DIFFERENCES IN USE OF AUDITORY FEEDBACK DO NOT ACCOUNT FOR AGE-RELATED CHANGES IN CHILDREN'S SINGING

By

Sara Lynn Beck

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Psychology

August, 2014

Nashville, Tennessee

Approved:

Professor John Rieser

Professor Bruce McCandliss

ACKNOWLEDGEMENTS

I would like to thank my advisor, John Rieser, for his tremendous ongoing support, as well as lab members Aysu Erdemir, Emelyne Bingham, David Johnson, Lionel Pomeranz, Jessica Siegel, Blake Tamez, and Brooke Harwell for their contributions to this work. I would also like to thank Bruce McCandliss for offering invaluable guidance as a member of my committee. Special thanks to my husband, Park Chisolm, and our families.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure | Page

INTRODUCTION

Children's ability to sing accurately increases with age such that older children show significantly better pitch matching abilities than younger children (Geringer, 1983), better pitch discrimination abilities (Bentley, 1969), and better interval accuracy and key stability when singing a melody from memory (Hornbach & Taggart, 2005). Theorists in music education have long maintained that learning to monitor one's auditory feedback is a key component of age-related singing improvement, but some have also suggested that monitoring one's own proprio-kinesthetic feedback is essential as well. In the current study, we investigated the extent to which the ability to use auditory and proprio-kinesthetic feedback to regulate singing accuracy develops with age. We asked children and adults to sing a familiar song under normal auditory feedback conditions and under masked auditory feedback conditions. In the masked condition, participants' singing performance indicated their ability to utilize proprio-kinesthetic feedback, since auditory feedback was blocked. Based on past research showing that self-regulatory feedback use in the auditory domain decreases with maturation, we predicted that younger children would be more disrupted by masked auditory feedback than older children and adults, demonstrating a higher reliance on auditory feedback and less skilled use of proprio-kinesthetic information. If this were confirmed, it would indicate that development of self-regulatory use of proprio-kinesthetic feedback was implicated in the development of singing skill that occurs reliably with age. If not, it would suggest that other areas of development – namely, improved vocal-motor control and improved auditory representations of target melodies – account for age-related improvements in singing.

Self-Monitoring: Auditory and Proprio-Kinesthetic Feedback

An individual's ability to monitor the sound of his or her own voice while singing is an essential skill if one wishes to sing accurately, and it is reasonable to ask how and when this skill develops in the course of normative development. Consider the process of learning to perform skilled actions other than singing. When one learns to throw a ball, for example, he or she must monitor where the ball lands visually and also notice how it feels to throw. Eventually, individuals learn how it feels to release the ball prematurely and begin to associate that actual result with the feeling of having let go too soon. Essentially, they are learning to compare efference, or their intentions and motor plans, with the afferent outcome of their actions. Similarly, when singing, one must learn to monitor his or her auditory feedback in real time and use that feedback to make adjustments to vocal production. Ultimately, one must learn to listen to oneself while performing the skilled action of singing and make fine vocal-motor adjustments in real time.

Additionally, we must learn to read cues from our bodies that correspond to desired and undesired results. These cues might indicate movement of air across the vocal folds or the positioning of the lips and soft palate. In the current study, the term proprio-kinesthetic feedback was used to capture both proprioceptive and kinesthetic cues. The difference between paying attention to bodily cues when singing as opposed to throwing a ball, however, is that proprio-kinesthetic cues when singing are arguably more subtle. It is possible that learning to attend to those cues develops normatively with age, and it is this hypothesis that is addressed in the current study. Alternatively, it is conceivable that using these cues effectively is a skill that emerges only with singing training or other experience that is outside the scope of normative development. The current study follows the logic that when auditory feedback is completely blocked, the degree of disruption to a participant's baseline singing accuracy reflects two things: the relative degree of reliance on auditory feedback (less disruption indicating less reliance) and the ability of the participant to utilize proprio-kinesthetic feedback (less disruption indicating greater use of bodily cues). The current study examined singing performance with and without auditory feedback among participants in three age groups: 5 to 8 years old, 9 to12 years old, and adults.

Research on Children's Use of Auditory Feedback in Speech Development

To our knowledge, no published studies have investigated children's accuracy when singing songs under masked auditory feedback conditions, but two bodies of research informed our predictions and methodology in the current study: self-regulatory use of auditory feedback in children's speech development and self-regulatory use of auditory feedback in adult singing. Early research in children's use of auditory feedback in speech development relied upon the logic that the more one utilizes auditory feedback during speech, the more one will experience disruption under altered auditory feedback conditions (Siegel, Fehst, Garber, & Pick, 1980). Specifically, several teams of researchers investigated children's use of auditory feedback at different ages by using a delayed auditory feedback (DAF) paradigm and examining resulting speech disfluencies. DAF has been shown to disrupt the speech of adults (Lee, 1951), causing slowed speech and stuttering. Several developmental studies have suggested that younger children are significantly more affected by DAF when speaking than older children and adults. This pattern emerged in work by MacKay (1968), Ratner, Gawronski, & Rice (1964), and Siegel, et al. (1980). These studies suggest that self-regulatory feedback use in the auditory domain decreases with maturation, and that older children and adults are not relying on auditory feedback as much as younger children. These studies likewise imply that one's ability to utilize proprio-kinesthetic feedback when speaking increases with maturation. Siegel and colleagues point out that in the context of speech – particularly with regard to the temporal elements that are disrupted by DAF – these findings make intuitive sense, since some elements of speech production likely become automatized with experience and increasing mastery of language (Siegel et al. 1980, p. 810). Therefore, if the same pattern held true for singing in the current study, one would expect that the youngest children would experience the greatest disruption to singing accuracy from masked auditory feedback, followed by older children and then adults.

Research on Adult Singing Without Auditory Feedback

Research on adults singing under masked auditory feedback conditions shows that pitch accuracy decreases to some degree when participants cannot hear themselves sing (Elliot & Niemoeller, 1970; Erdemir & Rieser, in review; Mürbe, Pabst, Hofmann, and Sundberg, 2002; Ward & Burns, 1978). This is true when participants are trained or untrained singers, and whether they are matching target tones (Elliot & Niemoeller, 1970), singing scales (Mürbe et al., 2002; Ward & Burns, 1978) or familiar songs from memory (Erdemir & Rieser, in review). Therefore, in addition to predicting age-related improvement in measures of singing accuracy, we expected that both children and adults in the current study would all experience disruption to baseline pitch accuracy under masked feedback conditions.

