Title: Understanding speech through the lens of rhythmic expertise: Speech-in-noise perception tracks with sequence-based drumming but not beat synchronisation

Running title: Speech-in-noise and rhythm skills

Author names: Jessica Slater,[1-2] Nina Kraus,[1-5] Kali Woodruff Carr,[1-2] Adam Tierney,[1-2†] Andrea Azem,[1-2] and Richard Ashley[6-7]

Author Affiliations:
[1]*Auditory Neuroscience Laboratory (www.brainvolts.northwestern.edu)*
[2]*Department of Communication Sciences*
[3]*Institute for Neuroscience*
[4]*Department of Neurobiology*
[5]*Department of Otolaryngology, Northwestern University, Evanston, Illinois USA*
[6]*School of Music, Northwestern University, Evanston, IL, USA*
[7]*Program of Cognitive Science, Northwestern University, Evanston, IL, USA*
†Current address: Department of Psychological Sciences, University of London Birkbeck, London, United Kingdom

Corresponding author:
[**]Nina Kraus, Ph.D.
Northwestern University
2240 Campus Drive
Evanston, IL 60208
USA
Email: nkraus@northwestern.edu
Phone: 847-491-3181
Fax: 847-491-2523

Title: Understanding speech through the lens of rhythmic expertise: Speech-in-noise perception tracks with sequence-based drumming but not beat synchronisation

Abstract

Rhythm guides speech perception, especially in noise. We recently revealed that percussionists outperform non-musicians in speech-in-noise perception, with better speech-in-noise perception associated with better rhythm discrimination across a range of rhythmic expertise. Here, we consider rhythm production skills, specifically drumming to a beat (metronome or music) and to rhythm sequences (metrical or jittered rhythms), as well as speech-in-noise perception in adult percussionists and non-musicians. Given the absence of a regular beat in speech, we hypothesise that processing of rhythm sequences is more important for speech-in-noise perception than the ability to entrain to a regular beat. Consistent with our hypotheses, we find that the sequence-based drumming measures predict speech-in-noise perception, above and beyond demographic factors, hearing thresholds and IQ, whereas the beat-based measures do not. Outcomes suggest rhythm sequences may help disambiguate speech under degraded listening conditions, extending theoretical considerations about speech rhythm to the everyday challenge of listening in noise.

Keywords: speech perception; music; rhythm; temporal processing

Introduction

Rhythm and timing cues facilitate speech perception in noise. For example, duration patterns help a listener segregate competing sound streams (Andreou, Kashino, & Chait, 2011; Shamma, Elhilali, & Micheyl, 2011), identify boundaries between words (Smith, Cutler, Butterfield, & Nimmo-Smith, 1989), and may bootstrap higher-level linguistic processing by providing cues about syntactic structure (Gordon et al., 2015). In a recent study we revealed that percussionists outperform non-musicians in the perception of speech in noise, and that better speech-in-noise perception is associated with better rhythm discrimination, across a range of musical expertise (Slater & Kraus, 2016). However, there are differences in the rhythmic characteristics of speech and music (see Patel, 2008 for review), and there is evidence for dissociable rhythm skills (for example, see Tierney & Kraus, 2015), supported by distinct underlying neural circuitry (Teki, Grube, Kumar, & Griffiths, 2011). Therefore it remains to be determined which specific aspects of rhythmic expertise may confer benefits for speech perception in noise.

Both speech and music contain patterns of durations or onsets, as well as "meter," the hierarchical organization of accented and unaccented elements into groups. However, musical meter is typically organised around a periodic pulse, or beat, whereas spoken language emerges as a flow of sequences that are governed by rules but not strictly constrained in time (Ding, Melloni, Zhang, Tian, & Poeppel, 2016; Liberman & Prince, 1977; Patel, 2008). The predictable structure of music is consistent with its role as a means of synchronization and coordination (for example, see Dalla Bella, Bialunska, & Sowinski, 2013), whereas the functional emphasis of language lies in semantic specificity (Cross, 1999). Although it has been proposed that isochronous timing intervals are present in speech (for example, Abercombie, 1967), attempts to demonstrate this empirically have been largely unsuccessful (Dauer, 1983; Lehiste, 1977; also see Patel, 2008 for some exceptions).

