JSLHR

Research Article

Rhythm Perception and Its Role in Perception and Learning of Dysrhythmic Speech

Stephanie A. Borrie,[a] Kaitlin L. Lansford,[b] and Tyson S. Barrett[c]

Purpose: The perception of rhythm cues plays an important role in recognizing spoken language, especially in adverse listening conditions. Indeed, this has been shown to hold true even when the rhythm cues themselves are dysrhythmic. This study investigates whether expertise in rhythm perception provides a processing advantage for perception (initial intelligibility) and learning (intelligibility improvement) of naturally dysrhythmic speech, dysarthria.
Method: Fifty young adults with typical hearing participated in 3 key tests, including a rhythm perception test, a receptive vocabulary test, and a speech perception and learning test, with standard pretest, familiarization, and posttest phases. Initial intelligibility scores were calculated as the proportion of correct pretest words, while intelligibility improvement scores were calculated by subtracting this proportion from the proportion of correct posttest words.
Results: Rhythm perception scores predicted intelligibility improvement scores but not initial intelligibility. On the other hand, receptive vocabulary scores predicted initial intelligibility scores but not intelligibility improvement.
Conclusions: Expertise in rhythm perception appears to provide an advantage for processing dysrhythmic speech, but a familiarization experience is required for the advantage to be realized. Findings are discussed in relation to the role of rhythm in speech processing and shed light on processing models that consider the consequence of rhythm abnormalities in dysarthria.

Speech rhythm, broadly defined herein as recurring spectral–temporal patterns within the acoustic signal, functions, universally, to reduce the computational load of recognizing spoken language. As summarized by Liss, Utianski, and Lansford (2013), the rhythm afforded in the speaker's productions enables the listener to "track, segment, anticipate, and focus attention on high yield aspects of the speech signal and to facilitate recognition and prediction of syntactic and semantic relationships among words and phrases in continuous speech" (p. 5). Central to recognizing spoken language is lexical segmentation, the process of parsing the continuous speech stream into its individual word components (Jusczyk & Luce, 2002). Rhythm perception plays an important role in this process. In stress-based languages, like English or Dutch, listeners exploit metrical stress to guide their segmentation decisions (Mattys, White, & Melhorn, 2005; McQueen, Norris, & Cutler, 1994). In particular, the presence of strong syllables, those receiving relative stress through longer duration, fundamental frequency change, increased loudness, and a relatively full vowel, informs the onset of a new word (Cutler & Butterfield, 1992; Cutler & Norris, 1988). Mattys et al. (2005) build a strong case that rhythm perception is especially key for speech processing in adverse listening conditions. Here, the authors propose a hierarchical model of cues to lexical segmentation, whereby cue utilization is modulated by the quality and quantity of acoustic and contextual information. According to this framework, when phonemic information is impoverished, rendering it an unreliable marker of word onsets and offsets, listeners rely on lower level cues, namely, rhythm cues, to make their segmentation decisions. Evidence supporting this hypothesis is observed in listener attempts to segment speech presented in noise (Smith, Cutler, Butterfield, & Nimmo-Smith, 1989) or at reduced intensity (Cutler & Butterfield, 1992), whereby lexical misperceptions, or missegmentations, reflect stress cue utilization.

Our previous work has shown that when phonemic information is degraded and rhythm cues corrupted, as is the case of the naturally dysrhythmic signal associated with dysarthria, listeners' lexical misperceptions are still guided by stress-initial interpretations (Borrie, 2015; Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne,

[a]Department of Communicative Disorders and Deaf Education, Utah State University, Logan
[b]School of Communication Sciences and Disorders, Florida State University, Tallahassee
[c]Department of Psychology, Utah State University, Logan
Correspondence to Stephanie A. Borrie: stephanie.borrie@usu.edu

Editor: Julie Liss
Associate Editor: Amy Neel
Received March 8, 2016
Revision received July 15, 2016
Accepted September 11, 2016
https://doi.org/10.1044/2016_JSLHR-S-16-0094

Disclosure: The authors have declared that no competing interests existed at the time of publication.

& Anderson, 2012; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2013; Liss et al., 2013; Liss, Spitzer, Caviness, Adler, & Edwards, 2000). It appears that listeners are able to glean something from the abnormal metrical stress to make weighted predictions regarding the onsets and offsets of words. Further, our studies demonstrate that listeners learn to better exploit the dysrhythmic properties of the dysarthric signal. Relative to a control group, listeners familiarized with dysarthric speech were not only better at recognizing the neurologically degraded speech but also relied more heavily on metrical stress to make their segmentation decisions (Borrie et al., 2013; Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, et al., 2012). This finding suggests that experience with dysarthric speech provides an opportunity for listeners to learn something about the abnormal rhythm patterns, which, in turn, helps them better navigate the signal in subsequent encounters.

Current models of perception and learning of dysarthric speech are based largely on group averages, yet evidence of substantial individual variability exists, suggesting that some listeners are better equipped to tackle the complex task of perceiving dysrhythmic speech (Borrie, 2015). Few studies, however, have explored the factors that may account for the variability observed. In a recent study examining the influence of vocabulary on older and younger adult listeners processing dysarthric speech, McAuliffe, Gibson, Kerr, Anderson, and LaShell (2013) observed a significant relationship between receptive vocabulary scores and speech intelligibility scores, concluding that the greater an individual's receptive vocabulary is, the more accurately they will recognize dysarthric speech. More recently, in an attempt to delineate the factors that contribute most significantly to word recognition performance of unfamiliar speech varieties, including dysarthric speech, Bent, Baese-Berk, Borrie, and McKee (2016) assessed three cognitive-linguistic areas—receptive vocabulary, cognitive flexibility, and selective attention. The study confirmed the finding of McAuliffe et al. (2013), reporting a significant relationship between receptive vocabulary and speech intelligibility; however, neither cognitive flexibility nor selective attention accounted for any of the observed variability. These results suggest other factors must be involved.

