

# The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians

Jessica Slater<sup>1,2</sup> · Nina Kraus<sup>1,2,3,4,5</sup>

Received: 21 April 2015 / Accepted: 17 September 2015  
© Marta Olivetti Belardinelli and Springer-Verlag Berlin Heidelberg 2015

**Abstract** The natural rhythms of speech help a listener follow what is being said, especially in noisy conditions. There is increasing evidence for links between rhythm abilities and language skills; however, the role of rhythm-related expertise in perceiving speech in noise is unknown. The present study assesses musical competence (rhythmic and melodic discrimination), speech-in-noise perception and auditory working memory in young adult percussionists, vocalists and non-musicians. Outcomes reveal that better ability to discriminate rhythms is associated with better sentence-in-noise (but not words-in-noise) perception across all participants. These outcomes suggest that sensitivity to rhythm helps a listener understand unfolding speech patterns in degraded listening conditions, and that

observations of a “musician advantage” for speech-in-noise perception may be mediated in part by superior rhythm skills.

**Keywords** Speech-in-noise perception · Language · Music · Rhythm · Temporal processing · Auditory · Listening

## Introduction

From the natural flow of conversation, to the complexities of bebop and Bach, rhythms guide our communication through sound. As we try to make sense of each new sentence we hear, we are not only processing the words themselves but the patterns of sound they create. Under challenging listening conditions, for example trying to hear a friend’s voice in a noisy restaurant, the timing patterns of speech can provide important perceptual cues. Expertise with rhythmic processing, such as developed through musical practice, could potentially confer benefits for processing the rhythm and timing cues in speech. Several cross-sectional comparisons of musicians and non-musicians have revealed a musician advantage for perceiving speech in noise (Parbery-Clark et al. 2009, 2011; Swaminathan et al. 2015; Zendel and Alain 2012; Zendel et al. 2015) when groups are matched for key factors such as age, sex and IQ. Further, recent research from our laboratory and others has revealed links between rhythm abilities and various language skills, including reading (Gordon et al. 2015; Huss et al. 2011; Strait et al. 2011; Thomson and Goswami 2008; Tierney and Kraus 2013; Woodruff Carr et al. 2014), suggesting that rhythm may play a role in the transfer from music to language processing (see also

---

Handling Editor: Marta Olivetti Belardinelli, Sapienza University of Rome.

Reviewers: Mari Tervaniemi, University of Helsinki; Riccardo Brunetti, European University of Rome.

---

✉ Nina Kraus  
nkraus@northwestern.edu;  
<http://www.brainvolts.northwestern.edu>  
Jessica Slater  
<http://www.brainvolts.northwestern.edu>

<sup>1</sup> Auditory Neuroscience Laboratory, Northwestern University, Evanston, IL, USA

<sup>2</sup> Department of Communication Sciences, Northwestern University, Evanston, IL, USA

<sup>3</sup> Institute for Neuroscience, Northwestern University, Evanston, IL, USA

<sup>4</sup> Department of Neurobiology and Physiology, Northwestern University, Evanston, IL, USA

<sup>5</sup> Department of Otolaryngology, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA

Shahin 2011). However, the specific contribution of rhythm skills to speech-in-noise perception is unknown.

Speech-in-noise perception is not only important for everyday communication, but also provides an informative measure of integrated auditory function. Understanding speech in noise requires that multiple timescales of information are combined together in real time, from the millisecond details that differentiate consonants to the unfolding patterns of syllables and stress. Timing patterns are important for parsing the speech stream into meaningful units, for example durational patterns help identify the boundaries between words when listening in noise (Smith et al. 1989). Temporal regularities can also help focus attention to points in time when meaningful information is most likely to occur and thereby maximize processing efficiency (Large and Jones 1999), for example, it has been demonstrated that linguistic discrimination judgments are made more rapidly if the event aligns with the expected stress patterns of a speech stream (Quene and Port 2005). Better targeted attention and more efficient processing may help to maximize comprehension when listening conditions are degraded.

Hearing in noise also requires a listener to separate the target signal from competing inputs. Temporal cues are important for grouping auditory features into objects, and segregating a complex soundscape into distinct streams, since sounds originating from a common source are likely to share temporal characteristics (Andreou et al. 2011; Shamma et al. 2011). In the case of fluctuating maskers or multiple talkers, the ability to detect patterns in the competing sound streams may help the listener to anticipate when “dips” will occur and focus attention accordingly. Sensitivity to temporal patterns may therefore be important not only in parsing the target signal, but in taking advantage of fluctuations in the competing elements of a complex auditory scene.

