Understanding dysrhythmic speech: When rhythm does not matter and learning does not happen

Stephanie A. Borrie, Kaitlin L. Lansford, and Tyson S. Barrett

Citation: The Journal of the Acoustical Society of America **143**, EL379 (2018); doi: 10.1121/1.5037620
View online: https://doi.org/10.1121/1.5037620
View Table of Contents: http://asa.scitation.org/toc/jas/143/5
Published by the Acoustical Society of America

Understanding dysrhythmic speech: When rhythm does not matter and learning does not happen

Stephanie A. Borrie[a)]
Department of Communicative Disorders and Deaf Education, Utah State University,
Logan, Utah 84322, USA
stephanie.borrie@usu.edu

Kaitlin L. Lansford
School of Communication Sciences and Disorders, Florida State University, Tallahassee,
Florida 32306, USA
klansford@fsu.edu

Tyson S. Barrett
Department of Kinesiology and Health Sciences, Utah State University, Logan, Utah
84322, USA
tyson.barrett@usu.edu

Abstract: A positive relationship between rhythm perception and improved understanding of a naturally dysrhythmic speech signal, ataxic dysarthria, has been previously reported [Borrie, Lansford, and Barrett. (2017). J. Speech Lang. Hear. Res. **60**, 3110–3117]. The current follow-on investigation suggests that this relationship depends on the nature of the dysrhythmia. When the corrupted rhythm cues are relatively predictable, affording some learnable acoustic regularity, the relationship is replicated. However, this relationship is nonexistent, along with any intelligibility improvements, when the corrupted rhythm cues are unpredictable. Findings highlight a key role for rhythm perception and distributional regularities in adaptation to dysrhythmic speech.

[DDO]
Date Received: March 5, 2018 **Date Accepted:** April 24, 2018

1. Introduction

Rhythm cues, or suprasegmental information, have been shown to be important for deciphering speech in adverse listening conditions (Mattys *et al.*, 2005). This statement appears to hold true even when the rhythm cues are corrupted, as is the case of the naturally dysrhythmic speech signal of dysarthria (Liss *et al.*, 2013). Furthermore, listeners demonstrate increased reliance on rhythm cues following familiarization with dysarthric speech (Borrie *et al.*, 2012). According to theoretical models of perceptual learning/adaptation, listeners build linguistic generative models for novel talkers based on knowledge about the distribution of acoustic cues associated with each linguistic category (Kleinschmidt and Jaeger, 2015). This cue-to-category mapping of incoming speech allows the listener to more successfully navigate that same signal in subsequent encounters. Thus, it appears that experience with a talker with dysarthria allows listeners an opportunity to learn useful structure, statistical contingencies, in the corrupted rhythm cues. To date, talker-specific perceptual learning of dysarthric speech has been evidenced with hypokinetic, spastic, and ataxic dysarthria (e.g., Borrie *et al.*, 2012; Borrie and Schafer, 2015; Liss *et al.*, 2002). It is nontrivial to note that while the rhythmic degradation of these three types of dysarthria is abnormal (e.g., reduced or equal/even stress), it is still, in most cases, relatively predictable, affording some level of distributional regularity.

Motivated by the recognized import of rhythm cues in deciphering degraded speech, we recently examined the relationship between rhythm perception and processing of a naturally dysrhythmic speech signal, ataxic dysarthria (Borrie *et al.*, 2017). Results revealed expertise in rhythm perception, operationalized as superior ability to perceive rhythm cues in the musical domain, provided no advantage with initial intelligibility of ataxic dysarthria but afforded a significant advantage with perceptual learning of the dysrhythmic speech signal, reflected in larger intelligibility

---

[a)] Author to whom correspondence should be addressed.

improvements following a familiarization experience. We speculated that listeners with expertise in rhythm perception are better able to exploit the familiarization experience, identifying underlying cue distributions in the aberrant rhythm patterns. Theoretically, this relationship between rhythm perception and adaptation to dysrhythmic speech would not be realized if the suprasegmental cues were irregular and unpredictable. Indeed, in such cases, statistical contingencies in speech rhythm would simply not exist.

