in mean contour errors between asynchronous and serially shifted feedback. The regression analysis, R^2 = .04, $F(2, 177) = 3.95$, $p <$ .05, did support the predicted dissociation effect where the only predictor accounting for a significant independent proportion of the variance was based on effects coding of disruption from serially shifted AAF, disruption from serially shifted AAF, β = .24, $t(177) = 2.78$, $p < .01$, disruption from asynchronous AAF, β = .12, $t(177) = 1.01$, p = .31. Thus, with the exception of one post hoc contrast, melodic contour error data upheld the predicted pattern.

With respect to mean percentage of syllabic errors, serially shifted feedback also led to the highest degree of accuracy disruption. The one-way ANOVA confirmed the effect feedback condition, $F(2, 28) = 7.16$, $p < .01$, η^2 = .34, and the Tukey's HSD (α = .05) test showed that serially shifted feedback (M = 12.4%, SD = 11.3%) resulted in significantly more syllabic errors than normal feedback. Again, the mean percentage of errors in the asynchronous feedback (M = 7.7%, SD = 5.9%) conditions were not significantly different from normal (M = 4.9%, SD = 3.9%) conditions and the difference between asynchronous and serially shifted feedback was not significant (contrary to the hypothesis). Again, the regression analysis, R^2 = .07, $F(2, 177) = 6.74$, $p <$ .01, supported the predicted dissociation effect such that the only predictor accounting for a significant independent proportion of the variance was based on effects coding of disruption from serially shifted AAF, disruption from serially shifted AAF, β = .30, $t(177) = 3.63$, $p < .001$, disruption from asynchronous AAF, β = .11, $t(177) = 1.32$, p = .19. Taken together, errors in Experiment 2 followed very similar patterns whether based on pitch content (contour errors) or syllabic content.

We next address the degree to which the different error analyses reported earlier reflect independent features of produced events. Given that the error analyses reported earlier are based on joint features of events, it is possible that each feature (pitch content and syllabic content) reflect an integrated planning process and are thus not truly independent events. If so, it is not appropriate to treat these features as separate in our analyses. Thus, we compared the probability of joint errors across trials (events that are errors with respect to both pitch and syllabic content) with the joint probability of each error type described above. There was a probability of making a syllabic error, $p_{(\text{syllable})}$ = .083, and the probability of making a contour error, $p_{(\text{contour})}$ = .129 across the trials. The joint probability of these events, $p_{(\text{syllable})} \times p_{(\text{contour})}$ = .011, was very close to the probability of simultaneous errors, $p_{(\text{simultaneous})}$ = .010, and within one standard error of the mean for simultaneous errors, SE = .003. Thus we concluded it was appropriate to treat contour and syllable errors as separate.

Action-feedback overlap analyses. One aspect of our experimental manipulations that requires further inspection involves the means by which we created serially shifted feedback. We knew from the outset that there would be potential deviations of production timing, which may result in desynchronizations between feedback and actions. Because altered feedback was always delivered at a temporally precise interval, if participants deviated from the prescribed production rate their altered feedback would have also deviated. It was thus necessary to assess whether such inadvertent timing variations affected performance accuracy.

To examine these effects we transformed timing data into a percentage of overlap between feedback and action timing. This transformation was conducted by employing equation:

$$\%\ \text{Overlap} = [(\text{IOI}_m - \text{Delay})/\text{IOI}_m] \times 100.$$

Here, IOI_m equals the mean IOI for a given trial and *Delay* equals the length of the delay on that trial. Thus, this overlap reflects the temporal relationship between perception and actions based on the participant's own tempo, rather than the prescribed tempo. An overlap percentage of 100 correspond to conditions in which onset and the feedback of an action are synchronous, such as during normal feedback performance. Conversely, 0% of overlap corresponds to instances where the participant does not hear any auditory feedback with its associated action, which is predicted to be the case during serially shifted feedback (when the delay ought to equal the produced mean IOI).