Given that rhythm perception plays a key role in recognizing spoken language under challenging listening conditions, it is plausible that expertise with rhythm perception, as developed through musical training, might predict processing of dysarthric speech. Indeed, this hypothesis is supported by the notion of the *musical advantage,* which suggests that musicians, relative to nonmusicians, are more successful at deciphering spoken language in challenging listening conditions. The empirical evidence for a cross-domain processing advantage is compelling. People with musical backgrounds demonstrate superior performance in a number of speech perception tasks, including recognizing speech in noise (e.g., Parbery-Clark, Skoe, & Kraus, 2009; Swaminathan et al., 2015; Zendel, Tremblay, Belleville, & Peretz, 2015) and nonnative speech discrimination (Marie, Delogu, Lampis, Belardinelli, & Besson, 2011) and learning (Lee & Hung, 2008; Wong & Perrachione, 2007). There also exists a large body of literature corroborating a neural basis for the musical advantage, demonstrating that musical training induces both structural and functional changes in the brain (e.g., Gaser & Schlaug, 2003; Schneider et al., 2002) and that the neuroplastic benefits associated with musical training extend beyond music processing to speech and language domains (Hannon & Trainor, 2007).

In the current study, we investigated whether expertise in rhythm perception provides an advantage for perception (initial intelligibility) and learning (intelligibility improvement) of dysarthric speech. In particular, we addressed the following two key questions: (a) do rhythm perception abilities predict initial intelligibility of dysarthric speech and (b) do rhythm perception abilities predict the magnitude of intelligibility improvement following familiarization with dysarthric speech? Here, we operationalized rhythm perception as a superior ability to perceive rhythm measured within the musical domain and hypothesized that expertise in rhythm perception will predict superior initial intelligibility of dysarthric speech. In addition, given our previous research demonstrating that listeners can learn to better utilize dysrhythmic speech cues, we hypothesized that expertise in rhythm perception will also predict superior intelligibility improvements for listeners familiarized with dysarthric speech. To quantify rhythm perception abilities, we used the rhythm subtest of the Musical Ear Test (MET; Wallentin, Nielsen, Friis-Olivarius, Vuust, & Vuust, 2010), a validated objective measurement tool for quantifying rhythm perception abilities, specifically sensitivity to temporal differences within musical signals. Superior performance on the MET is shown to predict superior recognition of speech in noise (Slater & Kraus, 2015). To draw links with the musical advantage literature, we obtained self-reports of musician status and examined whether musicians, relative to nonmusicians, performed better on the test of rhythm perception and on the intelligibility outcome measures associated with the study. Receptive vocabulary scores from the Peabody Picture Vocabulary Test–Fourth Edition (PPVT-4; Dunn & Dunn, 2007) were also included in our investigations due to recent evidence regarding their role in predicting initial intelligibility of dysarthric speech (Bent et al., 2016; McAuliffe et al., 2013). Thus, we also addressed the following secondary research questions: (c) do receptive vocabulary scores predict initial intelligibility of dysarthric speech and (d) do receptive vocabulary scores predict the magnitude of intelligibility improvement following familiarization with dysarthric speech? Here, we hypothesized that we would replicate previous findings regarding a relationship between vocabulary knowledge and perception of dysarthric speech and that this predictive correlation would extend to intelligibility improvements.

Method

Participants

Fifty young, healthy university students (34 women and 16 men), aged 18 to 29 years old ($M = 21.38$, $SD = 2.35$), participated in the experiment. All participants were native speakers of American English and passed a pure tone hearing screen at 20 dB HL for 500, 1000, 2000, and 4000 Hz in both the right and left ears. As per self-report, participants had no history of speech, language, or cognitive disorders and no prior experience with dysarthric speech. Participants were recruited from undergraduate classes at Utah State University, and institutional review board consent was obtained from each participant.

Speech Stimuli

Speech stimuli used in the current study were collected as part of a larger study conducted in the Motor Speech Disorders Lab at Arizona State University (Liss et al., 2009). The current study draws on an audio-recorded passage reading and a series of semantically anomalous phrases elicited from an 84-year-old male native speaker of American English with dysarthria secondary to cerebellar disease. The passage reading, an adapted version of the standard "Grandfather Passage" (Darley, Aronson, & Brown, 1975), consisted of 35 phrases, ranging in length from three to 12 words and containing between three and 14 syllables per phrase. This audio recording, paired with orthographic transcription, served as linguistically rich familiarization stimuli (familiarization speech set) in the speech perception and learning test.

The 80 semantically anomalous phrases (e.g., *amend estate approach* and *had eaten junk and train*) in their audio form served as the pre- and posttest speech sets in the speech perception and learning test. These test set phrases were syntactically plausible but semantically anomalous to reduce the influence of higher level cognitive cues on word recognition (Liss et al., 2000). In addition, all test set phrases contained six syllables with alternating metrical stress and ranged in length from three to five words.

Dysarthria, with its neurogenic origins and associated motor control deficits, corrupts rhythm at the level of articulatory implementation and disrupts the spectral–temporal flow of speech (Liss et al. 2009, 2013). The speaker who produced the speech stimuli in this study exhibited the cardinal features of an ataxic dysarthria of moderate severity, as diagnosed by three certified speech-language pathologists with expertise in assessment and differential diagnosis of motor speech disorders. His speech was characterized perceptually by excess and equal stress (scanning speech), prolonged phonemes and intervals, monotone, monoloudness, and imprecise articulation with irregular articulatory breakdowns (Darley, Aronson, & Brown, 1969). The presenting rhythm abnormalities were acoustically validated in earlier studies (Lansford & Liss, 2014a, 2014b; Liss et al., 2009). The unique presentation of ataxic dysarthria interferes with speech perception (Liss et al., 2013), perceptual learning (Liss, Spitzer, Caviness, & Adler, 2002), and entrainment processes involved in conversational speech (Borrie & Liss, 2014). Ataxic dysarthria, therefore, provides a good entry point for investigations into whether superior rhythm perception abilities afford an advantage for processing dysrhythmic speech.