The current study investigated whether the relationship between rhythm perception and processing of dysarthric speech, observed in our previous study with ataxic dysarthria, holds for other forms of dysarthria, even when the rhythm cues afforded in the degraded signal are not only corrupted but largely unstable, as is the case of hyperkinetic dysarthria resulting from Huntington's disease (Duffy, 2015). We addressed the following two research questions: (a) does expertise in rhythm perception provide an advantage for perception (initial intelligibility) and learning (intelligibility improvement) of the irregular and unpredictable degraded speech classified as *hyperkinetic* dysarthria, and (b) does expertise in rhythm perception provide an advantage for perception and learning of the regular and predictable degraded speech classified as *hypokinetic* dysarthria? Our first hypothesis was that expertise in rhythm perception would not predict initial intelligibility of either hyperkinetic or hypokinetic dysarthria, mirroring our initial findings with ataxic dysarthria. Our second and key hypothesis for this study was that expertise in rhythm perception would predict talker-specific intelligibility improvements when the speech degradation was relatively stable and predictable (i.e., hypokinetic dysarthria) but not when it was unstable and unpredictable (i.e., hyperkinetic dysarthria).

2. Method

2.1 Participants

A total of 98 participants (55 men and 43 women) aged 22 to 62 years old ($M = 36.13$), participated in the experiment. All participants were native speakers of American English and reported no history of speech, language, or hearing impairment, and no previous experience with talkers with dysarthria. Participants were recruited using the crowdsourcing website, Amazon Mechanical Turk (MTurk; http://www.mturk.com).[1] We employed MTurk setup options regarding participant prerequisites, limiting participation to individuals with a previous approval rate of greater than or equal to 99% and a confirmed status of being a U.S resident. This data collection method for perception of dysarthric speech has been shown to produce comparable results to data generated in the lab (Lansford *et al.*, 2016).

2.2 Speech stimuli

Speech stimuli for the speech perception and learning test consisted of audio recordings of testing material (semantically anomalous phrases) and familiarization material (passage reading), elicited from two male native speakers of American English with a clinical diagnosis of dysarthria. The testing material, used for the pretest and post-test speech sets, consisted of 80 syntactically plausible but semantically anomalous phrases (e.g., *amend estate approach* and *had eaten junk and train*). All phrases contained six syllables with alternating metrical stress and ranged in length from three to five words. The passage reading was an adapted version of the standard "Grandfather Passage" (Darley *et al.*, 1975). It consisted of 35 phrases, ranging in length from 3 to 12 words. The audio recording of the passage reading, paired with orthographic transcription, served as linguistically rich familiarization speech stimuli.

The two talkers who produced the speech stimuli in this study presented with different types of dysarthria—one with a moderate hypokinetic dysarthria secondary to Parkinson's disease and the other, a moderate hyperkinetic dysarthria secondary to Huntington's disease. Medical etiology was confirmed by medical records, and dysarthria diagnoses and severity ratings were made by three certified speech-language pathologists with expertise in assessment and differential diagnosis of motor speech disorders. The speech of the talker with Parkinson's disease represented a classic hypokinetic dysarthria, characterized perceptually by reduced stress, monotone, monoloudness, short rushes of speech, imprecise articulation, and a breathy voice. The speech of the talker with Huntington's disease was characterized perceptually by prolonged intervals and phonemes, variable rate, inappropriate silences, irregular articulatory breakdowns, and intermittent hypernasality. These presenting perceptual features and severity ratings of these two talkers have been acoustically validated in earlier studies (Lansford and Liss, 2014a,b; Liss *et al.*, 2009). Thus, while both talkers with dysarthria presented with dysrhythmic speech, the rhythm deficits of the talker with hypokinetic speech were largely predictable (i.e., reduced but stable) whereas the rhythm deficits of the talker with hyperkinetic speech were largely unpredictable—as is typically the case (Duffy, 2015).

2.3 Procedure

A description of the study, time commitment, and remuneration ($5 + $3 bonus[1]) was posted to MTurk. Interested and qualifying individuals were directed to a web page, loaded with the study procedure hosted on a secure university-based web server. In order to partake in the study, individuals were required to read through the institutional review board approved consent form. By clicking Agree, individuals indicated that they had read and understood the information in the consent form and voluntarily agreed to participate in the study. Participants were then required to complete a brief questionnaire regarding demographic information and questions related to inclusion/exclusion criteria. Upon completion of the questionnaire, participants were advanced to the two experimental tasks.