Figure 5 shows the relationship between mean percentage of overlap and the percentage of errors committed during both speech and singing tasks across all three experiments. One important result from this analysis is that the functions across task domains within each experiment are highly similar to one another. This

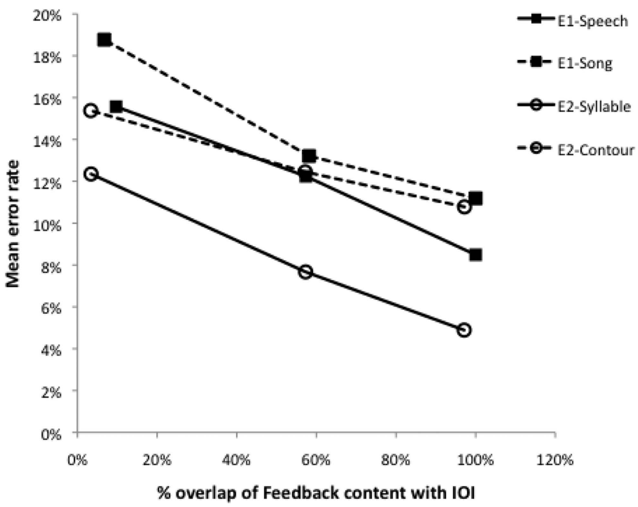

Figure 5. The relationship between percentage of action-feedback overlap and percentage of errors for both speech and singing tasks across the three experiments. E1 = Experiment 1; E2 = Experiment 2; IOI = inter-onset-intervals.

illustrates the similarities in feedback disruption across speech and singing tasks, where low percentage of overlap results in higher error rates and high percentage of overlap leads to lower error rates. Second, the relationship between overlap and error rates is fairly consistent across experiments. These plots verify that a similar influence of pitch content is found for production of speech and song, even given deviations in produced timing from the prescribed rate.

Discussion

The goal of Experiment 2 was to determine if the use of auditory feedback during combined speech/song tasks lead to results similar to either of the separate tasks in Experiment 1, reflecting domain-specific dominance, or integration. Overall, results suggest that any dominance effects come from the effect of temporal complexity associated with speech. Whereas patterns of error rates in Experiment 1 were consistent across tasks, effects of AAF differed in their influence on timing, with the dissociative effect not occurring for speech. Likewise, in Experiment 2, mean IOI data did not support a dissociative effect of AAF, whereas error data—whether analyzed with respect to pitch or syllable content—aligned with the dissociation hypothesis.

Beyond this, when one considers the magnitude of AAF effects overall in Experiment 2, relative to Experiment 1, it becomes clear that the results of Experiment 2 do not easily fit into a fully modular account. A modular account of music and language processing would assume that each cognitive system would need to coordinate with one another during a simultaneous production task (Peretz & Coltheart, 2003). Thus, disruptive effects of AAF would be exacerbated in Experiment 2 relative to Experiment 1. In contrast to this prediction, effect sizes associated with AAF were generally smaller in Experiment 2 than Experiment 1, as shown in Table 1. This reduction could reflect the possible reinforcement of compatible syllabic and pitch information during sequence learning (Schön et al., 2008, but see Larrouy-Maestri, Leybaert, & Kolinsky, 2013). However, there is one possibly significant exception in that error rates associated with syllable production were more vulnerable to AAF in Experiment 2 than in Experiment 1.

General Discussion

The primary purpose of this research concerns how action and perception are coordinated in different production domains. Our approach to this inquiry was to examine similarities in disruptive effects of AAF across speech and music production tasks. Participants sung melodies, spoke syllables, or sung syllable sequences during AAF manipulations of synchrony and content. We found that tasks which incorporate the production of sequences with phonetic variation (speaking or singing syllables) did not fully conform to the pattern of dissociative effects found in music production reported here and elsewhere (Pfordresher, 2003; Pfordresher & Mantell, 2012). In particular a double behavioral dissociation was found in music, where asynchronous feedback disrupts timing but not accuracy, while serially shifted feedback disrupts accuracy but spares timing. However, the present data only confirm a single dissociation in the domain of speech.