Procedure

Each participant attended a single experimental session held in the Human Interaction Lab at Utah State University. Upon obtaining informed consent, participants engaged in an identical experimental paradigm involving three key tests, including a rhythm perception test, a receptive language test, and a speech perception and learning test. They also completed a brief questionnaire regarding their musical background. The tests were administered via a computer preloaded with audio stimuli and task-specific instructions; the order of administration was randomized across participants. Stimuli were presented binaurally through sound-attenuating headphones (Sennheiser HD 650 PRO, Old Lyme, CT) at a comfortable listening level of 65 dB HL.

Rhythm Perception Test

The rhythm subtest of MET is validated for quantifying rhythm perception abilities in the musical domain and for distinguishing between musicians and nonmusicians (Wallentin et al., 2010). The test uses a forced choice paradigm for which participants must judge whether pairs of rhythmical phrases played with a wood block are the same or different. The rhythm sequences within each phrase contain 4–11 wood block beats and have a duration of one measure played at 100 beats per minute. When phrases are *different,* they differ from one another by a single rhythmic change (for further details on how rhythm complexity is varied across phrase pairs, see Wallentin et al., 2010). The test consists of 52 rhythmical phrase pairs for the participant to judge. No feedback regarding judgment accuracy is given at the time of testing.

Receptive Vocabulary Test

PPVT-4, is a norm-referenced, standardized assessment that provides a measure of receptive vocabulary for Standard American English (Dunn & Dunn, 2007). The test, administered online via Q-global (https://qglobal.pearsonclinical.com), has 228 items grouped into 19 sets of 12 items each. The sets are arranged in order of increasing difficulty, and participants start at a set (Start Set) predicted by chronological age. Participants are presented with a spoken word (e.g., *truck* and *lifting*) and four corresponding images and then are asked to identify the image that best illustrates the auditory production. If the participant makes one or zero errors in the set (Basal Set), then testing continues with more difficult sets. If, however, the participant makes more than one error, testing will drop back to an easier set until one or zero errors are made and a Basal Set for the participant is established. Participants continue to work through sets until they reach their Ceiling Set, which is defined as the highest set of items administered containing eight or more errors.

Speech Perception and Learning Test

To assess speech perception and learning of dysarthric speech, we used a three-phase testing protocol involving pretest, familiarization, and posttest parts, administered according to instructions from our previous studies (e.g., Borrie, McAuliffe, Liss, Kirk, et al., 2012; Lansford, Borrie, & Bystricky, 2016). During the pretest phase, participants were informed that they would be presented with short phrases produced by someone with a speech disorder and that while the phrases all contained real English words, they would not all necessarily make sense. Phrases of the pretest speech set were presented one at a time, and following each presentation, participants were asked to use the keyboard to type what they heard being said. Participants were encouraged to guess if unsure and given as much time as necessary to type a response. Following the pretest phase, participants engaged in a familiarization phase in which they listened to the passage reading stimuli and used written subtitles provided on the computer monitor to help them understand what was being said. After this, participants engaged in a posttest identical in structure to the pretest phase but using novel testing stimuli (i.e., the posttest set).

Data Analysis

For each participant, the following four scores were obtained: (a) rhythm perception score, (b) receptive vocabulary score, (c) initial intelligibility score, and (d) intelligibility improvement score. In addition, participants were categorized as either musician or nonmusician, according to the following criteria: musician status defined as playing an instrument or singing for at least 7 years and nonmusician status defined as no formal musical training or practice. Using a scoring sheet provided by Wallentin et al. (2010), rhythm perception scores were calculated as the percentage of correct responses on the 52 pairs of phrases that include the rhythm subtest of the MET. Using standardized scoring procedures outlined in Dunn and Dunn (2007), receptive vocabulary scores were calculated as the raw PPVT-4 score. In this calculation, a participant's total number of errors is subtracted from their ceiling item number, giving them credit for all items below their Basal Set but no credit for items above their Ceiling Set. Age-appropriate reference norms are provided to identify potential vocabulary deficits.

Initial intelligibility and intelligibility improvements score computations were based on listener transcriptions of the pretest and posttest speech sets during the speech perception and learning test. Transcription analysis involved a trained judge classifying words as correct or incorrect on the basis of previously established transcription scoring criteria (e.g., Borrie, McAuliffe, Liss, Kirk, et al., 2012; Liss et al., 2002). According to this criteria, words are scored as correct if they match the intended target precisely or if they differ only by tense (*-ed*) or plurality (*-s*). Word substitutions between *a* and *the* are also scored as correct. The percentage of words correct (PWC) score is then tabulated for each speech set for each participant. The pretest PWC score reflects a measure of intelligibility prior to familiarization and thus, with no further transformation, forms the initial intelligibility score. The intelligibility improvement score is calculated by subtracting the pretest PWC score from the posttest PWC score.