The first experimental task consisted of a rhythm perception test, the rhythm subtest of the Musical Ear Test (MET) (Wallentin *et al.*, 2010). For this test, listeners were presented with pairs of rhythmical phrases played with a wood block and were required to judge whether the phrase pairs were the same or different. The rhythm sequences within each phrase contain 4–11 wood block beats and had a duration of one measure played at 100 beats per minute (for more detailed information, see Wallentin *et al.*, 2010). Participants were given two practice trails before beginning the 52-item test. The rhythm perception test took approximately 10 min to complete.

The second experimental task consisted of a speech perception and learning test. For this portion of the experiment, participants were randomly assigned to one of two experimental speech conditions, *hyperkinetic-unpredictable* or *hypokinetic-predictable*. Using an identical protocol to a number of our previous studies, participants engaged in a three-phase testing paradigm involving pretest, familiarization, and posttest phases (e.g., Borrie *et al.*, 2017). During the pretest phase, participants were told that they would be presented with short phrases that would be difficult to understand because they would be produced by a person with a speech disorder. They were also told that phrases contained real English words but would not make sense. Phrases were presented one at a time, and following each presentation, participants were instructed to use the keyboard to type out exactly what they thought the talker was saying. Participants were strongly encouraged to make a guess at any words they did not recognize. Once they had finished typing their response, they were instructed to press the return key to move on to the next phrase. Immediately following the pretest, participants partook in the familiarization phase where they listened to the passage reading stimuli and followed along with an orthographic transcription. After this, participants completed the posttest phase, which was identical to the pretest phase but involved novel testing stimuli (i.e., the posttest set). The self-paced speech perception and learning test took, on average, 35 min to complete.

2.4 Data analysis

Three scores were obtained for each of the 98 participants: (a) rhythm perception score, (b) initial intelligibility score, and (c) intelligibility improvement score. Rhythm perception scores were calculated as the percentage of correct responses on the rhythm subtest of the MET. Initial intelligibility and intelligibility improvements score computations were based on listener transcripts of the pretest and post-test speech sets during the speech perception and learning test. Transcripts were analyzed for words correct, using previously established scoring criteria for the semantically anomalous phrases (Liss *et al.*, 1998; Borrie *et al.*, 2012) and an in-house computer scoring program. The program automatically scored words as correct if they matched the intended target exactly or differed only by tense (-ed) or plurality (-s). Homophones and obvious spelling errors were also scored as correct. A percentage words correct (PWC) score was tabulated for each listener participant for each speech set. 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. Twenty percent of the transcripts were randomly selected and reanalyzed by a human (trained research assistant) to examine reliability for coding words correct. Discrepancies between the computer and human revealed a high agreement with Pearson correlation r score above 0.98.

We used independent-samples *t*-tests to examine between condition differences in rhythm perception scores and paired-samples *t*-tests to assess within condition differences between pretest and posttest intelligibility scores. We also assessed the Pearson correlation between initial intelligibility scores and intelligibility improvement scores. Then, as per our previous study, we used beta regression models to test the predictive relationships between the intelligibility variables, the rhythm perception scores, and the speech condition (i.e., hyperkinetic-unpredictable or hypokinetic-predictable). Beta regression is particularly useful when the outcome is bounded, such as proportions of correct responses (i.e., the initial intelligibility and the improvement in intelligibility). It has two submodels: the first estimates the effects of the predictors on the mean (location submodel) and the other estimates those effects on the variability (dispersion submodel). The location submodel is much like linear regression but adjusts for the bounded outcomes and the non-normal distribution. The dispersion submodel provides information regarding how the predictors influence the variability in the outcome. Note, while the dispersion submodel forms an important part of beta regression, it is not critical for the question of interest. As we describe in our previous study, beta regression provides more accurate information than other common approaches (for a review on beta regression in the social sciences, see Smithson and Verkuilen, 2006).