Whereas the accuracy of speech was only influenced by alterations to speech content (hearing the previously produced syllable), speech timing was disrupted by both AAF conditions. Thus, unlike music, the use of auditory feedback in the regulation of timing in speech may be sensitive to a broader range of hierarchical levels that are both sensitive to onset timing but also to sequential information in feedback. Why?

As discussed in the introduction, we think the present results for speech bear on the importance of temporal variability in the speech signal in determining content. Whereas “content” for melodies may be explicitly linked to spectral information that defines pitch, “content” in the present speech stimuli is more complex and involves both spectral and temporal information. Moreover, previous theories have posited that auditory neural regions have become specialized due to the demands placed by the domain-relevant structural characteristics of the acoustic signal (Zatorre & Belin, 2001; Zatorre et al., 2002). This is especially pertinent to speech where greater temporal resolution is required to extract the rapid spectra-temporal changes important for identifying phonetic units. In the context of our observations, it seems that action-perception representations are sensitive to timing relationships corresponding to this perceptual level.

One way to conceptualize how this sensitivity might emerge is presented in Figure 6, which is a hypothetical illustration of the relationship between produced syllables and AAF. Implementing

Table 1
Summary of Effect Sizes Associated With the Main Effect of Feedback

	M IOI			Error rates		
Experiment	Overall	Singing	Speaking	Overall	Singing	Speaking
E1	.56	.58	.47	.53	.43	.24
E2	.40	N/A	N/A	.36	.31	.34
E1 – E2	.16			.17	.12	–.10

Note. Effect sizes are η_p^2. IOI = inter-onset-intervals; E1 = Experiment 1; E2 = Experiment 2.