Other findings from this body of work strengthened the prediction of a developmental trend such that younger children were more reliant on auditory feedback than older children and adults; namely, two studies (Erdemir & Rieser, in review; Mürbe, et al., 2004) found that musical training interacted with feedback condition such that trained singers experienced less disruption under masked singing conditions than untrained singers. Erdemir and Rieser asked participants who were either trained singers, trained instrumentalists, or nonmusicians, to sing “Happy Birthday” under normal and masked feedback conditions and found that trained singers experienced significantly less disruption to pitch when they couldn’t hear themselves than the other two groups. Disruption to pitch was measured in terms of average size of errors in cents and the variability of the errors. The authors suggested that as a result of vocal training, trained singers had increased their ability to utilize proprio-kinesthetic feedback when singing, or possibly decreased their reliance on auditory feedback. Similarly, Mürbe and colleagues (2004) tested university vocal students longitudinally and found that after three years of college-level singing education, participants’ ability to sing accurately when auditory feedback was masked improved significantly on specific types of melodic exercises. They concluded that the kinesthetic feedback circuit was improved after vocal training. Although neither study looked at age-related experience, both suggest that a particular kind of experience – singing training – may affect an individual’s reliance on auditory feedback and ability to utilize proprio-kinesthetic feedback to control pitch when singing. This is suggestive of the type of age by feedback interaction predicted in the current study, simply because older children as compared to younger children– like trained singers as compared to untrained singers – presumably have more experience singing.

An unanswered question, however, is whether experience alone is sufficient to decrease one's reliance on auditory feedback to regulate pitch accuracy, or if training itself and a subsequent increase in ability to utilize proprio-kinesthetic feedback accounts for trained singers' decreased reliance on auditory feedback to control pitch. In the work on children's speech development, it appears that experience alone results in the automatization of certain temporal elements of speech. It is possible that the elements of singing performance likely to be disrupted by masked auditory feedback – namely pitch accuracy and error variability – do not become automatized with maturation and experience in the same way as the temporal regularities of speech. If that is the case, one would expect that children's reliance on auditory feedback when singing – as measured by their level of disruption when auditory feedback is masked – would not vary based on age.

Current Study

When we began the current study, we had three hypotheses. First, we predicted that the youngest children (ages 5 to 8) would be the least accurate singers with the greatest error variability, followed by older children (ages 9 to 12) and then adults. Second, we predicted that all three age groups would be more accurate under normal auditory feedback conditions than under masked feedback conditions. Finally, we predicted an age by feedback condition interaction such that younger children were more disrupted by lack of auditory feedback than older children and adults. Such a finding would be consistent with past research on speech development showing that reliance on auditory feedback for vocal production decreases with maturity. It would also indicate that an increasing ability to effectively utilize proprio-kinesthetic feedback is a contributing factor in improved singing across development.

METHOD

Participants

Fourteen children ages five to eight (9 girls, 5 boys, M_{age}=6.93), fourteen children ages nine to twelve (9 girls, 5 boys, M_{age}=10.21), and fourteen adults (8 women, 6 men, M_{age}=19.86) participated in the current study. English-speaking participants with no diagnosed hearing loss or developmental delays were recruited through a public website associated with The Vanderbilt Kennedy Center called StudyFinder, as well as through fliers distributed within the Nashville community. Some children were recruited through fliers distributed to families involved in the Nashville Children's Choir. All children who participated received a small toy to thank them for their participation. Adult participants were Vanderbilt undergraduates participating for course credit, except for two who were graduate students. One participant (a seven-year-old child) did not complete the study due to shyness about singing alone.

Participants' parents (and adult participants themselves) were asked to complete a brief questionnaire describing any relevant experience with private music lessons and choral singing. Seventeen of the 42 participants had experience with private music instruction (4 children ages 5-8, 7 children ages 9-12, 6 adults) and 24 had sung in a choir (8 children ages 5-8, 9 children ages 9-12, 7 adults). Additionally, the questionnaire asked for information on diagnosed hearing loss or developmental delays, which had been established as excluding criteria due to our interest in normative age-related change. No participants were excluded after completing the tasks based on these factors.

Materials and Procedure

Participants were tested individually by the author in a quiet room. Participants wore Bose noise-canceling headphones and sang into a Blue Snowflake portable USB microphone, placed 12 horizontal inches from the participant's body on a table to ensure consistent within-subject input levels. All singing was recorded into GarageBand on a MacBook Pro laptop computer, and both participant and experimenter heard all unmasked singing amplified, but unaltered through the headphones. The experimenter also wore noise-canceling headphones in order to monitor exactly what the participant was hearing at any given time. Participants' voices were amplified even under normal feedback conditions in order to match conditions necessary for use of the delayed auditory feedback paradigm for finding individual masking levels, as detailed below.

Delayed Auditory Feedback Paradigm

One of the challenges of masking children's auditory feedback was verifying that children could not, in fact, hear themselves. Most masking studies with adults and children have either used a pre-determined or constant masking level (Mürbe et al., 2002; Siegel, Pick, Olsen, & Sawin, 1976; Schultz-Coulon, 1978; Watts, Murphy, & Barnes-

Burroughs, 2003) or asked participants to self-report the point at which they are unable to hear themselves (Erdemir & Rieser, in review). Furthermore, the masking stimulus generally consists of broadband white noise (Mürbe et al., 2002; Schultz-Coulon, 1978; Siegel et al., 1976; Watts et al., 2003). Our method differed in two important ways: the first is the use of a multi-talker babble mask rather than a white noise mask, and the second is use of a delayed auditory feedback (DAF) paradigm to establish and verify an individualized masking threshold for each participant.

Erdemir and Rieser (in review) reported that in pilot testing, they found that a multi-talker babble mask – which sounds like a crowd of people chatting in a cafeteria – more effectively masked participants' ability to hear themselves when singing than white noise. Grillo, Abbott, and Lee (2010) looked at the impact of auditory masking on laryngeal pressure, and they also report an advantage to using speech noise rather than white noise to mask self-singing. During piloting of the current experiment, adults reported that the multi-talker babble mask effectively masked self-singing while white noise did not. Additionally, the pitch range in multi-talker speech is broader than that of a white noise mask (which is often low-pass filtered) and more closely mirrors the range of singing, making it preferable to white noise for masking purposes.