The question of interest targeted whether the effect of rhythm perception scores on either initial intelligibility or intelligibility improvement (i.e., learning) depends on the type of dysrhythmic speech. Two beta regression models, one for each intelligibility outcome, were used to test the interaction between rhythm perception scores and speech condition. The interaction informs whether the relationship between rhythm perception and the outcomes differs based on the speech condition, thereby testing the question of interest. In contrast to our previous study, the best fitting distribution was logit, thereby producing results similar to logistic regression, albeit in relation to proportion of correct responses instead of a binary outcome.

3. Results

3.1 Group scores

The rhythm perception data, reported as percentages, for the hyperkinetic-unpredictable and hypokinetic-predictable speech conditions are presented in Fig. 1. Of note is the likeness between both the range and average rhythm perception scores of the participants in the hyperkinetic-unpredictable ($M = 73.31$) and hypokinetic-

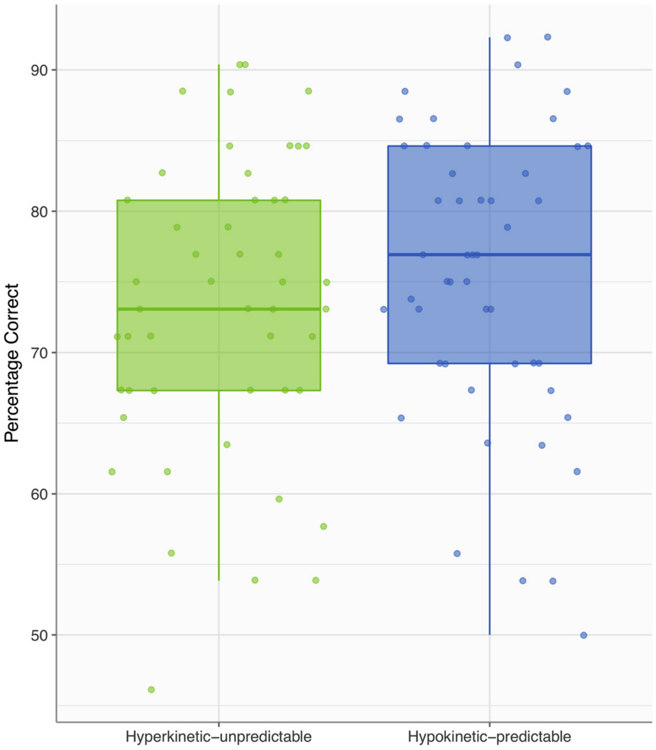

Fig. 1. (Color online) Average rhythm perception scores for listeners in the hyperkinetic-unpredictable and hypokinetic-predictable speech conditions.

predictable condition ($M = 75.41$). An independent-samples *t*-test revealed no significant difference in the rhythm perception scores between the two speech conditions.

Intelligibility data, reported as percentages, for hyperkinetic-unpredictable or hypokinetic-predictable speech conditions are presented in Fig. 2. Of note is the different pattern of learning, illustrated by the gradient of the line connecting the participants pretest and posttest scores. Results of a paired-samples *t*-test revealed that participants in the hypokinetic-predictable speech condition, on average, showed significant learning (intelligibility improvement) from pretest ($M = 50.49$, $\mathrm{SD} = 8.89$) to posttest ($M = 60.84$, $\mathrm{SD} = 8.95$), $t(48) = 20.38$, $p < 0.001$, whereas participants in the hyperkinetic-unpredictable speech condition did not show significant intelligibility improvement from pretest ($M = 52.33$, $\mathrm{SD} = 8.19$) to posttest ($M = 52.54$, $\mathrm{SD} = 8.09$), $t(48) = 0.37$, $p = 0.71$. Results of a correlation analysis demonstrated that, for both speech conditions, initial intelligibility scores and intelligibility improvement scores were unrelated.