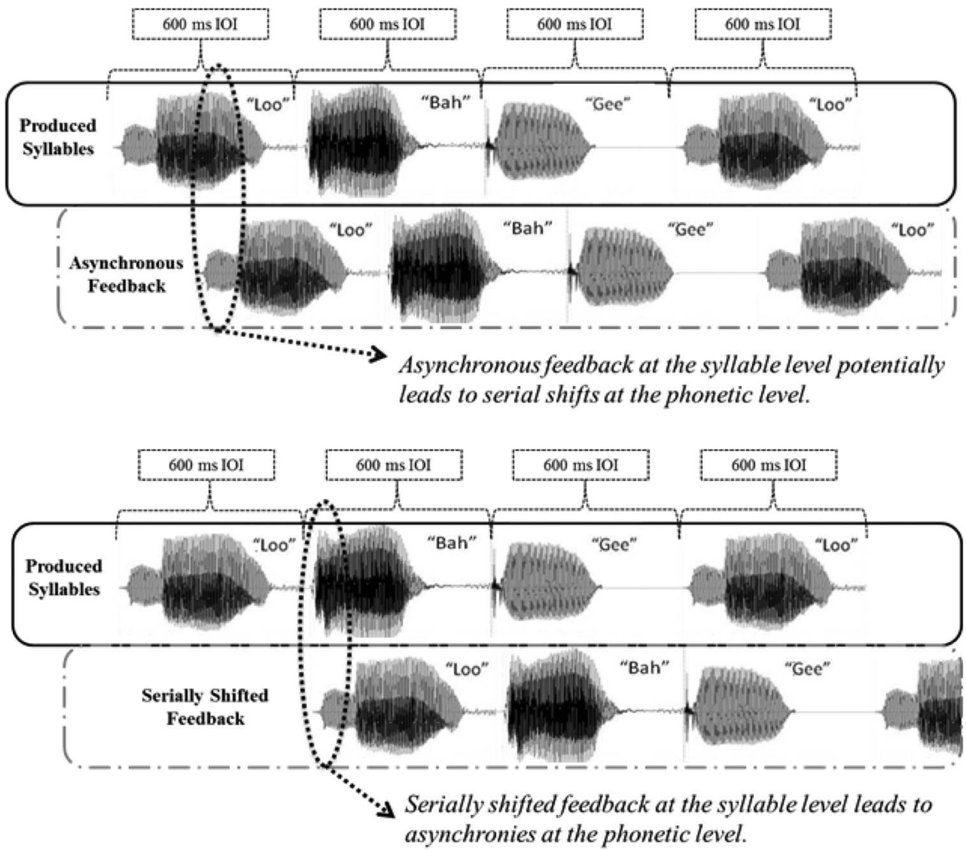

Figure 6. A hypothetical illustration of the relationship between produced syllables and altered auditory feedback. IOI = inter-onset-intervals.

asynchronous feedback at the superordinate syllable level may result in serial shifts of phonemes that compose said syllable. However, the results of the experiments reported here do not suggest that this scenario leads to significant performance disruption because we did not observe an elevation of production errors. On the other hand, serial shifts of syllabic information potentially create asynchronies at the phoneme level and such asynchronies do lead to the expected disruptive effects to performance timing.

An account for these observed timing disruptions, as a result of phonetic asynchronies, draws on assumptions outlined by Howell and colleagues' displaced rhythm hypothesis (DRH; Howell, 1983; Howell & Archer, 1984; Howell & Powell, 1987; Howell et al., 1983). Although DRH is oriented toward AAF's disruptive effects emerging from amplitude contour information at the syllabic level, the central assumptions could very well extend to phonetic amplitude variations as well. The potential difference between the amplitude profile of articulating a stop consonant (such as the /b/ in "Bah") while receiving feedback associated with a lateral consonant (such a /l/ in "Loo") may give rise to the observed timing disfluencies. Another example involves the asynchronies resulting from differences in voice onset timing, such as when articulating the /p/ in "Poh" but hearing the /b/ in "Bah," which would again result in a mismatch in amplitude contour between the produced utterance and feedback. However, ascertaining the validity of such an account is difficult given our manipulations and would require subsequent research with alteration to synchrony and content at the phonetic level.

Such observations cohere well with many cognitive models of speech production that account for sequencing at low-order representational level. For example, syllable-frame models assume that the production of speech sequences utilizes both syllabic and subsyllabic planning units that undergo a serial ordering process (Dell, 1986; Shattuck-Hufnagel, 1979). The onset, nucleus, and coda are independently represented constituents of the syllable unit. Evidence for this derives from speech error patterns, which commonly reflect an interaction among commensurate segments that occupy similar positions in the syllable. Moreover, the speech error data also suggests that these constituent units are further composed of phonological units, which are also subject to a serial ordering planning process (Dell, 1986; Fromkin, 1971; Garrett, 1982). Such potential misordering errors, at the syllabic, subsyllabic, and phonological level illustrate the gradated hierarchical structure of planning and motor execution in speech.

An additional motivation for this research involves suggestions by Howell and colleagues (e.g., Howell, 2001; Howell, 2004a, 2004b; Howell & Archer, 1984) concerning the role of feedback content. As mentioned in the introduction, various findings from the speech literature suggest that feedback content plays a negligible role in production (e.g., Howell, 2001; Howell, 2004a, 2004b; Howell & Archer, 1984). However, none of these papers incorporated manipulations similar to the serial shift manipulation used here. Rather, manipulations of content led to a sequence of auditory feedback events that were unrelated to the planned action sequence, except with respect to onset timing (e.g., Howell and Archer, 1984, converted the speech signal to an amplitude varying square-wave tone). Likewise, other studies that use similar delays of speech feedback during syllable production (Kaspar & Rübeling, 2011; Müller et al., 2000) did not control production timing enough to make a direct comparison to the present data. Thus, the fact that serial shifts of speech feedback are disruptive is a novel finding in its own right, and suggests—at least at the level of sequencing—more commonality in the use of feedback across domains than had been suggested in the past.