Here, we hypothesise that expertise with rhythm sequences transfers to speech perception and aids in the perception of speech in noise, whereas the ability to synchronise with a regular beat does not. We assessed beat-based skills (drumming with a metronome or the beat of music) and sequence-based skills (drumming with metrical and jittered rhythms), and speech-in-noise perception in adult percussionists and non-musicians. While the metrical rhythm condition assesses the participants' ability to produce correct sequences of hits and rests, the jittered condition assesses their ability to replicate fine timing deviations, more similar to those found in natural speech.

We assessed relationships between drumming skills and speech-in-noise perception, and then performed a hierarchical linear regression with speech-in-noise perception as the dependent variable. Given the absence of a regular beat in natural speech, we expected that the sequence-based skills would predict the ability to perceive speech in noise whereas the beat-based skills would not.

Material and methods

Participants comprised 31 young adults, split into two groups: percussionists (n=17, 5 females) and non-musicians (n=14, 4 females). Seventeen of the participants (8 percussionists, 9 non-musicians) had participated in an earlier study (Slater & Kraus, 2016) and returned for further testing. Percussionists were actively playing music and had at least five years of musical experience with drums and/or percussion as their primary instrument. Non-musicians had no more than three years of musical experience and no formal training within the seven years prior to the study. Participants were recruited with flyers on the Northwestern University campus and the Chicagoland area, and via postings on Craigslist. Participants had no external diagnosis of a neurological, language or attention disorder. Air-conducted audiometric thresholds $<$ 30 dB nHL for octaves from 125-8000 Hz. The groups did not differ on age, IQ **(**as measured by the Test of Nonverbal Intelligence (TONI) (Brown, Sherbenou, & Johnsen, 1997)**)** or hearing thresholds (pure tone averages) (all $p>0.4$). All procedures were approved by the Northwestern Institutional Review Board. Participants provided written consent and were compensated for their time.

Perceptual and cognitive tests

*Speech-in-noise perception****:*** Quick Speech-in-Noise Test (QuickSIN; Etymotic Research) (Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004) is a non-adaptive test of sentence perception in four-talker babble (three women and one man), presented in sound field at 55 dB SPL, with the first sentence starting at a SNR of 25 dB and each subsequent sentence presented with a 5dB SNR reduction down to 0 dB SNR. The sentences, which are spoken by a female, are syntactically correct yet have minimal semantic or contextual cues. Participants are instructed to repeat back each sentence, and their "SNR loss" is based on the number of target words correctly recalled. Sample sentences, with target words italicized, include "A *force equal* to that *would move* the *earth*." and "The *weight* of the *package* was *seen* on a *high scale*." Four lists were presented to each participant, with each list consisting of six sentences with five target words per sentence. (Returning participants were reassessed on this measure using a different set of four sentence lists from their first visit.) The total number of key words correctly recalled in the list (out of a possible 30) is subtracted from 25.5 to give the final SNR loss (see Killion et al., 2004 and QuickSIN User's Manual [Etymotic Research, 2001] for further details). The final score is the average SNR loss score from the four lists. A more negative SNR loss indicates better performance on the task (Killion et al., 2004).

Drumming tests

All the drumming tests used the same system for stimulus presentation, collection of drumming data, and marking of stimulus and drum onset times. Stimuli were presented with an iPod Nano (Apple) via headphones, and participants were asked to drum with one hand on a conga drum. The participant's drum hits were detected by a vibration-sensitive drum trigger pressed against the underside of the drum head. A copy of the audio signal presented to participants and the output of the drum trigger were recorded as two channels of a stereo input, using the audio recording program Audacity 2.0.5 (audacity.sourceforge.net). The two channels were saved together in a stereo sound file to provide a precise record of the timing relationship between stimuli and participant's drumming, while preserving the separate channels for analysis. Continuous stimulus and drum data were each converted to a list of onset times by a custom-written

MATLAB 7.5.0 (MathWorks, Inc.) program. The onset identification procedure is described in detail in Tierney and Kraus (2015). These stimulus and drum onsets were then subjected to further analyses for each rhythm test, as described below.