The second way in which the current auditory masking procedure differed from past research is that we were able to develop a paradigm to verify that participants could not, in fact, hear themselves when the masking stimulus was in use. As previously mentioned, DAF is known to create temporal speech disfluencies like hesitation and stuttering (Lee, 1951; Siegel et al., 1980). Children began by singing *The Alphabet Song* a cappella so that the experimenter was able to establish a baseline singing performance for a particular subject. During this initial performance, participants heard their own voices, amplified but unaltered in the headphones. The amplification was necessary so that when DAF was introduced (requiring amplification and digital manipulation of the singer's voice), the delay itself was the only new element. Next, a .63 second delay was applied to the participant's amplified voice using a GarageBand plug-in called AU Delay. Immediate, obvious disfluencies were observed in almost all cases; participants stuttered and hesitated, and many children were unable to get through the song due to giggling, starting and stopping. In the four cases where disfluencies were not immediately evident, the experimenter adjusted the delay (making it longer and louder) until the participant's singing was obviously compromised. Having established that performance under DAF created disfluencies, participants were asked to sing *The Alphabet Song* with DAF while the experimenter slowly increased the volume of the masking stimulus, termed the "babble mask." Our method for verifying the level at which subjects could no longer hear themselves was three-fold. First, we asked participants to indicate with a gesture (a horizontal hand motion signaling "enough") that they could no longer hear themselves. Second, the experimenter continued to increase the volume of the mask until it could be verified that the disfluencies created by the delay had disappeared. The disappearance of the DAF-created temporal disfluencies established that the participant was no longer able to hear his or her amplified voice through the headphones, which we can presume also masks the quieter, bone-conducted sound of self-singing. Third, the experimenter, seated next to the participant and approximately equidistant from the microphone, asked, "can you hear me?" after which no response signaled sufficient volume of the masking stimulus. The experimenter continued increasing the volume until these three criteria had all been met. The level of the masking stimulus at this point (which ranged from 72 to 81 decibels) was then utilized for the actual test trials, two of which were masked. Using the DAF paradigm enabled us to establish an individualized, effective masking level for each participant and verify its effectiveness prior to using it in test trials.

Procedure

Each participant began by completing consent documentation (and assent documentation in the case of children ages five to twelve). Each participant (or a parent of a participant) then completed a brief questionnaire describing experience with private music lessons, choral singing experience, and any hearing loss or diagnosed developmental delays. Next, the participant was seated next to the experimenter at a desk, in front of a stand-alone microphone. The experimenter explained that headphones would be worn throughout the task, but that the participant should feel free to ask any questions throughout and/or stop if he or she decided not to continue. Parents were allowed to remain in the room if it made the child feel more comfortable, but none elected to do so.

As detailed in Table 1, the experiment began with the experimenter and participant singing *The Alphabet Song* in unison to break the ice and also establish a precedent for continuing on through the end of the song, as some children were inclined to stop singing after the letter z. The experimenter let the participant begin, however, to allow the participant to select a comfortable key. Next, the participant sang *The Alphabet Song* once on his or her own so that the experimenter could establish a baseline performance before introducing the DAF. Next, the participant sang *The Alphabet Song* with DAF so that the experimenter could verify temporal disfluencies. As previously mentioned, in cases where the DAF was not immediately disruptive to the participant's singing, the experimenter adjusted it until disfluencies were evident. Next the participant sang *The Alphabet Song* with DAF while the experimenter increased the level of the babble mask. At this point the experimenter used the three converging measures detailed previously to verify that the mask was obscuring the participant's ability to hear his or her own voice. Before moving on, the experimenter tested the loudness of the individualized babble mask level to determine that it did not exceed 85 dB, the level that experimenters had established as a conservative maximum safe level for children participating in the study. All masking levels for all participants fell below the threshold. Next, the participant sang *The Alphabet Song* four more times alternating between normal auditory feedback and babble mask conditions, with half of the subjects beginning with normal auditory feedback and half beginning with babble mask. All nine performances of *The Alphabet Song* by each participant were recorded on individual tracks in GarageBand.

Following the singing, each participant was asked questions about his or her experience singing under altered auditory feedback. Participants' responses to these questions and their relations to the types of errors made will not be reported here.

Analysis

The fundamental frequency (F0) of each sung note in *The Alphabet Song* was extracted using Praat acoustic analysis software (Boersma & Weenik, 2008). In each sung performance, the experimenter identified the steady-state portion of each of the 26 letters of the alphabet for analysis, eliminating consonant onsets and stylistic sliding that occurred at the beginning of some notes. Consistent with recent work on adult singing (Dalla Bella, Giguère, & Peretz, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010, Ternström & Sundberg, 1988), our dependent measures of singing performance differentiated between pitch accuracy and precision, including separate measures of each. Pfordresher and colleagues (2010) define accuracy in singing as "the average difference between the pitch one sings and the actual target pitch" (p. 2182) whereas precision "relates to the consistency of the pitch one sings on repeated occasions, irrespective of whether any sung pitch meets its target (p.2183). They point out that although accuracy and precision are correlated, it is possible for a singer to be accurate and imprecise or precise yet inaccurate. Therefore, they were treated as separate constructs in the current study.

Since the primary focus of the current study was children's ability to control pitch under normal and masked feedback conditions, we chose not to measure intensity or perceived loudness, although intensity is a dependent variable in some developmental studies of children's speech under altered auditory feedback conditions (Siegel et al., 1980; Siegel et al., 1976). Another reason for excluding intensity as a dependent variable in the current study was that two factors in the experimental design may have impacted participants' intensity in opposite ways, potentially confounding interpretation of intensity data. Specifically, participants heard their voices amplified through headphones, a procedure sometimes called *sidetone amplification*, which has been shown to result in reduced intensity from speakers (Lane & Tranel, 1971, Siegel at al., 1976). We also used masking noise, which has been shown to increase intensity from speakers in a phenomenon called the *Lombard Effect* (Lombard, 1911, as cited in Lane & Tranel, 1971). Since neither sidetone amplification nor the Lombard Effect was the focus of this study, participants' intensity was not measured. In addition to measures of pitch accuracy and precision, consistency of tonal center and duration of sung performance were measured. The four dependent measures were extracted and computed as follows.

Mean Interval Error (MIE)

To assess each singer's accuracy, Mean Interval Error (MIE) –or the average size of absolute value errors – was calculated for each participant in each condition. Table 2 shows an example of these calculations for a single performance of *The Alphabet Song*. After identifying steady-state portions of each sung note as previously described, Praat was used to extract the median FO of each sung note in Hz. Each interval, or pair of adjacent sung notes, was then converted to cents, a logarithmic unit of measure based on twelve-tone equal temperament tuning in which 100 cents represents a semi-tone, or half-step. The formula used for calculating the distance between two adjacent sung pitches *a* and *b* (which are given in Hz) is $1200 * \log_2 (\frac{b}{a})$. Each resulting interval in cents was then compared against the target interval of a particular pair of notes, yielding a number reflecting the accuracy of a particular sung interval. These numbers were calculated using absolute values of differences between sung and target intervals, to avoid overshooting and undershooting errors canceling each other out. Across a single sung performance of the *The Alphabet Song* consisting of 26 sung letters, there were 25 interval errors calculated, and the MIE for a particular subject in a particular feedback condition (normal auditory feedback or babble mask) was calculated as an average of 50 interval errors across two sung performances. Lower values for MIE signify more accurate singing.