It is worth commenting on the roles of timing and sequencing in different tasks. The dissociation tested here is based on the assumption that timing and sequencing are both part of a structured event sequence, and that both time scales contribute to the information being communicated. Thus, when one of these aspects of production is perturbed, there is a disruption to the intended "message." This reflection bears on two important issues. First, the present results contribute to a broader literature having to do with speech timing. A longstanding debate has been found in the literature between characterization of speech timing as "event" based—invoking the construct of an internal timekeeper (Kidd, 1989)—versus the notion that speech timing is an "emergent" property of articulations (Fowler, 1977, 1980). Event-based timing models typically rest on the assumption that speech timing is based on syllables. Although our data do show a strong role of syllable timing (e.g., in the effects of serial shifts), they argue for some contribution of emergent temporal properties based on phonetics. A second issue concerning the role of timing relates to other Redundant research. Whereas we found no effects of alterations to content on timing when "timing" referred to the rate at which events are sequenced, other uses of time can show effects of such alterations. For example, Keller and Koch (2006, 2008) demonstrated that participants' expectations about the mapping of pitch content to action could influence the timing of short event sequences, including the time it takes to initiate the production of such sequences. More important, in their tasks, production timing included the time it takes to select one of four possible sequences, and thus was influenced by decision-making processes. A similar finding was recently reported by Pfordresher, Keller, Koch, Palmer, and Yildirim, 2011: When participants had to switch between one melody and another melody within a trial, temporal pauses associated with melody switches that participants had to suppress were longer when production of the initial melody was accompanied by feedback from the alternate melody (to which a participant may switch).

With respect to the organization of speech planning and production processes, a limitation of the experiments reported here concern the lack of control for temporal congruities at the phonetic level. We designed our stimuli and feedback manipulations so to treat the syllable unit in speech sequencing as an analogue to a single pitch in a melody. Nevertheless, we did observe a similarity across domains in that asynchronous feedback disrupts both speech and music timing, which is consistent with a domain general account. Thus, cognitive mechanisms that regulate production timing at this hierarchical level are likely similar across domains. However, speech inherently possesses lower order representations that guide planning and motor programs. Because serially shifted feedback manipulations at the syllable level perturbed timing at the phonetic level and the current experiments' pitch sequences do not have analogous structures at this smaller scale, further research designs should address this limitation to fully ascertain potential sequencing processes common across domains.

In conclusion, we found similar sensitivity in the ability to serially order syllables or pitches to manipulations of auditory feedback content. By contrast, we found that the timing of these sequences did not exhibit the same kind of sensitivity to different types of AAF. Whereas the timing of sung pitches on a single syllable exhibited sensitivity only to onset synchrony of AAF, the timing of spoken syllable sequences was also sensitive to variations in content. We see this difference arising from the greater complexity of temporal information in the speech signal, with potentially greater implications for how separable sequencing and timing are for speech planning in contrast to (nonverbal) music production. Thus, although the coordination of perception and action in both speech and music may be based on a hierarchy of time scales, the specific time scales involved may vary depending on the composition of event sequences from each domain.

References

Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: A review and a hypothesis. *Brain and Language, 89,* 320–328. http://dx.doi.org/10.1016/S0093-934X(03)00347-X

Black, J. W. (1951). The effect of delayed side-tone upon vocal rate and intensity. *The Journal of Speech and Hearing Disorders, 16,* 56–60. http://dx.doi.org/10.1044/jshd.1601.56

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer Version 5.3.83 [Computer program]. Retrieved August 16, 2014 from http://www.praat.org/

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.