Drumming to metronome: Participants were asked to synchronise their drumming to an auditory pacing stimulus. Each trial consisted of 40 repetitions of a snare drum stimulus (duration 99ms, acquired at freesound.org) with a constant inter-onset-interval (IOI). Two trials were presented with an IOI of 667ms (1.5 Hz) and two with an IOI of 500ms (2 Hz), for a total of four trials. The last twenty beats of each trial were analysed, to give the participant ample time to synchronise to the beat. The coefficient of variability was calculated for each participant as the standard deviation of the IOI of the drum hits, divided by the IOI. This was averaged across all four trials. A smaller score indicated better (i.e. less variable) performance.

Drumming to musical beat: The participants listened to a series of twelve 20-30 second clips of music and were asked to drum along to the beat of the music. The musical stimuli were based on a tapping test developed by Iversen and Patel (2008). The average IOI of the participant's drum hits was calculated and compared with the average IOI of the beats of the music (as indicated by a trained drummer synchronising to the music; for details see Iversen and Patel (2008)). The difference between the IOIs was computed as an "error" score, with a smaller score indicating that the participant was able to accurately match the tempo of the music, as described in Iversen and Patel (2008).

Drumming with rhythm sequences (metrical and jittered): The stimuli were based on 3.2-second four-measure sequences developed by Povel and Essens (1985). In each trial, the same four-measure sequence was repeated ten times, for a total of forty measures. Participants were asked to listen to the sequences and then, whenever they were comfortable, to align their drumming exactly with the sounds. In the metrical condition, each four-measure sequence consisted of the conga sound presented nine times and was based on the same set of IOIs: five 200ms, two 400ms, one 600ms, and one 800ms. The sequences differed in the order in which these IOIs were presented, which gave rise to different temporal patterns Two of the trials contained sequences taken from the set of strongly metrical sequences listed in Povel and Essens, while two of the trials were weakly metrical sequences which contained more rests in strongly metrical positions (greater syncopation).

Performance was calculated based on whether the sequence of hits and rests in the participant's drumming matched the stimulus. First, both the stimulus and drumming data were converted to a sequence of hits and rests. For each 200ms time interval, it was determined whether the stimulus track contained a hit or silence. The drumming data were similarly converted to a sequence of hits and rests: if the participant hit the drum within a given 200ms interval, a hit was added to the drum sequence, otherwise a rest was assumed. The test was scored by comparing the sequences of hits and rests between the stimulus and drumming tracks. For example, if the stimulus sequence was [0 1 1 0] and the drumming sequence was [1 1 1 0], where one indicates a hit and zero indicates a rest, the participant's score on this small section of the test would be 75%. The 200ms time intervals were centred on potential hit positions such that if a participant's drum hit fell within 100ms before or after the stimulus, it would be scored as correct. This condition therefore captured the participants' ability to produce correct sequences of hits and rests.

In the jittered condition, the timing of each conga sound had been randomly jittered by 100-300ms, with the amounts of jitter uniformly distributed across each rhythm. Here, performance was calculated based on whether the participant successfully hit the drum within 100ms of each stimulus onset (i.e. up to 50ms before or after). This score therefore captured the participants' ability to match fine timing deviations in the stimulus sequence.

In each condition, performance was calculated across the second through tenth repetitions of each rhythm to produce a percent correct score for each trial. The scores were averaged across the four trials to produce a composite score for each condition.

Statistical analyses

All statistical analyses were conducted with SPSS (SPSS Inc., Chicago, IL). The Shapiro Wilk test for normality revealed that performance on the metrical rhythm tests as well as the accuracy of drumming to the beat of music were not normally distributed ($p<0.05$). Performance on the rhythm sequence tasks was arcsine-transformed and accuracy in drumming to music was square-root-transformed (based on the characteristics of their distributions) after which these measures were normally distributed ($p>0.05$), and the transformed variables were used in subsequent analyses.

Results

Percussionists outperformed non-musicians in speech-in-noise perception ($F_{(1,29)}=5.005$, $p=0.033$, $\eta^2=0.147$, Percussionists: M=-1.04 dB/SNR, SD=0.70; Non-musicians: M=-0.36 dB/SNR, SD=1.00) and all drumming tasks (see Table 1).

Speech-in-noise perception was correlated with the two sequence-based tasks (drumming to metrical and jittered sequences) but not with the beat-based measures (drumming to metronome and music). See Table 1 for a summary of group comparisons and correlations.