Musical practice involves attention to many of the same aspects of timing that are important for speech, although there are important differences in how speech and music are structured in time: While music is typically organized around a regular beat, speech rhythms emerge dynamically over the course of an utterance. Articulatory and grammatical constraints provide some degree of predictability in the resulting patterns of speech, and certain spoken formats such as poetry or group chanting can impose metrical structure (Port 2003); however, everyday utterances do not typically exhibit strict periodic structure (Martin 1972; Patel 2010). Despite this difference, much of the expressive, communicative content in both live musical performance and natural speech is conveyed through the contrasting durations and subtle fluctuations of expressive timing within a phrase (Ashley 2002; Patel 2010), suggesting that the ability to parse non-periodic timing

features may provide an important bridge between music and speech perception (Cummins 2013; Patel 2010). A recent research study revealed enhanced timing abilities in trained percussionists not only for musical rhythms but also for rhythms without a musical beat (Cameron and Grahn 2014), suggesting that skills developed through musical practice could also transfer to the non-periodic rhythms of speech.

Speech-in-noise perception can be improved with computer-based training (Anderson et al. 2013; Song et al. 2011), suggesting that this ability is malleable with experience. A recent random assignment assessment of group music instruction reveals improved speech-in-noise perception in elementary school children after 2 years of group music class (Slater et al. 2015), indicating that the musician advantage observed in cross-sectional comparisons is not simply a result of preexisting differences between those who choose to pursue music training and those who do not. However, recent studies have reported either inconsistent (Fuller et al. 2014) or null (Boebinger et al. 2015; Ruggles et al. 2014) effects when comparing speech-in-noise perception in groups of musicians and non-musicians. For example, the Boebinger study set out to disentangle potential underlying mechanisms by assessing speech perception in a variety of maskers, yet did not find a musician advantage in any of the conditions. Beyond the characteristics of the stimuli and masking conditions, another possible contributing factor in the mixed experimental outcomes with musician versus non-musician comparisons may be that musical practice takes many forms, and it is still unknown which particular components of musical experience or expertise may confer advantages for speech perception. Even within categories such as classical or jazz performance, there is great variety in teaching approaches (e.g., learning to play from a score vs. learning to play by ear) that may influence the development of specific aspects of musical competence, such as rhythm perception and auditory memory. The same Boebinger study that failed to replicate a musician advantage for speech-in-noise perception also reported that musicians did not differ from non-musicians in duration discrimination, whereas previous studies have demonstrated enhanced duration discrimination in musicians (Jeon and Fricke 1997; Rammsayer and Altenmüller 2006), including a study with 2- to 3-year-olds engaged in informal musical activities in the home, which revealed larger neural responses to durational changes with more musical engagement (Putkinen et al. 2013). Given the relevance of durational patterns to speech perception, this provides an example of how differences in specific aspects of training and musical practice could influence the extent of skill transfer from music to speech processing and contribute to mixed

outcomes in cross-sectional studies comparing musicians and non-musicians.

In the present study, we depart from the musician versus non-musician dichotomy to focus on the specific role of rhythm discrimination in the perception of speech in noise. We assess speech-in-noise perception, musical competence (rhythmic and melodic) and auditory working memory performance in young adult percussionists, vocalists and non-musicians, with the goal of assessing relations between speech-in-noise perception and distinct aspects of musical competence across a range of musical abilities. Although rhythm is integral to musical activity regardless of instrument, previous research indicates that type of musical expertise does have some effect on temporal processing abilities. For example, percussionists outperform non-percussion instrumentalists in the perception of single time intervals (Cicchini et al. 2012) and in the reproduction of both beat-based and non-metric rhythms (Cameron and Grahn 2014), and jazz musicians show specialized neural processing of rhythm, including recruitment of left-hemisphere brain regions typically involved in linguistic processing (Herdener et al. 2014; Vuust et al. 2005), as well as more robust responses to rhythmic deviations than musicians who play other musical styles (Vuust et al. 2012). We therefore expected that our group of percussionists and drummers would show a relative strength in rhythm skills compared with the vocalists, due to the greater emphasis on rhythm in their musical practice. Our two musician groups had comparable amounts of musical experience and differed only with respect to their primary instrument.