Couchman, J. J., Beasley, R., & Pfordresher, P. Q. (2012). The experience of agency in sequence production with altered auditory feedback. *Consciousness and Cognition, 21,* 186–203. http://dx.doi.org/10.1016/j.concog.2011.10.007

Dalla Bella, S., Białuńska, A., & Sowiński, J. (2013). Why movement is captured by music, but less by speech: Role of temporal regularity. *PLoS ONE, 8,* e71945. http://dx.doi.org/10.1371/journal.pone.0071945

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321. http://dx.doi.org/10.1037/0033-295X.93.3.283

Fairbanks, G. (1955). Selective vocal effects of delayed auditory feedback. *The Journal of Speech and Hearing Disorders, 20,* 333–346. http://dx.doi.org/10.1044/jshd.2004.333

Fairbanks, G. (1960). The rainbow passage. In G. Fairbanks (Ed.), *Voice and articulation drill book* (2nd ed., pp. 124–139). New York, NY: Harper & Row.

Finney, S. A. (1997). Auditory feedback and musical keyboard performance. *Music Perception, 15,* 153–174. http://dx.doi.org/10.2307/40285747

Finney, S. A., & Warren, W. H. (2002). Delayed auditory feedback and rhythmic tapping: Evidence for a critical interval shift. *Perception & Psychophysics, 64,* 896–908. http://dx.doi.org/10.3758/BF03196794

Fowler, C. A. (1977). *Timing control in speech production*. Bloomington: Indiana University Linguistics Club.

Fowler, C. A. (1980). Co-articulation and theories of extrinsic timing. *Journal of Phonetics, 8,* 113–133.

Fromkin, V. (1971). The non-anomalous nature of anomalous utterances. *Language. 47,* 27–52.

Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 19–76). London, England: Academic Press.

Gates, A., & Bradshaw, J. L. (1974). Effects of auditory feedback on a musical performance task. *Perception & Psychophysics, 16,* 105–109. http://dx.doi.org/10.3758/BF03203260

Gates, A., Bradshaw, J. L., & Nettleton, N. C. (1974). Effect of different delayed auditory feedback intervals on a music performance task. *Perception & Psychophysics, 15,* 21–25. http://dx.doi.org/10.3758/BF03205822

Gold, B., Morgan, N., & Ellis, D. (2011). Pitch detection. In B. Gold, N. Morgan, & D. Ellis (Eds.), *Speech and audio signal processing: Processing and perception of speech and music* (pp. 455–472). Hoboken, NJ: Wiley. http://dx.doi.org/10.1002/9781118142882

Havlicek, L. L. (1968). Effect of delayed auditory feedback on musical performance. *Journal of Research in Music Education, 16,* 308–318. http://www.jstor.org/stable/3344070

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24,* 849–878. http://dx.doi.org/10.1017/S0140525X01000103

Hou, Y. C. (2014). Speech jammer (Version 1.4) [Mobile application software]. Retrieved from https://play.google.com/store/apps/details?id=com.icechen1.speechjammer

Howell, P. (1983). The effect of delaying auditory feedback of selected components of the speech signal. *Perception & Psychophysics, 34,* 387–396. http://dx.doi.org/10.3758/BF03203052

Howell, P. (2001). A model of timing interference to speech control in normal and altered listening conditions applied to the treatment of stuttering. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstjin (Eds.), *Speech motor control in normal and disordered speech* (pp. 291–294). Nijmegen, The Netherlands: Uitgeverij Vantilt

Howell, P. (2004a). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communicative Sciences and Disorders, 31,* 122–139.

Howell, P. (2004b). Effects of delayed auditory feedback and frequency-shifted feedback on speech control and some potentials for future development of prosthetic aids for stammering. *Stammering Research, 1,* 31–46.