MIE, however, fails to capture the variability with which a singer approached each interval, or whether they consistently made small or large errors For example, a given subject might have a MIE in the normal auditory feedback condition of 57.2 cents, but that doesn't convey whether the participant consistently make fifty-cent errors or made smaller errors on average with a few vastly undershot or overshot intervals. To capture error variability, standard deviation of signed interval error (sdSIE) was utilized.

Standard Deviation of Signed Interval Error (sdSIE)

Standard deviation of signed interval error (sdSIE) was intended to capture the precision with which a participant sang, or the error variability of a particular singer. It answers the question, how variable are a participant's errors in each feedback condition? Table 3 shows an example of these calculations for a single performance of *The Alphabet Song*. As in calculating MIE, the extracted F0 values from Praat were used to calculate the magnitude of actual sung intervals in cents. Instead of using absolute value errors, however, sdSIE reflects the use of signed error terms since for each feedback condition, we were interested in a standard deviation of fifty numbers spanning the zero point (25 from each performance in each condition). Each sung interval was compared with the corresponding target interval to yield a signed interval error, and a standard deviation of all fifty interval errors in each feedback condition (normal auditory feedback or babble) was calculated for each participant.

To give an example which illustrates the complex relationship of accuracy and precision as reflected by these measures, imagine a participant who has a MIE of 57.2 cents in the normal auditory feedback condition with a sdSIE of 32 cents, while in the babble condition, he or she has a comparable MIE of 59 cents but a sdSIE of 87 cents. Although the participant's accuracy appears stable across feedback condition, his or her error variability has increased, giving us more information about the role of auditory feedback in controlling these different aspects of singing performance. In this hypothetical case, lack of auditory feedback is affecting the participant's precision, but not his or her pitch accuracy.

Standard Deviation of Tonic (sdT)

Another key component of singing performance that is not captured by either MIE or sdSIE is stability of tonal center, or the extent to which a singer is able to stay in the key in which he or she began singing. To measure this, Praat was used to extract F0 from three notes in each sung performance of *The Alphabet Song*: the letter "a," the letter "p," and the word "me" which ends the sung phrase, "next time won't you sing with me." These three notes represent the three appearances of the tonic, or tonal center, of the song. Regardless of a participant's chosen key, if he or she has a consistent tonal center, the frequency of these three notes will be nearly identical. If a participant is inconsistent, perhaps drifting up or down or making inconsistent errors in both directions, these three notes will show more variability. Thus, a standard deviation of these three notes in each performance of *The Alphabet Song* was calculated as a measure of stability of tonal center. A participant's sdT for a particular feedback condition (normal auditory feedback or babble mask) represents an average of two standard deviations, given in Hz. Smaller values indicate higher stability of tonic, a reflection of superior singing.

Duration

Duration of sung performance was utilized as a reflection of tempo. Duration was extracted in Praat, and an average duration was calculated for each participant in each condition.

RESULTS

We utilized two-way (Age X Feedback Condition) analysis of variance (ANOVA) with repeated measures applied separately to each dependent measure in turn. The age groups were 5 to 8 year-olds, 9 to 12 year-olds, and adults. The feedback conditions were normal auditory feedback (A) and babble mask (B). Criterion for significance was set at the .05 level unless otherwise indicated.

Mean Interval Error (MIE)

As shown in Figure 1, the analysis showed a significant effect of feedback condition on pitch accuracy (as measured by MIE) such that the average size of absolute value errors increased in the babble mask condition ($F=19.604$, $p<.01$). There was also a significant effect of age on MIE ($F=4.359$, $p<.01$) such that older children and adults made smaller errors on average than younger children. Post hoc testing with Bonferroni correction (adjusted $\alpha=.0125$) revealed that children ages 5 to 8 differed from children ages 9 to 12 in their MIE under both normal auditory feedback and babble mask conditions (normal auditory feedback: $t=2.445$, $p=.013$; babble mask: $t=1.717$, $p=.011$), but that children ages 9 to 12 did not differ significantly from adults in terms of MIE in either condition (normal auditory feedback: $t=-.26$, $p=.127$; babble mask: $t=.826$, $p=.719$). Therefore, it appears that 9 to 12 year-olds are demonstrating adult-like pitch accuracy, at least on this particular song. The interaction between age and feedback condition was non-significant, indicating that masking auditory feedback disrupted the pitch accuracy of participants across age groups equally.

Standard Deviation of Signed Interval Error (sdSIE)

As shown in Figure 2, the pattern of results for our measure of precision (sdSIE) was similar to the pattern of results for accuracy (MIE). There was a significant effect of feedback condition on sdSIE ($F=21.409$, $p<.01$), demonstrating that error variability increased overall in the babble condition as compared to the normal auditory feedback condition. There was also a significant effect of age on sdSIE, such that older children and adults showed less variable errors than younger children ($F=3.737$, $p=.033$). Similar to the pattern of results with MIE, post hoc testing with Bonferroni correction (adjusted $\alpha=.0125$) revealed that although 5 to 8 year-olds differed from 9 to 12 year-olds in normal feedback conditions ($t=2.150$, $p=.006$) and were nearly significantly different in the babble mask condition ($t=1.624$, $p=.019$), older children were not significantly less variable in their errors than adults (normal auditory feedback: $t=-.125$, $p=.252$; babble mask: $t=.802$, $p=.462$). The interaction between age and feedback condition on sdSIE was non-significant, indicating that error variability increased uniformly across all age groups when auditory feedback was masked.

Standard Deviation of Tonic (sdT)

As shown in Figure 3, there was a near-significant effect of feedback condition on sdT such that participants' tonal center was more stable in the normal auditory feedback condition than in the babble mask condition (F=3.815, p=.058). There was also a significant effect of age of sdT, showing that older children and adults stay "in key" better than younger children (F=5.165, p=.01). The interaction between age and feedback condition on sdT was non-significant, indicating that the ability of participants in all age groups to maintain a stable tonal center was equally disrupted by not being able to hear themselves sing (F=2.517, p=.094).

Duration

The effect of feedback condition on duration of sung performance was non-significant (F=.134, p=.716). Likewise, the effect of age on duration was non-significant (F=1.166, p=.322). The interaction between age and feedback condition on duration was also non-significant (F=.566, p=.572). Thus, the tempo at which participants sang *The Alphabet Song* did not vary systematically as a function of age or feedback condition.

DISCUSSION

We hypothesized that children's singing performance would improve with age, and we found a significant effect of age on singing accuracy (MIE), precision (sdSIE), and stability of tonal center (sdT) such that the youngest children were the least accurate, precise, and stable in their chosen key. Interestingly, older children (ages nine to twelve) did not make significantly larger or more variable errors than adults, suggesting one of three things: either children have reached adult-like levels of singing accuracy by this age, children have reached adult-like levels of accuracy on this particular, frequently rehearsed song, or *The Alphabet Song* was not sufficiently difficult to challenge adults and older children, thus creating a ceiling effect. The possibility of a ceiling effect was addressed by running The Shapiro-Wilk Test for Normality on MIE and sdSIE data distributions for 9 to 12-year-olds and adults in both feedback conditions. The results indicated that data from all groups were distributed normally, which rules out the possibility of a ceiling effect on performance of this particular song. We can conclude that older children and adults performed comparably in terms of pitch accuracy and precision in the current study. Regarding whether or not children ages nine to twelve have truly reached adult-like levels of accuracy in general or only with *The Alphabet Song*, replication using a more challenging, less-rehearsed song would give a more definitive answer.