<Table 1 about here>

To further investigate the relationships between speech-in-noise perception and drumming skills, a three-step hierarchical linear regression was performed with speech-in-noise perception as the dependent variable. In the first step, the independent variables age, sex, non-verbal IQ and hearing thresholds did not significantly predict variance in speech-in-noise perception (adjusted $R^2 = .047$, $F_{(4,26)}=1.371$, $p=.271$). Next we added group (Percussionists and Non-musicians), which significantly improved the model [$\Delta R^2 = .188$, $p=.012$; overall model: adjusted $R^2 = .234$, $F_{(5,25)}=2.834$, $p=.037$]. Finally, we added the drumming measures, which further improved the model ($\Delta R^2 = .275$, $p=.015$). Overall, the model predicted 48% of variance in speech-in-noise perception (adjusted $R^2 = .481$, $F_{(9,21)}=4.088$, $p=.004$). The sequence measures both contributed significantly to the model, above and beyond demographic factors and group membership, while the beat-based measures did not. Further, group was no longer a significant predictor once the drumming measures were added. See Table 2 for a statistical summary of the regression analysis.

<Table 2 about here>

Discussion

Here, we provide the first evidence that the ability to perceive speech in noise is linked with rhythm production skills, across a range of rhythmic expertise. Specifically, drumming skills involving rhythm sequences predict speech-in-noise perception whereas beat-based measures do not. These outcomes build from our previous study in which we demonstrated that better speech-in-noise perception is associated with better rhythm discrimination (Slater & Kraus, 2016), and highlight rhythm as an important bridge between speech and music.

When listening to speech in noise, a listener may discern the rhythm of what is said, even when the specific words are unclear. This “rhythm template” may help in the process of disambiguating speech by constraining the candidate word patterns to those that match the perceived rhythm. The listener may therefore be able to resolve ambiguities by drawing on temporal cues, including prosody (Fear, Cutler, & Butterfield, 1995; Turk & Sawusch, 1997), phonological information (Klatt, 1976), phrase boundaries (Choi, Hasegawa-Johnson, & Cole, 2005; Scott, 1982), and syntactic structure (Gordon et al., 2015; Schmidt-Kassow & Kotz, 2008).

Sensitivity to timing relies upon both the ability to track patterns and the ability to detect deviations from those patterns. For example, deviations from expected timing provide an important means of musical expression, and live musical performance often departs from the formal regularity of the written score (Ashley, 2002; Palmer, 1997; Repp, 1992, 1995). Detailed analyses of live performances reveal variations in note onsets and durations on the order of hundreds of milliseconds (Ashley, 2002; Repp, 1995), comparable to the timescale of meaningful variations in syllable durations and prosodic stress patterns in speech, and within the same range as the 100-300ms deviations in our jittered rhythms task. Given both the metrical and jittered rhythm measures contributed unique explanatory power in our regression model, we propose that understanding a novel sentence in noise calls upon the ability to track temporal structure within the signal, as well as sensitivity to subtle timing deviations that may provide important clues about what was said.

It is important to note that the relevance of specific rhythmic skills to the perception of speech in noise may also be influenced by the temporal characteristics of the masker. For example, speech reception thresholds are lowered when listening to speech with a fluctuating vs. continuous masker (Festen & Plomp, 1990), which may be due in part to the ability to anticipate dips in fluctuating background noise. In the present study, the background noise comprised four-talker babble, therefore tracking the complex sequences of speech could help the listener anticipate dips and boost comprehension. However, in the case of a periodic masker, different rhythmic skills may come into play (i.e. the ability to track a periodic beat) and further research is needed to investigate these relationships in different listening conditions.

Rhythm is an integral part of musical practice and it is possible that non-percussionist musicians would demonstrate similar patterns of enhancement in both rhythm skills and speech-in-noise perception. Enhanced rhythm skills have been observed in non-percussionist instrumentalists (Rammsayer & Altenmüller, 2006; Slater, Tierney, & Kraus, 2013; Thompson, White-Schwoch, Tierney, & Kraus, 2015). Matthews et al. (2016) found no significant differences between percussionists, pianists, vocalists, and string players (Matthews, Thibodeau, Gunther, & Penhune, 2016) on several drumming tasks, but did identify a percussionist advantage over all other groups (musician and non-musician) for processing complex meter, and several studies report enhanced rhythm skills in percussionists (Cameron & Grahn, 2014; Ehrlé & Samson, 2005; Krause, Schnitzler, & Pollok, 2010; Manning & Schutz, 2016).