Howell, P. (2007). The effects of gated speech on the fluency of speakers who stutter. *Folia Phoniatrica et Logopaedica, 59,* 250–255. http://dx.doi.org/10.1159/000104463

Howell, P., & Archer, A. (1984). Susceptibility to the effects of delayed auditory feedback. *Perception & Psychophysics, 36,* 296–302. http://dx.doi.org/10.3758/BF03206371

Howell, P., & Powell, D. J. (1987). Delayed auditory feedback with delayed sounds varying in duration. *Perception & Psychophysics, 42,* 166–172. http://dx.doi.org/10.3758/BF03210505

Howell, P., Powell, D. J., & Khan, I. (1983). Amplitude contour of the delayed signal and interference in delayed auditory feedback tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 772–784. http://dx.doi.org/10.1037/0096-1523.9.5.772

Howell, P., & Sackin, S. (2002). Timing interference to speech in altered listening conditions. *The Journal of the Acoustical Society of America, 111,* 2842–2852. http://dx.doi.org/10.1121/1.1474444

Kalmus, H., Denes, P., & Fry, D. B. (1955). Effect of delayed acoustic feed-back on some non-vocal activities. *Nature, 175,* 1078. http://dx.doi.org/10.1038/1751078a0

Kaspar, K., & Rübeling, H. (2011). Rhythmic versus phonemic interference in delayed auditory feedback. *Journal of Speech, Language, and Hearing Research, 54,* 932–943. http://dx.doi.org/10.1044/1092-4388(2010/10-0109)

Keller, P. E., & Koch, I. (2006). The planning and execution of short auditory sequences. *Psychonomic Bulletin & Review, 13,* 711–716. http://dx.doi.org/10.3758/BF03193985

Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 61,* 275–291. http://dx.doi.org/10.1080/17470210601160864

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 736–748. http://dx.doi.org/10.1037/0096-1523.15.4.736

Kolinsky, R., Lidji, P., Peretz, I., Besson, M., & Morais, J. (2009). Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition, 112,* 1–20. http://dx.doi.org/10.1016/j.cognition.2009.02.014

Krampe, R. T., Mayr, U., & Kliegl, R. (2005). Timing, sequencing, and executive control in repetitive movement production. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 379–397. http://dx.doi.org/10.1037/0096-1523.31.3.379

Kurihara, K., & Tsukada, K. (2012). *SpeechJammer: A system utilizing artificial speech disturbance with delayed auditory feedback*. Available at http://arxiv.org/ftp/arxiv/papers/1202/1202.6106.pdf

Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behavior Research Methods, Instruments, & Computers, 25,* 238–241. http://dx.doi.org/10.3758/BF03204504

Larrouy-Maestri, P., Leybaert, J., & Kolinsky, R. (2013). The benefit of musical and linguistic expertise on language acquisition in sung material. *Musicae Scientiae, 17,* 217–228. http://dx.doi.org/10.1177/1029864912473470

Lee, B. S. (1950). Effects of delayed speech feedback. *The Journal of the Acoustical Society of America, 22,* 824–826. http://dx.doi.org/10.1121/1.1906696

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38. http://dx.doi.org/10.1017/S0140525X99001776

MacKay, D. G. (1968). Metamorphosis of a critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *The Journal of the Acoustical Society of America, 43,* 811–821. http://dx.doi.org/10.1121/1.1910900

MacKay, D. G. (1987). *The organization of perception and action*. New York, NY: Springer-Verlag. http://dx.doi.org/10.1007/978-1-4612-4754-8

Müller, K., Aschersleben, G., Esser, K. H., & Müsseler, J. (2000). Effects of delayed auditory feedback on speech: Just a problem of displaced rhythm. In P. Desain & L. Windsor (Eds.), *Rhythm perception and production* (pp. 249–261). Lisse, The Netherlands: Swets & Zeitlinger.

Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 457–470. http://dx.doi.org/10.1037/0278-7393.19.2.457

Palmer, C., & van de Sande, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 947–962. http://dx.doi.org/10.1037/0096-1523.21.5.947

Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6,* 688–691. http://dx.doi.org/10.1038/nn1083

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology, 56,* 89–114. http://dx.doi.org/10.1146/annurev.psych.56.091103.070225

Pfordresher, P. Q. (2003). Auditory feedback in music performance: Evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 949–964. http://dx.doi.org/10.1037/0096-1523.29.5.949

Pfordresher, P. Q. (2005). Auditory feedback in music performance: The role of melodic structure and musical skill. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1331–1345. http://dx.doi.org/10.1037/0096-1523.31.6.1331

Pfordresher, P. Q. (2006). Coordination of perception and action in music performance. *Advances in Cognitive Psychology, 2,* 183–198. http://dx.doi.org/10.2478/v10053-008-0054-8

Pfordresher, P. Q. (2008). Auditory feedback in music performance: The role of transition-based similarity. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 708–725. http://dx.doi.org/10.1037/0096-1523.34.3.708

Pfordresher, P. Q. (2012). Musical training and the role of auditory feedback during performance. *Annals of the New York Academy of Sciences, 1252,* 171–178. http://dx.doi.org/10.1111/j.1749-6632.2011.06408.x

Pfordresher, P. Q., & Brown, S. (2007). Poor-pitch singing in the absence of "tone deafness." *Music Perception: An Interdisciplinary Journal, 25,* 95–115. http://dx.doi.org/10.1525/mp.2007.25.2.95

Pfordresher, P. Q., Keller, P. E., Koch, I., Palmer, C., & Yildirim, E. (2011). Activation of learned action sequences by auditory feedback. *Psychonomic Bulletin & Review, 18,* 544–549. http://dx.doi.org/10.3758/s13423-011-0077-x

Pfordresher, P. Q., & Kulpa, J. D. (2011). The dynamics of disruption from altered auditory feedback: Further evidence for a dissociation of sequencing and timing. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 949–967. http://dx.doi.org/10.1037/a0021435

Pfordresher, P. Q., & Mantell, J. T. (2012). Effects of altered auditory feedback across effector systems: Production of melodies by keyboard and singing. *Acta Psychologica, 139,* 166–177. http://dx.doi.org/10.1016/j.actpsy.2011.10.009

Pfordresher, P. Q., Mantell, J. T., Brown, S., Zivadinov, R., & Cox, J. L. (2014). Brain responses to altered auditory feedback during musical keyboard production: An fMRI study. *Brain Research, 1556,* 28–37. http://dx.doi.org/10.1016/j.brainres.2014.02.004

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., & Grodd, W. (2000). Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. *NeuroReport, 11,* 1997–2000. http://dx.doi.org/10.1097/00001756-200006260-00038

Rosenbaum, D. A. (2010). Psychological foundations. In *Human motor control* (pp. 93–134). Waltham, MA: Academic Press. http://dx.doi.org/10.1016/B978-0-12-374226-1.00004-8

Ruhm, H. B., & Cooper, W. A., Jr. (1963). Some factors that influence pure-tone delayed auditory feedback. *Journal of Speech and Hearing Research, 6,* 223–237. http://dx.doi.org/10.1044/jshr.0603.223

Saito, Y., Ishii, K., Yagi, K., Tatsumi, I. F., & Mizusawa, H. (2006). Cerebral networks for spontaneous and synchronized singing and speaking. *NeuroReport, 17,* 1893–1897. http://dx.doi.org/10.1097/WNR.0b013e328011519c

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition, 106,* 975–983. http://dx.doi.org/10.1016/j.cognition.2007.03.005

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Somerset, NJ: Wiley.

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin, 136,* 943–974. http://dx.doi.org/10.1037/a0020541

Wong, P. C., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *The Journal of Neuroscience, 24,* 9153–9160. http://dx.doi.org/10.1523/JNEUROSCI.2225-04.2004

Yates, A. J. (1963). Delayed auditory feedback. *Psychological Bulletin, 60,* 213–232. http://dx.doi.org/10.1037/h0044155

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946–953. http://dx.doi.org/10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6,* 37–46. http://dx.doi.org/10.1016/S1364-6613(00)01816-7

Received December 6, 2013
Revision received September 26, 2014
Accepted September 29, 2014 ■