Also as predicted, we found a significant main affect of feedback condition, demonstrating that masking auditory feedback resulted in larger and more variable errors, as well as a tendency (trending toward significance) toward a more variable tonal center –

or being less able to stay “in key”. We also predicted an interaction between age and feedback condition such that the youngest children would be most dependent on auditory feedback to sing well, and this was not what we found. Instead, we found that all participants were equally disrupted by being unable to hear themselves sing. This suggests that self-regulatory use of auditory feedback cannot account for normative, age-related improvement in singing, and in fact, may play very little role at all.

Different Feedback Loops for Different Aspects of Vocal Production

Although a statistical interaction between age and feedback condition was not found in the current study of singing, past work has uncovered such an interaction for speaking (MacKay, 1968; Ratner et al., 1964; Siegel et al., 1980). The fact that we see a statistical interaction between age and use of auditory feedback in some cases but not others may suggest that different feedback systems operate simultaneously for different aspects of vocal development, a possibility raised by Siegel et al. (1980). When singing is compared to speaking, with its increased emphasis on steady state vocal production and pitch, it is possible that additional different feedback systems operate to help the singer regulate his or her voice. This could explain why younger children might be significantly more affected by DAF when speaking than older children and adults, as shown in work by MacKay (1968), Ratner et al. (1964), and Siegel et al. (1980), but not more affected by auditory masking when singing than older children and adults in the current study.

Furthermore, it is possible that pitch control – in contrast with the temporal features of speech examined in past research – may not be an aspect of vocal production that naturally becomes automatized with maturation. Rather, perhaps automatization of pitch control using proprio-kinesthetic awareness in the absence of feedback only occurs with extensive vocal training or a naturally excellent ear – a possibility which is supported by other work showing trained singers show less disruption than untrained singers (Erdemir & Rieser, in review; Mürbe et al., 2004; Schultz-Coulon, 1978; Watts et al., 2002) and that untrained talented singers are less affected by masking than untrained non-talented singers (Watts et al., 2002).

Further support for the idea that different feedback loops may support different facets of vocal production lies in developmental work on the so-called Lombard effect, or the tendency of individuals to spontaneously increase the intensity of volume of their speech under conditions of excessive noise (Lombard, 1911, as cited in Lane & Tranel, 1971). Developmentally, this effect has been shown in children as young as three and four (Siegel et al., 1976), but like the current study, past work on the Lombard effect shows no interaction between age and auditory feedback condition. Siegel and colleagues (1976) showed a significant Lombard effect on participants of three age groups such that vocal intensity (or loudness) increased with masking level. However, younger children (three year-olds) were not more or less affected than older children (four-year-olds) and adults. This suggests that there is no difference in the degree to which preschoolers and adults utilize self-regulatory auditory feedback mechanisms to control their speaking intensity. The lack of a developmental trend in Lombard research and in the current study, combined with the presence of a developmental trend in work on DAF and speech development, supports the idea that different feedback loops may operate simultaneously for different aspects of vocal development. Perhaps some elements of vocal production – like temporal features of speech – become automatized with normative experience while others – like pitch and intensity – may become automatized only with specialized training and otherwise continue to be dependent on an individual's ability to hear his or her own voice when producing sound. The evidence from the current study and the reviewed literature on developmental change in how children use self-regulatory auditory feedback for different tasks (including speaking and singing) suggests that there are several feedback systems in place which may operate independently and follow different developmental trajectories.

Vocal-Motor Control and Strength of Auditory Representations

If self-regulatory use of auditory feedback is not one of the skills or components of singing that improves with age, what other explanations exist? Children's vocal-motor control and the strength of their auditory representation of familiar songs are two factors not addressed in the current study that could contribute to age-related improvements in singing. A 2012 study by Hutchins and Peretz provides support for investigating these two factors. Hutchins and Peretz conducted a series of studies investigating the causes of poor singing in adults, and concluded that although several factors seem to be at play simultaneously, poor motor control and sensorimotor mapping errors seem to be larger contributors than a purely perceptual deficit. By using a manually operated instrument called a slider, they enabled participants to match pitches without using their voices. Their results showed that adults were significantly more accurate using the slider than they were matching pitches vocally, indicating that vocal production may be the primary limiting factor on singing ability in adults – not perceptual deficits. In other words, poor singers may have a perfectly accurate representation of a pitch that they wish to reproduce, but they lack the necessary vocal-motor control to reproduce it.

This work suggests two possible directions for future developmental work on singing centered on children's vocal-motor development and the strength of their auditory representations of songs. Improvements in vocal-motor control are likely to contribute to improvements in singing accuracy with increased age over the course of development, and future studies are needed to examine its relative contribution. One approach could involve using a slider with younger children to see if their pitch matching accuracy is significantly improved over their ability to match pitch vocally at the same age. Could young children using a slider demonstrate accuracy comparable to older children matching pitch vocally? If so, it would provide support for the idea that vocal-motor control is a key area of development that accounts for age-related singing improvement.

Additionally, this work suggests a possible role of auditory representation of melody as a skill that may improve with age. Hutchins and Peretz (2012) asked adults to match target pitches, whereas in the current study, children were asked to reproduce a well-known song from memory. It is unclear how much the accuracy of children's auditory representations might have varied, thus affecting the accuracy of their efferent motor plans for reproducing the song in the first place. For example, some children may have heard more or less accurately rendered versions of *The Alphabet Song* over the course of their childhoods due to family members being more or less accurate singers. Additionally, it's unknown whether children's ability to encode and represent the melodic and rhythmic elements of the song varies greatly between individuals or more normatively with age. The salience of a child's auditory representation of a song likely affects their efference, or the intentions and motor plans that govern their attempts to reproduce that song. If, as indicated by the current study, children's use of auditory feedback is not significantly improving with age and contributing to the normative improvements in singing accuracy that we see in general, growth in vocal-motor control and increased strength and salience of auditory representation are areas which deserve further investigation.