Evidence for a musician enhancement in speech-in-noise perception has been mixed (Boebinger et al., 2015; Parbery-Clark, Lam, & Kraus, 2009; Ruggles, Freyman, & Oxenham, 2014; Swaminathan, Mason, Streeter, Kidd Jr, & Patel, 2014), and it is possible this could be due to heterogeneity within the musician groups with respect to rhythm skills. Although the percussionists in the present study did not differ from non-percussionist instrumental musician groups in previous studies on the same speech-in-noise perception task (for example, see Parbery-Clark et al., 2009), this may also reflect stricter musicianship criteria in earlier studies with respect to age of training onset and years of musical practice. As previous work has emphasised, speech-in-noise perception relies upon a dynamic integrated network of cognitive and sensory processing (Anderson, White-Schwoch, Parbery-Clark, & Kraus, 2013; Mattys, Davis, Bradlow, &

Scott, 2012; Pichora-Fuller, Schneider, & Daneman, 1995). Enhanced speech-in-noise perception in musicians has previously been associated with stronger auditory cognitive skills (Parbery-Clark et al., 2009; Parbery-Clark, Tierney, Strait, & Kraus, 2012; Strait & Kraus, 2011), and further research is needed to determine whether advantages in speech-in-noise perception in non-percussionist instrumentalists are also mediated by rhythmic expertise, in addition to cognitive and sensory factors.

There is evidence that complex rhythm processing occurs in brain areas typically associated with language (Vuust, Roepstorff, Wallentin, Mouridsen, & Østergaard, 2006), and the recruitment of language areas for rhythm processing may also be increased in expert musicians (Herdener et al., 2014; Vuust et al., 2005). Patel's OPERA hypothesis proposed that speech perception advantages in musicians may reflect an experience-based adaptation whereby language networks are increasingly engaged and strengthened with musical practice (Patel, 2011). Our present findings suggest that rhythm may play an especially important role in mediating these benefits.

Brain regions traditionally associated with motor coordination, such as the cerebellum and basal ganglia, are also increasingly understood to play an important role in perception and timing (Graybiel, 1997; Ivry & Keele, 1989; Kotz, Schwartze, & Schmidt-Kassow, 2009). There is evidence of increased cerebellar activation when listening to speech in noise (Salvi et al., 2002), which could reflect an increased importance of temporal cues in suboptimal listening conditions. An interesting direction for future research is to investigate whether the transfer of rhythmic expertise to speech perception is driven by engagement with musical rhythm (irrespective of instrument), or by specific motor activities associated with drumming.

Conclusions

These outcomes suggest that sensitivity to rhythm sequences may be helpful in disambiguating the patterns of speech under degraded listening conditions. Although the present study cannot speak to the causal effects of training, our cross-sectional findings provide a basis for further investigation into the potential for rhythm-based training to strengthen building blocks of communication. The complex overlap between the rhythms of music and speech provides fertile ground for further research into the dynamic interaction between the brain and its environment, and how this may be shaped by experience.

Acknowledgements

The authors wish to thank Britta Swedenborg and Manto Agouridou for assistance with data collection and processing, and Trent Nicol, Elaine Thompson and Travis White-Schwoch who provided comments on an earlier version of this manuscript. This work was supported by the National Institutes of Health grant F31DC014891-01 to J.S. and the National Association of Music Merchants (NAMM) and the Knowles Hearing Center. The authors declare no competing financial interests.