Both initial intelligibility and intelligibility improvement scores are proportions of correct items bounded between 0 and 1. Therefore, beta regression models were used to test the predictive relationships between the two intelligibility scores and the receptive vocabulary and rhythm perception scores. Beta regression, although a relatively new method, is built specifically for bounded outcomes, such as proportions of correct items (for a review on beta regression in the social sciences, see Smithson & Verkuilen, 2006). It has two submodels: one for the mean (location submodel) and the other for the variability (dispersion submodel). The location submodel works in a similar way as linear regression but for bounded outcomes. In addition, the dispersion submodel allows information from the predictors to better estimate the nonnormal distribution that is common in proportional data. This improved understanding of the dispersion provides more accurate estimates of the standard errors (Cribari-Neto & Zeileis, 2010; Simas, Barreto-Souza, & Rocha, 2010). In these ways, both the estimated effect and the level of significance are more accurate with beta regression than when using other common methods (e.g., other generalized linear models).

Reliability Analysis

Twenty percent of the transcriptions (10 pretest and 10 posttest) were randomly selected according to computer-generated random number lists and reanalyzed by the original judge (intrajudge) and a second trained judge (interjudge) to obtain reliability estimates for the PWC scores. Reliability analysis confirmed that the agreement rate between the reanalyzed data and the original data was high (all correlations $r > .98$).

Results

Participant Scores

All 50 participants scored within normal limits on the PPVT-4 assessment ($M = 205.82$, $SD = 8.57$), denoting that although a range of scores were observed, receptive vocabulary abilities were age appropriate for each participant. Thus, we can objectively confirm that the study participants did not present with receptive language deficits that might interfere with processing dysarthric speech. Rhythm perception scores on the MET rhythm subtest, reported as percentage scores, ranged from 53.85% to 88.46% ($M = 71.92$, $SD = 8.31$). As anticipated, results of a correlation analysis demonstrated that participant's receptive vocabulary scores and rhythm perception scores were unrelated (see Table 1 for results of a bivariate correlation analysis among study variables).

Intelligibility data, reported as percentages, are presented in Figure 1. Of note is the large individual variability evident in both outcome measures for initial intelligibility scores and intelligibility improvement scores. The former range from 23.17% to 57.32% ($M = 48.25$, $SD = 8.79$), and the latter range from 8.78% to 33.33% ($M = 19.63$, $SD = 5.89$), illustrated by the gradient of the line connecting

Table 1. Correlations among study variables.

Variable	1	2	3	4
1. Receptive vocabulary score	—			
2. Rhythm perception score	−.029	—		
3. Initial intelligibility	.389*	.047	—	
4. Intelligibility improvement	−.020	.721*	−.218	—

*$p < .01$.

an individual's pretest and posttest scores. Results of a correlation analysis demonstrated that a participant's initial intelligibility scores and intelligibility improvement scores were unrelated (see Table 1).

Prediction Analysis

Two beta regression models with a log–log link were conducted to assess the relationships between the scores for receptive vocabulary and rhythmic perception and the scores for initial intelligibility and intelligibility improvement. The final models were selected on the basis of the highest log likelihood, as suggested by Smithson and Verkuilen (2006). These final models included receptive vocabulary and rhythmic perception scores as predictors in the dispersion submodels to better account for the heteroscedasticity. Table 2 shows the estimates of the final models for both intelligibility outcome measures. Initial intelligibility scores were significantly predicted by receptive vocabulary scores ($p = .024$) but not rhythm perception scores ($p = .758$). For a 10% increase in the receptive vocabulary score, there was, on average, a 6% increase in the initial intelligibility score. Receptive vocabulary scores were marginally significant ($p = .071$) in the dispersion model, with higher scores predicting lower variability. In contrast, intelligibility improvement scores were significantly predicted by rhythm perception scores ($p < .001$) but not receptive vocabulary scores ($p = .317$). For a 10% increase in the rhythm perception score, there was, on average, a 13% increase in the intelligibility improvement score. Both rhythm perception and receptive vocabulary were significant predictors of the variability in improvement ($p = .001$ and $p < .001$, respectively), with higher scores of both predicting lower variability. Overall, with the inclusion of the dispersion submodels, each model is more accurate. The results of the location submodel parallel those of the correlation reported in Table 1.

Analysis by Musician Status

When grouped according to the musician ($n = 24$) versus nonmusician ($n = 26$) dichotomy, a two-tailed, independent samples *t* test, $t(48) = 8.12$, $p < .001$, Cohen's $d = 2.31$, revealed that the musicians ($M = 78.28$, $SD = 5.99$) achieved significantly higher rhythm perception scores than the nonmusicians ($M = 65.09$, $SD = 5.41$). Group averages for intelligibility outcomes are displayed in Figure 2. An independent samples *t* test on initial intelligibility scores revealed no significant difference between the musicians ($M = 47.92$, $SD = 8.19$) and nonmusicians ($M = 49.58$, $SD = 4.79$). In contrast, a significant difference between the musicians ($M = 22.65$, $SD = 4.56$) and nonmusicians ($M = 16.83$, $SD = 3.59$) was revealed for intelligibility improvement scores. These findings provide additional validation that the MET is capable of distinguishing between musicians and nonmusicians (Wallentin et al., 2010). Moreover, they support the results of our prediction analyses

Figure 1. Pretest and posttest intelligibility scores, by each participant, reflect large individual variability in both initial intelligibility scores and intelligibility improvement scores (slope of connecting line). Participants are categorized according to musician versus nonmusician dichotomy to facilitate ease of interpretation.