We hypothesized that sensitivity to temporal patterns aids in the perception of speech in noise and can be strengthened by musical experience. We predicted that better performance on a rhythm discrimination test would relate to better perception of sentences in noise, but not words in noise.

## Methods

### Participants

Participants were 54 young adult males, aged 18–35 years and split into three groups: non-musicians ( $n = 17$ ),

vocalists ( $n = 21$ ) and percussionists ( $n = 16$ ). Musician participants (percussionists and vocalists) had been actively playing for at least the past 7 years with either drums/percussion or vocals as their primary instrument, based on self-report. Non-musician participants had no more than 3 years of musical experience across their lifetime, with no active music making within the 3 years prior to the study. All participants completed an audiological screening at the beginning of the testing session (pure tone thresholds at octave frequencies 0.125–8 kHz) and had normal hearing (any participant with thresholds above 25 dB nHL for more than one frequency in either ear did not continue with further testing). The three groups did not differ on age, IQ, as measured by the Test of Nonverbal Intelligence (TONI) (Brown et al. 1988), and the percussionist and vocalist groups did not differ with regard to years of musical experience or age of onset of musical training (see Table 1).

### Testing

#### *Musical competence*

The Musical Ear Test (MET) (Wallentin et al. 2010) uses a forced choice paradigm in which participants are presented with pairs of musical phrases and asked to indicate whether they are the same or different. There are 52 pairs of melodic phrases and 52 pairs of rhythmic phrases that vary greatly in level of difficulty such that performance is well distributed across a wide range of musical experience. The percentage of correct responses was calculated for each subtest, and a total score was also calculated as the average of the two subtest scores.

#### *Speech-in-noise perception*

Two standard clinical measures of speech-in-noise perception were used to measure the perception of both sentences and words in noise. Both were presented binaurally through insert earphones (ER-2; Etymotic Research). The Quick Speech-in-Noise Test (QuickSIN; Etymotic Research) (Killion et al. 2004) is a non-adaptive test of sentence perception in four-talker babble. Sentences are presented at 70 dB SPL, with the first

**Table 1** Participant characteristics

	Non-musicians ( $n = 17$ )	Vocalists ( $n = 21$ )	Drummers ( $n = 16$ )	Statistic
Age (years)	Mean 23.2 (SD = 3.8)	23.4 (3.6)	25.4 (5.7)	$F(2,51) = 1.217, p = 0.304$
Non-verbal IQ (percentile score)	71.6 (19.1)	74.1 (22.9)	77.3 (14.9)	$F(2,51) = 0.352, p = 0.705$
Years of musical experience	2.35 (1.5)*	14.7 (4.4)	16.7 (7.1)	$F(1,35) = 1.138, p = 0.293*$

\* Ten non-musician participants had some musical training; these participants were not included in the group comparison for years of musical experience

sentence starting at a signal-to-noise ratio (SNR) of 25 dB and with each subsequent sentence being presented with a 5 dB 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 (Wilson et al. 2007). Participants repeat each sentence (e.g., “The square peg will settle in the round hole”), and their SNR score is based on the number of correctly repeated target words. For each participant, four lists are presented, with each list consisting of six sentences with five target words per sentence. Each participant’s final score, termed “SNR loss,” is calculated as the average score across each of the four administered lists. A more negative SNR loss is indicative of better performance on the task.

Words in noise (WIN) is a non-adaptive test of speech-in-noise perception in four-talker babble noise (Wilson et al. 2007). Single words are presented following a carrier phrase, “Say the word...” and participants are asked to repeat the target words, one at a time. Thirty-five words are presented at 70 dB SPL with a starting SNR of 24 dB, decreasing in 4 dB steps until 0 dB with five words presented at each SNR. Each subject’s threshold was based on the number of correctly repeated words, with a lower score indicating better performance.

### Auditory working memory

Auditory working memory (AWM) was assessed using the *Woodcock Johnson III Test of Cognitive Abilities* Auditory Working Memory subtest (Woodcock et al. 2001). Participants listened to a series of intermixed nouns and digits presented on a CD player and were asked to repeat first the nouns and then the digits in their respective sequential orders (e.g., the correct ordering of the following sequence, “4, salt, fox, 7, stove, 2” is “salt, fox, stove” followed by “4, 7, 2”). Age-normed standard scores were used for all

statistical analyses. Three participants did not complete this test due to time constraints.