Abercombie, D. (1967). *Elements of general phonetics*: Aldine Pub. Company.
Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research, 300*, 18-32. doi: 10.1016/j.heares.2013.03.006
Andreou, L.-V., Kashino, M., & Chait, M. (2011). The role of temporal regularity in auditory segregation. *Hearing Research, 280*(1), 228-235
Ashley, R. (2002). Do [n't] change a hair for me: The art of jazz rubato. *Music Perception, 19*(3), 311-332
Boebinger, D., Evans, S., Rosen, S., Lima, C. F., Manly, T., & Scott, S. K. (2015). Musicians and non-musicians are equally adept at perceiving masked speech. *Journal of the Acoustical Society of America, 137*(1), 378-387
Brown, L., Sherbenou, R., & Johnsen, S. K. (1997). Test of nonverbal intelligence. *A language free measure of cognitive ability*
Cameron, D. J., & Grahn, J. A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Frontiers in Human Neuroscience, 8*, 1003. doi: 10.3389/fnhum.2014.01003
Choi, J. Y., Hasegawa-Johnson, M., & Cole, J. (2005). Finding intonational boundaries using acoustic cues related to the voice source. *J Acoust Soc Am, 118*(4), 2579-2587
Cross, I. (1999). Is music the most important thing we ever did? Music, development and evolution. *Music, Mind and Science*, 10-39
Dalla Bella, S., Bialunska, A., & Sowinski, J. (2013). Why movement is captured by music, but less by speech: role of temporal regularity. *PLoS One, 8*(8), e71945. doi: 10.1371/journal.pone.0071945
Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *J Phonetics, 11*, 51-62
Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience, 19*(1), 158-164. doi: 10.1038/nn.4186
Ehrlé, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain and Cognition, 58*(1), 133-147
Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *J Acoust Soc Am, 97*(3), 1893-1904
Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *J Acoust Soc Am, 88*(4), 1725-1736
Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev Sci, 18*(4), 635-644. doi: 10.1111/desc.12230
Graybiel, A. M. (1997). The basal ganglia and cognitive pattern generators. *Schizophrenia bulletin, 23*(3), 459-469
Herdener, M., Humbel, T., Esposito, F., Habermeyer, B., Cattapan-Ludewig, K., & Seifritz, E. (2014). Jazz drummers recruit language-specific areas for the processing of rhythmic structure. *Cerebral Cortex, 24*(3), 836-843
Iversen, J. R., & Patel, A. D. (2008). The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population. In K. Miyazaki (Ed.), *10th*

International Conference on Music Perception and Cognition (pp. 465–468). Adelaide, SA: Causal Productions.
Ivry, R. B., & Keele, S. W. (1989). Timing functions of the cerebellum. *J Cogn Neurosci, 1*(2), 136-152. doi: 10.1162/jocn.1989.1.2.136
Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am, 116*(4 Pt 1), 2395-2405
Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *J Acoust Soc Am, 59*(5), 1208-1221
Kotz, S. A., Schwartze, M., & Schmidt-Kassow, M. (2009). Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex, 45*(8), 982-990. doi: 10.1016/j.cortex.2009.02.010
Krause, V., Schnitzler, A., & Pollok, B. (2010). Functional network interactions during sensorimotor synchronization in musicians and non-musicians. *Neuroimage, 52*(1), 245-251
Lehiste, I. (1977). Isochrony reconsidered. *J Phonetics, 5*, 253-263
Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry, 8*(2), 249-336
Manning, F. C., & Schutz, M. (2016). Trained to keep a beat: movement-related enhancements to timing perception in percussionists and non-percussionists. *Psychological Research, 80*(4), 532-542. doi: 10.1007/s00426-015-0678-5
Matthews, T. E., Thibodeau, J. N., Gunther, B. P., & Penhune, V. B. (2016). The Impact of Instrument-Specific Musical Training on Rhythm Perception and Production. *Frontiers in Psychology, 7*, 69. doi: 10.3389/fpsyg.2016.00069
Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang Cogn Process, 27*(7-8), 953-978
Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48*(1), 115-138
Parbery-Clark, A., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing, 30*(6), 653-661. doi: 10.1097/AUD.0b013e3181b412e9
Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience, 219*, 111-119
Patel, A. D. (2008). *Music, language, and the brain*: Oxford University Press, USA.
Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in Psychology, 2*
Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *J Acoust Soc Am, 97*(1), 593-608
Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*(4), 411-440
Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception, 24*(1), 37-48
Repp, B. H. (1992). Diversity and commonality in music performance: an analysis of timing microstructure in Schumann's "Traumerei". *J Acoust Soc Am, 92*(5), 2546-2568
Repp, B. H. (1995). Expressive timing in Schumann's ''Träumerei:''An analysis of performances by graduate student pianists. *The Journal of the Acoustical Society of America, 98*(5), 2413-2427

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One, 9*(1), e86980

Salvi, R. J., Lockwood, A. H., Frisina, R. D., Coad, M. L., Wack, D. S., & Frisina, D. R. (2002). PET imaging of the normal human auditory system: responses to speech in quiet and in background noise. *Hearing Research, 170*(1-2), 96-106