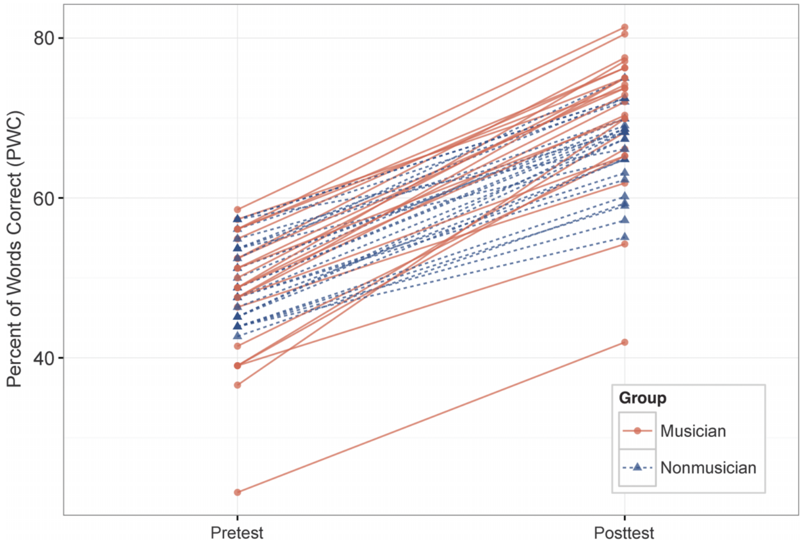


Table 2. Summary of beta regression analysis.

Score	Initial intelligibility			Intelligibility improvement		
	b	*SE*	*p*	b	*SE*	*p*
Location submodel						
Receptive vocabulary score	0.006	0.003	.024	0.002	0.002	.317
Rhythm perception score	0.001	0.003	.758	0.013	0.002	<.001
Dispersion submodel						
Receptive vocabulary score	−0.042	0.023	.071	−0.093	0.023	<.001
Rhythm perception score	0.004	0.023	.864	−0.074	0.023	.001

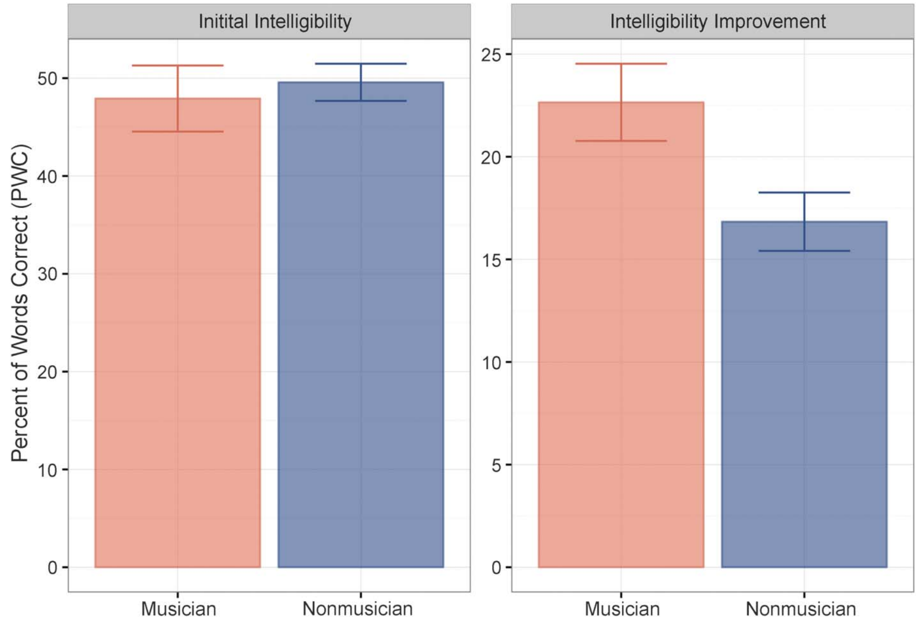
Note. B is the estimate of the effect of the variable on the outcome in logged units; *SE* is the standard error of B. The location submodel had a log–log link, and the dispersion submodels had a log link.

and, thus, bridge this study to the musical advantage literature.

Discussion

Here, we provide the first empirical evidence of the relationship between rhythm perception and processing of a naturally dysrhythmic speech signal. Our first key hypothesis, that expertise in rhythm perception would predict superior initial intelligibility of dysarthric speech, was not confirmed—rhythm perception scores did not predict initial intelligibility scores. Thus, it appears that superior rhythm perception abilities, as measured within the musical domain, do not aid listeners in their very initial attempts to navigate a dysrhythmic speech signal. Considered in isolation, this finding was not expected, particularly given previous work demonstrating that listeners utilize corrupted rhythm cues to segment dysarthric speech (e.g., Borrie, McAuliffe, Liss, Kirk, et al., 2012; Liss et al., 2000) and that superior rhythm skills, also measured by the MET, predict superior intelligibility of speech in noise (Slater & Kraus, 2015). However, the current findings confirmed our second key hypothesis, that expertise in rhythm perception would predict superior intelligibility improvements for listeners familiarized with dysarthric speech—rhythm perception scores were a robust predictor of intelligibility improvement scores. Thus, it appears that superior rhythm perception abilities aid listeners in adapting to a dysrhythmic speech signal. Taken together, the findings suggest that expertise in rhythm perception provides a perceptual advantage for processing of dysrhythmic

Figure 2. Group means of intelligibility outcome measures when categorized according to the musician versus nonmusician dichotomy: the panels reflect no group difference in initial intelligibility of dysarthric speech (left) but a significant group difference in intelligibility improvement following familiarization with dysarthric speech (right). Error bars delineate ±1 *SEM* (standard error of the mean).



speech; however, a brief opportunity to learn something about the aberrant signal properties is required for the advantage to be realized.[1]

Borrie and colleagues (Borrie et al., 2013; Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, et al., 2012) have previously demonstrated that familiarization with dysarthric speech and the associated written targets affords an opportunity for listeners to learn something about the aberrant suprasegmental properties by drawing attention to relevant acoustic information. In this series of complementary studies, Borrie and colleagues observed that listeners familiarized with dysarthric speech, relative to listeners familiarized with neurologically healthy speech, were not only better able to decipher the degraded speech but also relied more heavily on metrical stress to inform their speech segmentation decisions. This was taken as evidence that given the opportunity, listeners can learn to better process the abnormal speech rhythm cues that characterize dysarthric speech. Novel here is the idea that listeners with expertise in rhythm perception, or put another way, an aptitude for recognizing rhythm cues in the musical domain, are better able to exploit familiarization as an opportunity to learn something useful about corrupted rhythm cues in the speech domain.