Implications for Teaching

The results of the current study suggest that children's ability to utilize proprio-kinesthetic feedback to control pitch when singing may not develop in the absence of specialized training, while the evidence from adult singing studies (Erdemir & Rieser, in review; Mürbe et al., 2004; Schultz-Coulon, 1978) suggests that singing training can facilitate increased use of proprio-kinesthetic cues in adults when auditory feedback is unavailable. Therefore, it is reasonable to ask whether singing training might facilitate decreased reliance on auditory feedback in children as well. The practical advantage of decreased reliance on auditory feedback is that singers are often in environments in which hearing oneself is difficult, from loud choral settings to live singing with a band. As early as elementary school, children may sing well alone with a music teacher, for example, but struggle to control pitch when singing in a group setting. Additionally, there is intrinsic value in singers learning to attend to proprio-kinesthetic cues, which give additional information about the action in which the singer is engaged and can only strengthen an individual's efferent motor plan once he or she is attuned to such cues. The fact that proprio-kinesthetic awareness of the act of singing does not seem to develop normatively among children ages nine to twelve and adults who are not trained singers, as indicated by the current study, suggests a need for singing training for children that focuses on bodily cues and attending to proprio-kinesthetic feedback. One approach to this could involve practicing singing while auditory feedback is masked.

The current study also raises important questions regarding hearing-impaired children and users of cochlear implants who may wish to improve their singing. Children with severe hearing loss may depend to an unusual degree on the use of proprio-kinesthetic cues for singing and prosody when speaking. It is clear that cochlear implants do not accurately preserve auditory frequencies, so intervals like third, fifths, and octaves are distorted in the representation along the cochlea. One question is whether children who are hearing-impaired and children with cochlear implants still rely on auditory feedback as much as hearing children, in spite of their impaired frequency discrimination. If they do not, it would suggest that some types of ongoing auditory experience (namely, impaired frequency discrimination) may increase reliance on proprio-kinesthetic feedback even without singing training.

Despite evidence that cochlear implant users (both children and adults) tend to have severely compromised pitch discrimination and pitch production abilities (Nakata, Trehub, Mitani, & Kanda, 2006; Vongpaisal, Trehub, Schellenberg, 2006), children with cochlear implants are known to listen to and enjoy music informally, as well as participate in a variety of music-related activities like singing, dancing, and playing a instrument (Gfeller, Witt, Spencer, Stordahl, & Tomblin, 1998; Stordahl, 2002). It could be the case that children with cochlear implants rely more on proprio-kinsthetic feedback to participate in and enjoy music as compared to peers with normal hearing. Another important question is whether visual aids and/or verbal instruction could help children with impaired hearing utilize bodily cues to control their pitch when singing. In cases where auditory feedback is unreliable, how can teachers help children learn to attend to proprio-kinesthetic cues?

Conclusion

In conclusion, the current study suggests that self-regulatory use of auditory feedback does not account for the improvements that we see in children's singing over the course of development, which are reliably reflected in pitch accuracy, precision, and tonal stability. Stated simply, children ages five to twelve and adults rely equally on being able to hear themselves in order to control their pitch when singing. Conversely, when it comes to the specialized skill of controlling pitch when singing as opposed to the more universal skill of regulating temporal features of speech, older children and adults are no better than younger children at utilizing proprio-kinesthetic feedback when they are unable to hear themselves. This suggests the need for future work investigating the development of vocal-motor control as well as the strength of children's auditory representations of familiar songs.

Table 1

Study Procedure

1.	Participant and researcher sing the ABC song together to break the ice.
2.	Participant sings the ABC song solo (before ever hearing DAF).
3.	Participant sings the ABC song with DAF to verify that it creates disfluencies.
4.	Participant sings the ABC song with DAF while adjusting babble mask to the level at which disfluencies disappear.
5.	Participant sings the ABC song with DAF and babble mask at level determined by Step 4.
6.	**Participant sings the ABC song *a cappella* under normal conditions.**
7.	**Participant sings the ABC song with the babble mask (level determined by Step 4).**
8.	**Participant sings the ABC song *a cappella* under normal conditions.**
9.	**Participant sings the ABC song with the babble mask (level determined by Step 4).**
10.	Participant verbally answers questions about their experience.

Note: Steps 6-9 were counterbalanced such that half of the participants began with singing *a cappella* and half began singing with the babble.

Table 2

Mean Interval Error (MIE) Calculation

Sung Letter	Median F0 in Hz	Actual Change in Cents	Desired Change in Cents	Absolute Value Error in Cents	Mean Interval Error (MIE)
a	243				
b	264	143.497	0	143.497	
c	372	593.717	700	106.282	
d	383	50.450	0	50.450	
e	404	92.413	200	107.586	
f	400	-17.226	0	17.226	
g	352	-221.309	-200	21.309	
h	314	-197.773	-200	2.226	
i	309	-27.789	0	27.789	
j	289	-115.844	-200	84.155	
k	293	23.797	0	23.797	
l	269	-147.953	-200	52.046	
m	279	63.190	0	63.190	
n	279	0	0	0	
o	278	-6.216	0	6.216	
p	237	-276.237	-200	76.237	
q	377	803.612	700	103.612	
r	371	-27.774	0	27.774	
s	337	-166.404	-200	33.595	
t	308	-155.781	-200	44.218	
u	310	11.205	0	11.205	
v	288	-127.439	-200	72.560	
w	370	433.747	500	66.252	
x	338	-156.602	-200	43.397	
y	292	-253.265	-200	53.265	
z	271	-129.210	-200	70.789	52.347

Note: This table represents a single performance of *The Alphabet Song*. The formula used for calculating actual change in cents, or the distance between two adjacent sung pitches *a* and *b* (which are given in Hz), is $1200 * \log_2 (\frac{b}{a})$. For any given participant, a single average MIE score was computed for each feedback condition.

Table 3

Standard Deviation of Signed Interval Error (sdSIE) Calculation

Sung Letter	Median F0 in Hz	Actual Change in Cents	Desired Change in Cents	Signed Interval Error in Cents	Standard Deviation of Signed Interval Error (sdSIE)
a	243				
b	264	143.497	0	143.497	
c	372	593.717	700	-106.282	
d	383	50.450	0	50.450	
e	404	92.413	200	-107.586	
f	400	-17.226	0	-17.226	
g	352	-221.309	-200	-21.309	
h	314	-197.773	-200	2.226	
i	309	-27.789	0	-27.789	
j	289	-115.844	-200	84.155	
k	293	23.797	0	23.797	
l	269	-147.953	-200	52.0466	
m	279	63.190	0	63.190	
n	279	0	0	0	
o	278	-6.216	0	-6.216	
p	237	-276.237	-200	-76.237	
q	377	803.612	700	103.612	
r	371	-27.774	0	-27.774	
s	337	-166.404	-200	33.595	
t	308	-155.781	-200	44.218	
u	310	11.205	0	11.205	
v	288	-127.439	-200	72.560	
w	370	433.747	500	-66.252	
x	338	-156.602	-200	43.397	
y	292	-253.265	-200	-53.265	
z	271	-129.210	-200	70.789	64.064

Note: This table represents a single performance of *The Alphabet Song*. The formula used for calculating actual change in cents, or the distance between two adjacent sung pitches *a* and *b* (which are given in Hz), is $1200 * \log_2 (\frac{b}{a})$. For any given participant, a single sdSIE score was computed for each feedback condition.