3.2 Prediction analysis

The beta regression results are presented in Table 1, showing the interaction effect and the main effects of each model. For the location submodel (i.e., the submodel concerned with changes in the mean level of correct responses or improvement in correct responses), there was no significant interaction between rhythm perception scores and speech condition on initial intelligibility scores ($p = 0.769$). Thus, regardless of speech condition, rhythm perception scores did not predict initial intelligibility of a dysrhythmic speech signal. However, there was a significant interaction between rhythm perception scores and speech condition on intelligibility improvement scores ($p = 0.015$) in the location submodel. Specifically, rhythm perception scores predicted intelligibility improvement only for the hypokinetic-predictable condition whereas there was no effect in the hyperkinetic-unpredictable condition (Fig. 2 illustrates this finding). The model results suggest that, for a 10-percentage point increase in rhythm perception scores, there was, on average, an

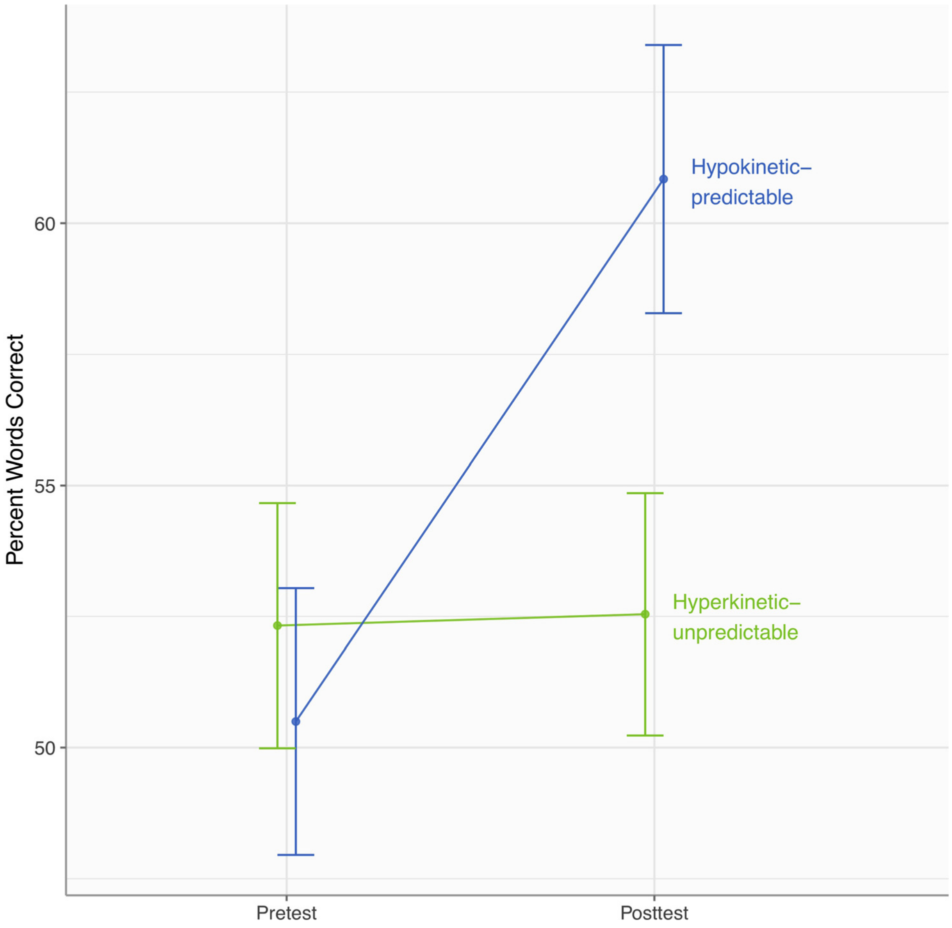

Fig. 2. (Color online) Average pretest (initial intelligibility) and posttest intelligibility scores for listeners in the hyperkinetic-unpredictable and hypokinetic-predictable speech conditions. Intelligibility improvement is reflected in the slope of the connecting line.

Table 1. Results of the beta regression models. Note: B is the estimate of the effect of the variable in log-odds units (location submodel) and in the outcome's units (dispersion submodel). SE is the standard error of B. The location submodel had a logit link, and the dispersion submodels had an identity link.