The idea that rhythm perception skills in the musical domain enable an individual to more successfully navigate rhythm cues in the speech domain forms a central assumption underlying the musical advantage (Kraus & Chandrasekaran, 2010). Indeed, the results of the present study also contribute to this body of literature. Here, we see that despite formal musical training, with links to expertise in rhythm perception confirmed,[2] musicians were no better than nonmusicians at deciphering dysarthric speech for the first time. This finding, in direct contrast to the literature evidencing a musical advantage in challenging listening conditions, suggests that musical training may be of little initial value when the spectral–temporal cues in the speech domain are themselves corrupted. Such a hypothesis is supported, in part, by a recent study examining degraded pitch conditions of cochlear implant simulations, observing that under some noise settings, the musical advantage for word and sentence identification was notably absent (Fuller, Galvin, Maat, Free, & Baskent, 2014).

However, the present data demonstrate that musical training does aid in learning to better understand dysarthric speech—musicians significantly outperformed nonmusicians on scores reflecting magnitude of intelligibility improvement following familiarization with dysarthric speech. Thus, the anticipated cross-domain processing advantage induced by musical training is only realized when listeners are given a chance to get acquainted with the dysrhythmic speech signal. Although the mechanisms that give rise to the musical advantage are not well understood, the cross-domain processing benefits certainly make intuitive sense. From an acoustic view, music and speech share many similar qualities. Both are hierarchically organized, rule-based systems whose processing requires a precise representation of acoustic features, such as pitch, timbre, and timing (Kraus, Skoe, Parbery-Clark, & Ashley, 2009). Further, both signals recruit similar auditory processes for analysis, and musicians have intensive training of these processes (e.g., Cummins, 2013). For example, musicians outperform nonmusicians in auditory-perceptual tasks, including syllable discrimination (Zuk et al., 2013), pitch discrimination (Micheyl, Delhommeau, Perrot, & Oxenham, 2006), duration discrimination (Jeon & Fricke, 1997), and encoding of rhythm contour (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004). The results of the current study suggest that the pathological rhythm patterns that characterize dysarthric speech may interfere with the mechanism underlying the musical advantage but that this interference is assuaged by a brief opportunity to learn about the abnormal speech rhythm cues. Further, the findings also implicate superior rhythm perception abilities as contributing, at least in part, to a mechanistic account of the musical advantage.

An alternative explanation of our findings is that during the familiarization experience, individuals with expertise in rhythm sensitivity are quick to recognize that the dysrhythmic cues are not particularly useful for processing and so turn their attention to the acoustic–phonetic regularities in the speech signal. This would, in turn, support a retuning of their phonological categories to include the disordered articulatory productions and result in improved intelligibility. That familiarization aids listeners in processing segmental information also supported by earlier studies examining perceptual learning of dysarthric speech (e.g., Borrie, McAuliffe, Liss, Kirk, et al., 2012; Liss et al., 2002). Note that because rhythm sensitivity was not targeted in the previous literature with perceptual learning of dysarthric speech, it is unknown if the listeners possessed strong rhythm perception skills. Indeed, consideration of this listener-related factor may have yielded a different pattern of results. Work in our labs is now examining how rhythm perception abilities may differentially influence listeners' use of segmental and suprasegmental speech cues in perception and learning of dysrhythmic speech.

We turn our attention now to our secondary research questions investigating the influence of receptive vocabulary on intelligibility outcomes associated with processing a dysrhythmic speech signal. Here, we observed that receptive vocabulary scores, as measured by the standardized PPVT-4 assessment tool, positively predicted initial intelligibility of dysarthric speech. This finding is consistent with two earlier studies in which superior receptive vocabulary scores, also measured using the PPVT-4, tracked superior initial intelligibility of dysarthric speech (Bent et al., 2016; McAuliffe et al., 2013). A positive relationship between receptive

[1]Although there is no control group in the present study to confirm that the catalyst for improved intelligibility was the familiarization task, there is a preponderance of data to support this assumption (Borrie et al., 2013; Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, et al., 2012; Borrie & Schäfer, 2015; Lansford et al., 2016; Liss et al., 2002).

[2]Rhythm perception scores were significantly greater in the musician group relative to the nonmusician group.

vocabulary and speech recognition has also been identified with speech in noise (Tamati, Gilbert, & Pisoni, 2013) and unfamiliar accents (Banks, Gowen, Munro, & Adank, 2015). Bent et al. (2016) speculate that larger vocabularies may be the result of experience with a greater variety of exemplars for each lexical item and that this experience may aid in word identification, even when the spoken realizations deviate from typical norms.

The processing advantage associated with superior vocabulary knowledge was not evident following familiarization with dysarthric speech—receptive vocabulary scores failed to predict intelligibility improvement scores. Thus, although it appears that receptive vocabulary aids in initial processing of dysarthric speech, this knowledge does not seem to offer any benefit for learning to adapt to the disordered signal. This finding supports the idea that the magnitude of intelligibility improvement following familiarization with dysarthric speech is guided by expertise in rhythm sensitivity rather than vocabulary knowledge. To our knowledge, the relationship between vocabulary size and perceptual learning of dysarthric speech has not been formerly examined. However, consistent with the findings of the current study, Banks et al. (2015) found no relationship between receptive vocabularies and perceptual learning for young, healthy listeners adapting to a novel accent. In contrast, Janse and Adank (2012) observed that the rate at which older listeners adapted to a novel accent was influenced by vocabulary knowledge—higher receptive vocabularies correlated with greater improvements in recognition accuracy over blocks of trials. In an attempt to account for the discrepancy, Banks et al. (2015) speculate that vocabulary knowledge may become an important strategy for older adults to use when adapting to unfamiliar speech as a subconscious effort to compensate for age-associated cognitive declines. It, therefore, seems essential that if we are to fully understand the mechanisms that underlie recognition and learning of speech in challenging conditions, including that of dysarthria, processing models must consider not only the nature of the signal degradation but also the unique profile of the listener.