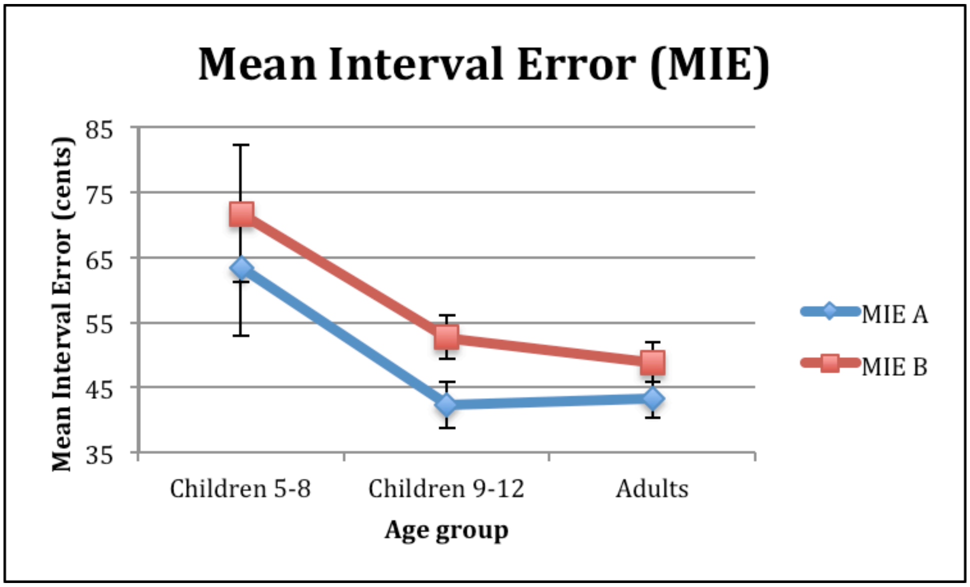

Figure 1. Mean Interval Error (MIE) is a measure of average size of absolute value errors for each participant in each feedback condition, and it reflects pitch accuracy. MIE *A* refers to the MIE in the normal auditory feedback condition and MIE *B* refers to the babble mask condition. Analysis showed a significant effect of age (F=4.359, p<.01) and a significant effect of feedback condition (F=19.604, p<.01). Post hoc testing with Bonferroni correction (adjusted α=.0125) revealed that children ages 5 to 8 differed from children ages 9 to 12 in their MIE under both normal auditory feedback and babble mask conditions (normal auditory feedback: t=2.445, p=.013; babble mask: t=1.717, p=.011), but that children ages 9 to 12 did not differ significantly from adults in terms of MIE in either condition (normal auditory feedback: t=-.26, p=.127; babble mask: t=.826, p=.719).

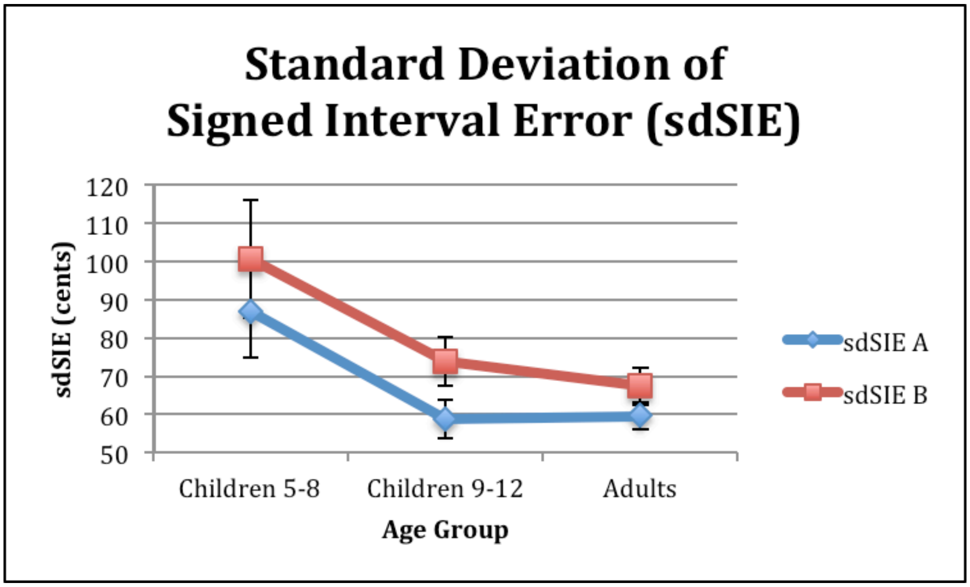

Figure 2. Standard deviation of signed interval error (sdSIE) is a measure of precision; it answers the question, how variable were a participant's errors in each feedback condition? *A* refers to the sdSIE in the normal auditory feedback condition and *B* refers to the babble mask condition. There was a significant effect of feedback condition on sdSIE (F=21.409, p<.01) and a significant effect of age on sdSIE (F=3.737, p=.033). Similar to the pattern of results with MIE, post hoc testing with Bonferroni correction (adjusted α=.0125) revealed that although 5 to 8 year-olds differed from 9 to 12 year-olds in normal feedback conditions (t=2.150, p=.006) and were nearly significantly different in the babble mask condition (t=1.624, p=.019), older children were not significantly less variable in their errors than adults (normal auditory feedback: t=-.125, p=.252; babble mask: t=.802, p=.462).