All procedures were approved by the Northwestern Institutional Review Board. Participants provided written consent and were compensated for their time.

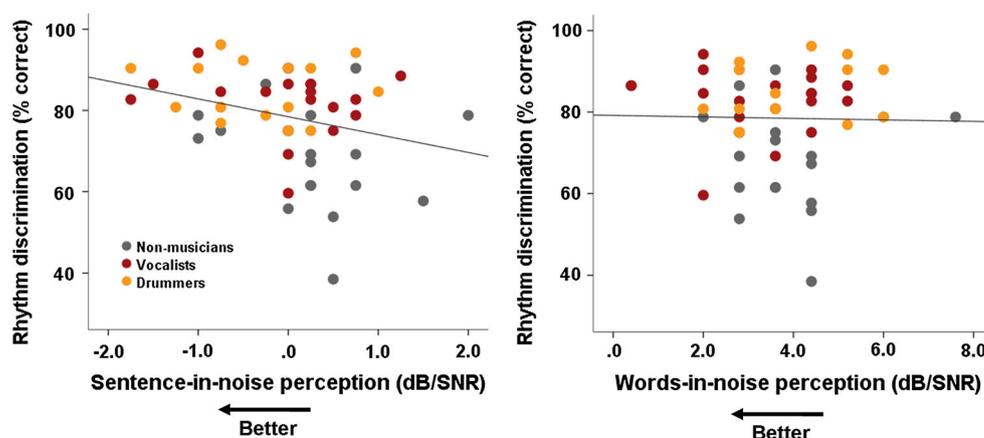
## Results

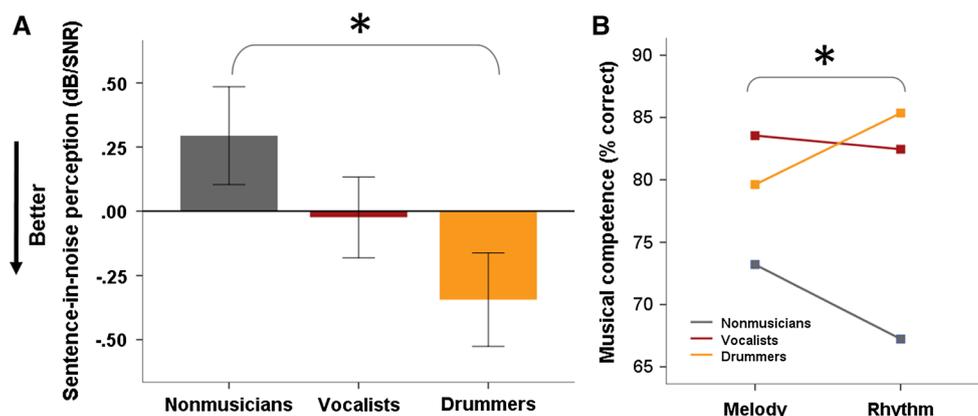
Better performance on the musical competence rhythm subtest was associated with better performance on the QuickSIN across all participants ( $r = -0.285$ ,  $p = 0.037$ , see Fig. 1). There was no relationship between melodic competence and QuickSIN ( $r = -0.150$ ,  $p = 0.278$ ). The ability to perceive words in noise (WIN) did not relate to either rhythmic or melodic competence (both  $p > 0.7$ ).

The combined musician group outperformed the non-musician group on the QuickSIN test of sentence-in-noise perception ( $F_{(52,1)} = 4.321$ ,  $p = 0.043$ ). Post hoc pairwise comparisons between the three groups revealed that the percussionists significantly outperformed the non-musicians (mean difference: 0.638 dB/SNR,  $p = 0.045$ , see Fig. 2), whereas the vocalists did not differ significantly from either group ( $p > 0.3$ ).

The musician group also outperformed the non-musician group on the musical competence test (melody subtest:  $F_{(52,1)} = 24.213$ ,  $p < 0.001$ ; rhythm subtest:  $F_{(52,1)} = 29.233$ ,  $p < 0.001$ ; total score:  $F_{(52,1)} = 36.167$ ,  $p < 0.001$ ) and on the test of auditory working memory ( $F_{(49,1)} = 4.929$ ,  $p = 0.031$ ). There was no difference between musicians and non-musicians in their performance on the WIN (words-in-noise) test ( $F_{(52,1)} = 1.204$ ,  $p = 0.278$ ). The percussionist and vocalist groups did not differ in their performance on the musical competence test (rhythm or melody subtests), speech-in-noise perception (words or sentences) or auditory working memory ( $p > 0.3$  for all pairwise comparisons).