To reduce the idea that putative differences in intelligence between participants could influence the data in one direction or another, we recruited participants engaged in university-level education. We also used semantically anomalous testing phrases to minimize utilization of higher level functions. Future work will, however, more directly examine cognitive factors as a potential source of variability in processing dysrhythmic speech signal. Indeed, processing both music and speech signals demands higher level cognitive functions, such as working memory and attention (Besson, Chobert, & Marie, 2011), and musical training has been shown to enhance these functions (e.g., George & Coch, 2011). Thus, individuals with superior rhythm perception skills, conceivably associated with musical training, may also possess superior cognitive skills, which may, in turn, support perception and learning of dysrhythmic speech.

Here, we used speech stimuli produced by a single speaker whose speech features represented the cardinal features of ataxic dysarthria. Although this was designed for a high level of experimental control, namely, reducing a potential source of variability that may interact with intelligibility outcomes, the potential limitation of using a single speaker should be raised. Perhaps more important to acknowledge, however, is that the results of this study may be limited to perception and learning of dysarthria with the classic ataxic presentation. However, listeners have been shown to utilize corrupted rhythm cues to segment speech regardless of dysarthria type and severity (Liss et al., 2000, 2009), which suggests that the results observed here with one type of dysarthria may hold true for other types and/or severities. It is, however, plausible that the results of this study may be limited to stress-based languages with metrical regularities similar to that of English. Therefore, whether expertise in rhythm perception provides processing benefits with varying patterns of rhythm abnormality across languages with different rhythmic structures provides important future directions for this work.

While largely of theoretical import, the current study offers some interesting clinical considerations. There exists abundant empirical evidence to suggest that listener-targeted perceptual learning paradigms may be a viable clinical tool for reducing the intelligibility burden of dysarthria (Borrie et al., 2013; Borrie, McAuliffe, Liss, et al., 2012; Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie & Schäfer, 2015; Lansford et al., 2016). Here, we extend this work, proposing the novel hypothesis that individuals with expertise in rhythm sensitivity may be best suited for such perceptual learning interventions. Or viewed from another vantage point, that prior rhythm training may elevate perceptual learning intervention outcomes. Although not currently established in the clinical domain, the notion of listener-targeted remediation in dysarthria should not be underestimated. Treatment that focuses on the primary communication partners of the patient (e.g., caregivers, family members, and friends) may be a valuable adjunct to the traditional patient-based interventions or a stand-alone alternative when a patient is unable to improve speech behaviorally (see Borrie, McAuliffe, & Liss, 2012, for a review). The knowledge that rhythm perception skills advantage a listener to adapt to dysarthric speech introduces novel lines of investigation in this area.

Conclusion

In sum, the current study offers new insight into the relationship between rhythm perception and processing of a dysrhythmic speech signal. Here, we see that expertise in rhythm perception affords listeners a perceptual advantage for processing dysarthric speech, but a brief opportunity to learn something about the pathological signal is required for the advantage to be realized. On the other hand, while receptive vocabulary scores offered listeners a perceptual benefit for initial processing, they did not aid in improved processing in subsequent encounters. Taken together, these findings add support for the value of rhythm perception in recognizing spoken language in challenging listening conditions, as well as highlight the need for processing models that account for both the speaker and the listener.

Acknowledgments

We gratefully acknowledge research assistants Ashely Hurst and Michelle Parker in the Human Interaction Lab at Utah State University for their assistance with data collection and analysis.

References

Banks, B., Gowen, E., Munro, K. J., & Adank, P. (2015). Cognitive predictors of perceptual adaptation to accented speech. *The Journal of the Acoustical Society of America, 137,* 2015–2024. doi.org/10.1121/1.4916265

Bent, T., Baese-Berk, M., Borrie, S. A., & McKee, M. (2016). Individual differences in the perception of unfamiliar regional, nonnative, and disordered speech varieties. *The Journal of the Acoustical Society of America, 140,* 3775–3886.

Besson, M., Chobert, J., & Marie, C. (2011). Transfer of training between music and speech: Common processing, attention, and memory. *Frontiers in Psychology, 2,* 94. doi:10.3389/fpsyg.2011.00094

Borrie, S. A. (2015). Visual speech information: A help or hindrance in perceptual processing of dysarthric speech. *The Journal of the Acoustical Society of America, 137,* 1473–1480. doi.org/10.1121/1.4913770

Borrie, S. A., & Liss, J. M. (2014). Rhythm as a coordinating device: Entrainment with disordered speech. *Journal of Speech, Language, and Hearing Research, 57,* 815–824. doi:10.1044/2014_JSLHR-S-13-0149

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55,* 290–305. doi:10.1044/1092-4388(2011/10-0349)

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27,* 1039–1055. doi:10.1080/01690965.2011.610596

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. (2012). A follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech. *The Journal of the Acoustical Society of America, 132,* EL102–EL108. doi.org/10.1121/1.4736952

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. (2013). The role of linguistic and indexical information in improved recognition of dysarthric speech. *The Journal of the Acoustical Society of America, 133,* 474–482. doi.org/10.1121/1.4770239

Borrie, S. A., & Schäfer, M. C. M. (2015). The role of somatosensory information in speech perception: Imitation improves recognition of disordered speech. *Journal of Speech, Language, and Hearing Research, 58,* 1708–1716.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software, 34*(2), 1–24.