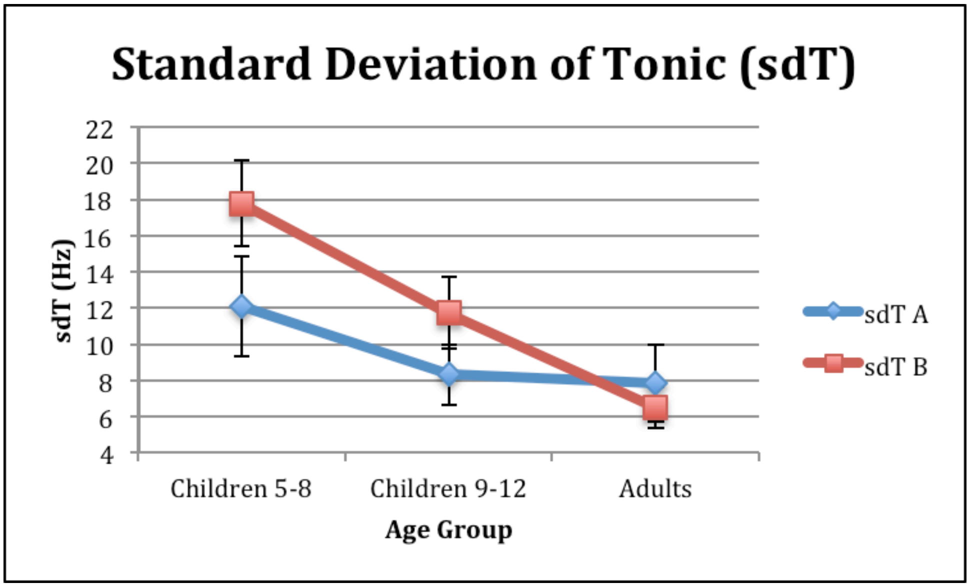

Figure 3. Standard deviation of tonic (sdT) is a measure of the extent to which a participant stayed in the key he or she started in. Smaller numbers indicate more accurate tonal center. *A* refers to the sdT in the normal auditory feedback condition and *B* refers to the babble mask condition. There was a near-significant effect of feedback condition on sdT such that participants' tonal center was more stable in the normal auditory feedback condition than in the babble mask condition (F=3.815, p=.058). There was also a significant effect of age of sdT, showing that older children and adults stay "in key" better than younger children (F=5.165, p=.01). The interaction between age and feedback condition on sdT was non-significant, indicating that the ability of participants in all age groups to maintain a stable tonal center was equally disrupted by not being able to hear themselves sing (F=2.517, p=.094).

REFERENCES

Bentley, A. (1969). Papers of the international seminar on experimental research in music education, University of Reading, Reading, England, July 9-16, 1968. *Journal of Research in Music Education, 17*(1), 3-159.

Boersma, P., & Weenik, D. (2008). "Praat: Doing phonetics by computer (Version 5.3.81)," http://222.praat.org/ (Last viewed 7/15/2014).

Dalla Bella, S., Giguère, J. F., & Peretz, I. (2007). Singing proficiency in the general population. *Journal of the Acoustical Society of America*, 121, 1182-1189.

Elliott, L., & Niemoeller, A. (1970). The role of hearing in controlling voice fundamental frequency. *International Audiology*, IX, 47-52.

Erdemir, A., & Rieser, J. (2014). Singing without hearing: The use of auditory and motor information when singers, instrumentalists, and non-musicians sing a familiar tune. Manuscript in review.

Geringer, J. M. (1983). The relationship of pitch-matching and pitch-discrimination abilities of preschool and fourth-grade students. *Journal of Research in Music Education, 31*(2), 93-99.

Gfeller, K., Witt, S. A., Spencer, L. J., Stordahl, J., & Tomblin, B. (1998). Musical involvement and enjoyment of children who use cochlear implants. *The Volta Review, 100*(4), 213-233.

Grillo, E. U., Abbott, K. V., & Lee, T. D. (2010). Effects of masking noise on laryngeal resistance for breathy, normal, and pressed voice. *Journal of Speech, Language, and Hearing Research, 53*(4), 850-861. doi:http://dx.doi.org/10.1044/1092-4388(2009/08-0069.

Hornbach, C. M., & Taggart, C. C. (2005). The relationship between developmental tonal aptitude and singing achievement among kindergarten, first-, second-, and third-grade students. *Journal of Research in Music Education, 53*(4), 322-331.

Hutchins, S. M., & Peretz, I. (2012). A frog in your throat or in your ear? searching for the causes of poor singing. *Journal of Experimental Psychology: General, 141*(1), 76-97. doi:http://dx.doi.org/10.1037/a0025064.

Lane, H. L., & Tranel, B. (1971) The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, 14, 677-709.

Lee, B. S. (1951). Artificial stutter. *Journal of Speech & Hearing Disorders, 16*, 53-55.

MacKay, D. G. (1968). Metamorphosis of critical interval: Age-linked changes in the delay in auditory feedback that produces maximal disruption of speech. *Journal of the Acoustical Society of America*, 43, 811-821.

Mürbe, D., Pabst, F., Hofmann, G., & Sundberg, J. (2002). Significance of auditory and kinesthetic feedback to singers' pitch control. *Journal of Voice, 16*(1), 44-51. doi:http://dx.doi.org/10.1016/S0892-1997(02)00071-1.

Mürbe, D., Pabst, F., Hofmann, G., & Sundberg, J. (2004). Effects of a professional solo singer education on auditory and kinesthetic feedback--A longitudinal study of singers' pitch control. *Journal of Voice, 18*(2), 236-241. doi:http://dx.doi.org/10.1016/j.jvoice.2003.05.001.

Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *Journal of the Acoustical Society of America*, 128(4), 2182-2190.

Nakata, T., Trehub, S. E., Mitani, C., & Kanda, Y. (2006). Pitch and timing in the songs of deaf children with cochlear implants. *Music Perception, 24*(2), 147-154.

Ratner, S. C., Gawronski, J. J., & Rice, F. E. (1964). The variable of concurrent action in the language of children: Effects of delayed speech feedback. Psychological Record, 14, 47-56.

Schultz-Coulon, H. J. (1978). The neuromuscular phonatory control system and vocal function. *Acta Otolaryngol*, 86, 142-153.

Siegel, G. M., Fehst, C. A., Garber, S. R., & Pick, H. L. (1980). Delayed auditory feedback with children. *Journal of Speech & Hearing Research, 23*(4), 802-813.

Siegel, G. M., Pick, H. L., Olsen, M. G., & Sawin, L. (1976). Auditory feedback on the regulation of vocal intensity of preschool children. *Developmental Psychology, 12*(3), 255-261. doi:http://dx.doi.org/10.1037/0012-1649.12.3.255.

Stordahl, J. (2002). Song recognition and appraisal: A comparison of children who use cochlear implants and normally hearing children. *Journal of Music Therapy, 39*(1), 2-19. doi:http://dx.doi.org/10.1093/jmt/39.1.2.

Ternström, S., & Sundberg, S. J. (1988) Intonation precision of choir singers. *Journal of the Acoustical Society of America*, 84, 59-69.

Vongpaisal, T., Trehub, S. E., & Schellenberg, E. G. (2006). Song recognition by children and adolescents with cochlear implants. *Journal of Speech, Language, and Hearing Research, 49*(5), 1091-1103. doi:http://dx.doi.org/10.1044/1092-4388(2006/078.

Watts, C., Murphy, J., & Barnes-Burroughs, K. (2003). Pitch matching accuracy of trained singers, untrained subjects with talented singing voices, and untrained subjects with nontalented singing voices in conditions of varying feedback. *Journal of Voice, 17*(2), 185-194. doi:http://dx.doi.org/10.1016/S0892-1997(03)00023-7.

Ward, D., & Burns, E. (1978). Singing without auditory feedback. *Journal of Research in Singing*, 1, 4-44.