**Fig. 1** Better rhythm competence relates to better sentence-in-noise perception, but not words-in-noise perception





**Fig. 2** **a** Percussionists/drummers are better at perceiving sentences in noise than non-musicians. **b** The percussionists/drummers show an advantage for the rhythm versus melody subtest, and vocalists show an advantage for melody versus rhythm subtest, although these two groups do not differ in their average scores (\*Significant interaction

between musician type and musical subtest at  $p < 0.05$ . Non-musicians perform significantly worse on both subtests and were excluded from the repeated measures ANOVA; their mean scores are displayed for reference)

### Comparison of percussionists versus vocalists

To further investigate the impact of primary instrument, we performed two additional sets of repeated measures analyses including only the musician groups. First, we assessed the effect of musician type on melody versus rhythm subtest performance, covarying for years of musical practice. This revealed an interaction between musician type and music subtest (RMANOVA musician group  $\times$  subtest:  $F_{(34,3)} = 5.608$ ,  $p = 0.024$ , see Fig. 2), with percussionists performing relatively better on the rhythm versus melody subtest and vocalists performing relatively better on melody versus rhythm. There was also a main effect of years of musical practice ( $F_{(34,3)} = 4.222$ ,  $p = 0.048$ ). Secondly, we asked whether musician type influences relative performance on the perception of sentences versus words in noise. This analysis revealed a significant interaction between musician type and speech-in-noise test (RMANOVA musician group  $\times$  SIN test:  $F_{(35,2)} = 4.302$ ,  $p = 0.045$ ), with the percussionist group performing relatively better on the QuickSIN than the WIN, compared with the vocalists.

### Auditory working memory

Given there is an auditory working memory component to the musical competence tests and previous work has implicated auditory working memory as an important factor in speech-in-noise perception (Kraus et al. 2012; Parbery-Clark et al. 2009), we also assessed correlations between auditory working memory, the musical competence subtests and speech-in-noise performance. We found that auditory working memory correlated with rhythm competence ( $r = 0.361$ ,  $p = 0.009$ ), but not with melody

competence ( $r = 0.116$ ,  $p = 0.417$ ). However, the relationship between QuickSIN performance and auditory working memory was not significant in this dataset ( $r = -0.207$ ,  $p = 0.145$ ), suggesting that the relationship between rhythm discrimination and speech-in-noise perception was not simply driven by auditory working memory.

### Discussion

We provide the first evidence for a relationship between rhythm discrimination and the ability to perceive sentences in noise. Making sense of a novel speech pattern involves complex temporal processing and may engage similar processes to those involved in the perception and production of musical phrases. Musical practice with a focus on rhythmic skills may therefore increase sensitivity to timing patterns that are important for speech perception, and improve the ability to perceive speech under degraded listening conditions.

Rhythm perception may contribute to the perception of speech in noise in various ways. Accurate encoding of durational patterns not only helps to identify word boundaries in noise (Smith et al. 1989), but may also help to narrow the range of possible word sequences to those that match the perceived rhythmic pattern. Further, rhythm skills have been shown to predict a significant amount of individual variance in grammatical abilities in children (Gordon et al. 2015). It is therefore plausible that rhythm serves as a kind of proxy for grammatical processing: A rhythm pattern may imply a certain grammatical structure, allowing candidate word sequences to be eliminated if they do not fit that grammatical template. There is also neural

evidence that temporal regularity in speech influences both syntactic and semantic processing (Roncaglia-Denissen et al. 2013; Rothermich et al. 2012; Schmidt-Kassow and Kotz 2009), and a greater sensitivity to temporal patterns may bootstrap these higher levels of linguistic processing. These previous findings are consistent with the lack of relationship between rhythm discrimination and the perception of words in noise observed in the present study. The words-in-noise (WIN) test sets up a very clear temporal expectation by presenting a consistent carrier phrase: “Say the word...”, followed by the target word. However, this format does not provide any timing cues that would help disambiguate the target word itself, since there is no variation in grammatical structure, context or prosodic characteristics that could aid in the identification of a monosyllabic target word.