Cummins, F. (2013). Joint speech: The missing link between speech and music? *PERCEPTA–Revista de Cognicao Musical, 1*(1), 17.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969). Differential diagnostic patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246–269. doi:10.1044/jshr.1202.246

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). *Motor speech disorders* (3rd ed.). Philadelphia, PA: W. B. Saunders.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test*. Circle Pines, MN: AGS.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience, 16,* 1010–1021. doi.org/10.1162/0898929041502706

Fuller, C. D., Galvin, J. J., III, Maat, B., Free, R. H., & Baskent, D. (2014). The musician effect: Does it persist under degraded pitch conditions of cochlear implant simulations? *Frontiers in Neuroscience, 8,* 179. doi:10.3389/fnins.2016.00075

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 23,* 9240–9245.

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia, 49,* 1083–1094. doi:10.1016/j.neuropsychologia.2011.02.001

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11,* 466–472. doi:10.1016/j.tics.2007.08.008

Janse, E., & Adank, P. (2012). Predicting foreign-accent adaptation in older adults. *The Quarterly Journal of Experimental Psychology, 65,* 1563–1585. doi:10.1080/17470218.2012.658822

Jeon, J. Y., & Fricke, F. R. (1997). Duration of perceived and performed sounds. *Psychology of Music, 25,* 70–83. doi:10.1177/0305735697251006

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing, 23,* 2–40.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11,* 599–605. doi:10.1038/nrn2882

Kraus, N., Skoe, E., Parbery-Clark, A., & Ashley, R. (2009). Experience-induced malleability in neural encoding of pitch, timbre, and timing. *Annals of the New York Academy of Sciences, 1169*(1), 543–557. doi:10.1111/j.1749-6632.2009.04549.x

Lansford, K. L., Borrie, S. A., & Bystricky, L. (2016). Use of crowdsourcing to assess the ecological validity of perceptual training paradigms in dysarthria. *American Journal of Speech-Language Pathology, 25,* 233–239.

Lansford, K. L., & Liss, J. M. (2014a). Vowel acoustics in dysarthria: Mapping to perception. *Journal of Speech, Language, and Hearing Research, 57,* 68–80. doi:10.1044/1092-4388(2013/12-0263)

Lansford, K. L., & Liss, J. M. (2014b). Vowel acoustics in dysarthria: Speech disorder diagnosis and classification. *Journal of Speech, Language and Hearing Research, 57,* 57–67. doi:10.1044/1092-4388(2013/12-0262)

Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America, 124,* 3235–3248. doi.org/10.1121/1.2990713

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112,* 3022–3030. doi.org/10.1121/1.1515793

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. W. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 107,* 3415–3424. doi.org/10.1121/1.429412

Liss, J. M., Utianski, R., & Lansford, K. L. (2013). Crosslinguistic application of English-centric rhythm descriptors in motor speech disorders. *Folia Phoniatrica et Logopaedica, 65,* 3–19. doi:10.1159/000350030

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., & Caviness, J. N. (2009). Quantifying speech rhythm deficits in the dysarthrias. *Journal of Speech, Language, and Hearing Research, 52,* 1334–1352. doi:10.1044/1092-4388 (2009/08-0208)

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience, 23,* 2701–2715. doi:10.1162/jocn.2010.21585

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134,* 477–500.

McAuliffe, M. J., Gibson, E. M. R., Kerr, S. E., Anderson, T., & LaShell, P. J. (2013). Vocabulary influences older and younger listeners' processing of dysarthric speech. *The Journal of the Acoustical Society of America, 134,* 1358–1368. doi.org/10.1121/1.4812764

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 621–638.

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research, 219*(1–2), 36–47. doi:10.1016/j.heares.2006.05.004

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *The Journal of Neuroscience, 29,* 14100–14107. doi:10.1523/JNEUROSCI.3256-09.2009

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5,* 688–694. doi:10.1038/nn871

Simas, A. B., Barreto-Souza, W., & Rocha, A. V. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis, 54,* 348–366. doi:10.1016/J.CSDA.2009.08.017

Slater, J., & Kraus, N. (2015). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17,* 79–87. doi:10.1007/s10339-015-0740-7

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech and Hearing Research, 32,* 912–920.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*(1), 54–71.

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd, G., Jr., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. *Scientific Reports, 5,* 11628. doi:10.1038/srep11628

Tamati, T. N., Gilbert, J. L., & Pisoni, D. B. (2013). Some factors underlying individual differences in speech recognition on PRESTO: A first report. *Journal of the American Academy of Audiology, 24,* 616–634. doi:10.3766/jaaa.24.7.10

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The Musical Ear Test, a new reliable test for measuring musical competence. *Learning and Individual Differences, 20,* 188–196. doi:10.1016/j.lindif.2010.02.004

Wong, P. & Perrachione, T. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics, 28,* 565–585. doi.org/10.1017/S0142716407070312

Zendel, B. R., Tremblay, C. D., Belleville, S., & Peretz, I. (2015). The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *Journal of Cognitive Neuroscience, 27,* 1044–1059. doi:10.1162/jocn_a_00758

Zuk, J., Ozernov-Palchik, O., Kim, H., Lakshminarayanan, K., Gabrieli, J. D., Tallal, P., & Gaab, N. (2013). Enhanced syllable discrimination thresholds in musicians. *PLoS One, 8,* e80546. doi:10.1371/journal.pone.0080546