Schmidt-Kassow, M., & Kotz, S. A. (2008). Entrainment of syntactic processing? ERP-responses to predictable time intervals during syntactic reanalysis. *Brain Research, 1226*, 144-155

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *J Acoust Soc Am, 71*(4), 996-1007

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences, 34*(3), 114-123

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17*(1), 79-87. doi: 10.1007/s10339-015-0740-7

Slater, J., Tierney, A., & Kraus, N. (2013). At-risk elementary school children with one year of classroom music instruction are better at keeping a beat. *PLoS One, 8*(10), e77250

Smith, M. R., Cutler, A., Butterfield, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *J Speech Hear Res, 32*(4), 912-920

Strait, D. L., & Kraus, N. (2011). Can you hear me now? Musical training shapes functional brain networks for selective auditory attention and hearing speech in noise. *Frontiers in Psychology, 2*, 113. doi: 10.3389/fpsyg.2011.00113

Swaminathan, J., Mason, C. R., Streeter, T. M., Kidd Jr, G., & Patel, A. D. (2014). Spatial release from masking in musicians and non-musicians. *Journal of the Acoustical Society of America, 135*(4), 2281-2282

Teki, S., Grube, M., Kumar, S., & Griffiths, T. D. (2011). Distinct neural substrates of duration-based and beat-based auditory timing. *J Neurosci, 31*(10), 3805-3812

Thompson, E. C., White-Schwoch, T., Tierney, A., & Kraus, N. (2015). Beat Synchronization across the Lifespan: Intersection of Development and Musical Experience. *PLoS One, 10*(6), e0128839. doi: 10.1371/journal.pone.0128839

Tierney, A., & Kraus, N. (2015). Evidence for multiple rhythmic skills. *PLoS One, 10*(9), e0136645

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *J Phonetics, 25*(1), 25-41

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., & Østergaard, L. (2005). To musicians, the message is in the meter: pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage, 24*(2), 560-564

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing…: Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage, 31*(2), 832-841

Table 1. Group comparisons and correlations between speech-in-noise perception and drumming tasks.

	Percussionists vs. non-musicians		**Correlation with speech-in-noise perception**
	F value (*p*)	Effect size (η2)	R value (*p*)
Speech-in-noise perception (dB/SNR)	5.005 (.033)	0.147	-
BEAT-BASED DRUMMING			
Drumming to metronome (coeff var)	35.603 (<.001)	0.5	.172 (.354)
Drumming to beat of music (error, ms)	6.274 (.018)	0.178	.278 (.130)
SEQUENCE-BASED DRUMMING			
Drumming to rhythm sequences (% correct)	8.928 (.006)	0.235	-.500 (.004)
Drumming to jittered sequences (% correct)	6.222 (.019)	0.176	-.491 (.005)

Table 2. Summary of regression analyses predicting speech-in-noise perception.

Regression model	**Speech-in-noise perception** Standardized beta (p value)
STEP 1	
Age	.321 (.115)
Sex	-.334 (.187)
Non-verbal IQ	-.102 (.659)
Hearing thresholds	.434 (.051)
	R2=.174, adjusted R2=.047, $F_{(4,30)}$=1.371, p=.271
STEP 2	
Age	**.395 (.036)**
Sex	-.356 (.119)
Non-verbal IQ	-.131 (.529)
Hearing thresholds	**-.462 (.023)**
Group (Percussionists vs. non-musicians)	**.438 (.012)**
	R2change=.188, F change=7.346, p=.012
	R2=.362, adjusted R2=.234, $F_{(5,30)}$=2.834, p=.037
STEP 3	
Age	**.408 (.016)**
Sex	-.278 (.150)
Non-verbal IQ	-.028 (.878)
Hearing thresholds	**-.465 (.013)**
Group	-.242 (.366)
Drumming to metronome	-.211 (.348)
Drumming to musical beat	.011 (.946)
Drumming to metrical rhythms	**-.417 (.016)**
Drumming to jittered rhythms	**-.349 (.031)**
	R2 change=.275, F change=3.971, p=.015
Overall model:	**R2=.637, adjusted R2=.481,$F_{(9,30)}$=4.088, p=.004**

Figure 1. Correlations between speech-in-noise perception and the sequence-based drumming measures.