	Initial Intelligibility			Intelligibility Improvement		
	B	SE	*P*-Value	B	SE	*P*-Value
Location Submodel						
Rhythm Perception	0.004	0.010	0.663	0.007	0.012	0.567
Condition	0.180	1.149	0.876	−0.817	0.981	0.405
Rhythm Perception × Condition	−0.004	0.015	0.769	0.033	0.014	0.015
Dispersion Submodel						
Rhythm Perception	0.031	0.019	0.108	−0.011	0.019	0.539
Condition	1.478	2.021	0.465	3.300	1.995	0.098
Rhythm Perception × Condition	−0.028	0.027	0.301	−0.032	0.027	0.227

increase of 48% in the odds of improvement in intelligibility for the hypokinetic-predictable condition.[2] For the dispersion submodel, there were no interactions between rhythm perception and speech condition on either the variability in initial intelligibility or intelligibility improvement. Notably, the Psuedo-R squared for the initial intelligibility was near zero (0.003) but was large for the intelligibility improvement (0.449).

4. Discussion

The current study expands on an earlier investigation on the role of rhythm perception in perception and learning of dysrhythmic speech (Borrie *et al.*, 2017). Consistent with our earlier report with ataxic dysarthria, expertise in rhythm perception did not predict initial intelligibility of either hyperkinetic or hypokinetic dysarthria. This finding, realized now for three different types of dysarthric speech, is in contrast to a body of literature reporting a relationship between rhythm expertise and deciphering spoken language in others types of challenging listening conditions. Slater and Kraus (2015), for example, demonstrated that superior rhythm perception skills, also measured by the MET, predicted superior intelligibility of speech in noise. A critical difference, however, is that while speech in noise is characterized by disruption to a number of segmental features, suprasegmental cues remain relatively intact (e.g., Parikh and Loizu, 2005). Dysarthria, on the other hand, is characterized by both segmental and suprasegmental degradation. Thus, it appears that when the suprasegmental rhythm cues are degraded, expert rhythm perception abilities do not advantage listeners in their initial attempts to decipher the disordered speech.

Expert rhythm perception abilities do, however, advantage listeners in their ability to adapt to the disordered speech. We replicated our earlier findings with ataxic dysarthria, demonstrating a positive relationship between rhythm perception scores and improved understanding of hypokinetic dysarthria following a familiarization experience. This relationship was not observed with the irregular and unpredictable dysrhythmia of hyperkinetic dysarthria. Thus, the advantage afforded by skills in rhythm perception is only apparent when the rhythm cues which, while degraded, offered some level of regularity. Taken together, the findings suggest that listeners who are better able to process rhythm information are also better able to navigate a dysrhythmic speech signal, identifying and learning distributional regularities in the aberrant rhythm patterns for improved processing of the signal during subsequent encounters. The processing advantage is simply not realized when there are no distributional regularities to be learned.

Indeed, even the seemingly robust phenomenon of perceptual learning is not realized in the case of the unpredictable dysrhythmia of hyperkinetic dysarthria. Listeners do not learn to better understand this type of degraded speech following a familiarization experience. While in contrast to our previous work in perceptual learning of dysarthric speech (e.g., Borrie *et al.*, 2012; Borrie and Schafer, 2015; Lansford *et al.*, 2016), the finding highlights the critical role of statistical contingencies in mapping speech input to linguistic categories, generating generative models for improved understanding of novel talkers/speech patterns (Kleinschmidt and Jaeger, 2015). Our earlier work has demonstrated that listeners rely more heavily of rhythm cues following familiarization with the degraded acoustic signal, reflected in lexical segmentation decisions based largely on metrical stress patterns (Borrie *et al.*, 2012). The same listeners are also better able to process segmental cues in subsequent encounters with the speech signal. With hyperkinetic dysarthria, even the segmental acoustic cues are irregular and unpredictable (e.g., irregular consonant and vowel distortion). Thus, we show that perceptual learning will not transpire when there are no learnable regularities available in the speech signal.

In sum, the present investigation extends on the findings of our earlier report, adding critical insight and understanding of the role of rhythm perception in learning of dysrhythmic speech. Specifically, results suggest that the relationship between expertise in rhythm perception and perceptual learning of a dysrhythmic speech signal is dependent on the nature of the dysrhythmia. When the perceptual speech rhythm features are relatively stable across the talker's utterances, affording some level of learnable acoustic regularity, as in the case of hypokinetic dysarthria, a positive relationship exists. However, this relationship disappears entirely when the perceptual speech rhythm features are unstable and unpredictable, as in the case of hyperkinetic dysarthria. Indeed, in this case, no intelligibility improvements were observed. While the limitation of using speech stimuli from single talkers may be raised, this was intentionally done for a high level of experimental control (see Borrie *et al.*, 2017, for a discussion). Research going forward will further explicate how acoustic cue degradation, both type and severity, impact statistical learning of dysrhythmic speech.