There are important differences between the temporal characteristics of speech and music, such as the absence of a steady beat in everyday speech. However, there is evidence for some overlap in the neural circuitry involved in perceiving speech and music, particularly with respect to rhythm and pattern processing (for example, see Patel 2011; Patel and Iversen 2014). Brain regions typically associated with movement and motor planning have been implicated in rhythmic processing (see Grahn 2012 for review), even in the absence of any overt movement (Chen et al. 2008). Similarly, input from the motor systems involved in speech production may also inform predictions about an incoming speech stream (Davis and Johnsruide 2007), since characteristics of the human vocal mechanism place inherent constraints on the sound sequences likely to be generated in the course of a typical human utterance. It has been proposed that the auditory-motor connections that evolved in conjunction with vocal production for language may also have enabled certain aspects of rhythmic processing that are typically found only in humans and other vocal learning species, such as the ability to synchronize to a beat (Patel and Iversen 2014). Although the present findings cannot speak directly to the involvement of motor systems in mediating the connections between rhythmic processing and speech perception, recent work reveals increased recruitment of the motor system when listening in adverse conditions, suggesting the motor system may help to compensate for impoverished sensory representations in noise (Alain and Du 2015).

It has been proposed that tracking of temporal regularities in both speech and music relies on the entrainment of neural oscillators (Bastiaansen and Hagoort 2006; Giraud and Poeppel 2012; Harle et al. 2004; Large and Snyder 2009; Nozaradan et al. 2012; Peelle and Davis 2012). The ability to perceive speech in noise may be supported by the synchronization of low-frequency cortical activity to the slow temporal modulations of speech, which results in a

background-invariant neural representation of the speech signal (Ding and Simon 2013), and motor regions of the brain, such as the basal ganglia, also play a critical role in the coordination of this oscillatory neural activity (Kotz et al. 2009). Further investigation into the rhythms of music and speech may provide invaluable insight into these deeper mechanisms of neural synchrony, which not only underlie motor planning but also facilitate the exquisite choreography of real-time communication.

Previous research has suggested that frequency discrimination could be a contributing factor in the musician advantage for speech-in-noise perception (Parbery-Clark et al. 2009), and it is perhaps surprising that melodic discrimination did not relate to speech-in-noise perception in the present study. Interestingly, a recent study demonstrated enhanced pitch perception in musicians yet did not observe any advantage for speech-in-noise perception (Ruggles et al. 2014). A study with amusic Mandarin speakers revealed that amusics performed worse than normal participants on a speech intelligibility task in both quiet and noise (Liu et al. 2015), yet this disadvantage for speech perception was unrelated to performance on a pitch perception task. Further, the disadvantage was observed even when the pitch contour of the target sentence was flattened, suggesting that the speech perception deficit in amusics was not due to impaired processing of the pitch contour. Rather, the results led the authors to deduce that the deficit may be due to difficulties segmenting speech elements, consistent with previous research associating amusia with impaired segmental processing (Jones et al. 2009), and providing further support for the importance of timing cues for perceiving speech in noise. Another recent study that was unable to replicate the musician advantage for speech-in-noise perception also reported that the musician and non-musician groups did not differ in their ability to discriminate durations (Boebinger et al. 2015), in contrast to previous studies showing enhanced duration discrimination in musicians (Jeon and Fricke 1997; Rammsayer and Altenmüller 2006). It therefore seems conceivable that some of the inconsistency in outcomes across studies comparing musicians and non-musicians in their ability to perceive speech in noise could be due to variations in rhythmic expertise in the musician groups, given the great diversity of musical training and practice styles, and their differing degrees of emphasis on rhythm skills.

Working memory has also been implicated as an important factor for speech-in-noise perception (Kraus et al. 2012; Parbery-Clark et al. 2009). Although working memory performance tracks with rhythm (but not melody) competence in the present study, it is not correlated with speech-in-noise performance. These findings indicate that the relationship between rhythm discrimination and

speech-in-noise perception is not simply driven by working memory capacity. However, the lack of relationship between working memory performance and melody discrimination suggests that processing rhythm patterns places additional demands on working memory, beyond simply retaining two phrases in memory to compare them (which is required in both the rhythm and melody subtests, but also reflects the challenge of storing temporal sequences, since this aspect was only present in the rhythm test (the melodic phrases were identical with respect to note durations). Further research is needed to disentangle these relationships between working memory, rhythm processing and speech perception, and how they may be shaped by musical experience.