Acknowledgments

This research was supported by the National Institute of Deafness and Other Communication Disorders, National Institutes of Health Grant No. R21DC016084, awarded to S.A.B.

References and links

[1] Participants were informed that they would be paid $5 for completing the experiment but that they would receive a $3 bonus payment if there was evidence that they had fully engaged in the tasks. Two participants did not receive the bonus payment and thus their data were not included in the study.

[2] This was calculated using the estimate for both the main effect of rhythm perception and the interaction. That is, we calculated $e^{(0.007+0.033)*10}$, giving us the odds of a 10 unit increase in rhythm perception scores.

Borrie, S. A., Lansford, K. L., and Barrett, T. S. (**2017**). "Rhythm perception and its role in recognition and learning of dysrhythmic speech," J. Speech Lang. Hear. Res. **60**, 561–570.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., and Anderson, T. (**2012**). "Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech," Lang. Cogn. Process. **27**, 1039–1055.

Borrie, S. A., and Schäfer, M. C. M. (**2015**). "The role of somatosensory information in speech perception: Imitation improves recognition of disordered speech," J. Speech Lang. Hear. Res. **58**, 1708–1716.

Darley, F. L., Aronson, A. E., and Brown, J. R. (**1975**). *Motor Speech Disorders* (W. B. Saunders Company, Philadelphia, PA).

Duffy, J. R. (**2015**). *Motor Speech Disorders: Substrates, Differential Diagnosis, and Management* (Elsevier Mosby, St. Louis, MO).

Kleinschmidt, D. F., and Jaeger, T. F. (**2015**). "Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel," Psychol. Rev. **122**, 148–203.

Lansford, K. L., Borrie, S. A., and Bystricky, L. (**2016**). "Use of crowdsourcing to assess the ecological validity of perceptual training paradigms in dysarthria," Am. J. Speech Lang. Pathol. **25**, 233–239.

Lansford, K. L., and Liss, J. M. (**2014a**). "Vowel acoustics in dysarthria: Speech disorder diagnosis and classification," J. Speech Lang. Hear. Res. **57**, 57–67.

Lansford, K. L., and Liss, J. M. (**2014b**). "Vowel acoustics in dysarthria: Mapping to perception," J. Speech Lang. Hear. Res. **57**, 68–80.

Liss, J. M., Spitzer, S. M., Caviness, J. N., and Adler, C. (**2002**). "The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria," J. Acoust. Soc. Am. **112**, 3022–3030.

Liss, J. M., Spitzer, S., Caviness, J. N., Adler, C., and Edwards, B. (**1998**). "Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech," J. Acoust. Soc. Am. **104**, 2457–2466.

Liss, J. M., Utianski, R., and Lansford, K. L. (**2013**). "Crosslinguistic application of English-centric rhythm descriptors in motor speech disorders," Folia Phoniatr. Logop. **65**, 3–19.

Liss, J. M., White, L., Mattys, S. L., Lansford, K., Lotto, A. J., Spitzer, S. M., and Caviness, J. N. (**2009**). "Quantifying speech rhythm deficits in the dysarthrias," J. Speech Lang. Hear. Res. **52**, 1334–1352.

Mattys, S. L., White, L., and Melhorn, J. F. (**2005**). "Integration of multiple speech segmentation cues: A hierarchical framework," J. Exp. Psychol. Gen. **134**, 477–500.

Parikh, G., and Loizou, P. (**2005**). "The influence of noise on vowel and consonant cues," J. Acoust. Soc. Am. **118**, 3874–3888.

Slater, J., and Kraus, N. (**2015**). "The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians," Cog Process. **170**, 79–87.

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

Wallentin, M., Nielsen, A. H., Friis-Olivariusa, M., Vuust, C., and Vuust, P. (**2010**). "The Musical Ear Test, a new reliable test for measuring musical competence," Learn. Individ. Differ. **20**, 188–196.