The present study investigates common mechanisms involved in rhythm processing and speech-in-noise perception, through the lens of musical expertise. We interpret our findings within a framework of experience-based plasticity, consistent with the accumulating body of research demonstrating that musical experience shapes neural and cognitive function (for review, see Kraus and White-Schwoch in press). However, given the cross-sectional design of this study, we are limited in the conclusions that can be drawn regarding causal effects of music training or primary instrument, since there may be preexisting differences between individuals who choose to play music versus those who do not, or between those who choose to sing versus those who choose to play drums or percussion. There is evidence of differential effects of chosen instrument on brain function that are unlikely to exist prior to training, such as more robust neural responses to the timbre of an individual's instrument (Pantev et al. 2001; Strait et al. 2012); on the other hand, there is evidence highlighting the contribution of factors such as genetics and personality traits to musical practice habits and resulting expertise (Butkovic et al. 2015; Mosing et al. 2014; Schellenberg 2015). Therefore, it is likely that both nature and nurture contribute to the shaping of musical and linguistic capabilities in a given individual. Although we cannot draw conclusions about causal effects of training from this study, outcomes provide an important step toward identifying the specific aspects of musical practice that may be of particular relevance to understanding speech in noise. Longitudinal evidence supports the potential for music training to improve speech-in-noise perception (Slater et al. 2015); however, comparisons of musicians and non-musicians have yielded mixed results and the diversity of musical experience may well be a factor in these inconsistent outcomes. The specific importance of rhythm in language processing is reflected in the success of intervention studies using rhythm-based training to strengthen reading skills (Bhide et al. 2013; Overy 2000, 2003), as well as evidence suggesting that the benefits of melodic intonation therapy

for improving speech production in aphasic patients may in fact be attributed to rhythmic aspects of the therapy (Stahl et al. 2011). The present findings provide a basis for investigating the potential benefits of rhythm-based training to strengthen other important communication skills, such as the perception of speech in noise.

In conclusion, rhythm and timing cues are important for the perception of novel speech patterns in degraded listening conditions. Superior rhythm skills track with better performance on a standard clinical test of sentence-in-noise perception, suggesting that rhythm skills developed through regular musical practice may confer benefits for speech processing. Further, primary instrument may influence the development of rhythmic versus melodic competence, although further research is necessary to disentangle the effects of experience-based plasticity versus inherent differences that could lead an individual to be drawn to one instrument over another, as well as other factors such as the impact of learning and playing styles (e.g., reading from a score vs. improvising, playing with others vs. playing alone). There is much still to be understood about how different aspects of temporal information are tracked and integrated by the nervous system during the perception of speech and other communication signals. Future research should continue to explore the role of distinct aspects of rhythm processing in speech perception, with particular regard to natural stimuli that may not exhibit strict periodic structure but which do, nonetheless, contain predictable patterns over time.

**Acknowledgments** The authors wish to thank Britta Swedenborg, Emily Spitzer and Andrea Azem for assistance with data collection and processing, and Trent Nicol, Kali Woodruff Carr, Travis White-Schwoch and Adam Tierney 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., the National Association of Music Merchants (NAMM) and Knowles Hearing Center, Northwestern University.

## References

- Alain C, Du Y (2015) Recruitment of the speech motor system in adverse listening conditions. *J Acoust Soc Am* 137:2211
- Anderson S, White-Schwoch T, Parbery-Clark A, Kraus N (2013) Reversal of age-related neural timing delays with training. *Proc Natl Acad Sci* 110:4357–4362
- Andreou L-V, Kashino M, Chait M (2011) The role of temporal regularity in auditory segregation. *Hear Res* 280:228–235
- Ashley R (2002) Do [n't] change a hair for me: the art of jazz rubato. *Music Percept* 19:311–332
- Bastiaansen M, Hagoort P (2006) Oscillatory neuronal dynamics during language comprehension. *Prog Brain Res* 159:179–196
- Bhide A, Power A, Goswami U (2013) A rhythmic musical intervention for poor readers: a comparison of efficacy with a letter-based intervention mind. *Brain Educ* 7:113–123
- Boebinger D, Evans S, Rosen S, Lima CF, Manly T, Scott SK (2015) Musicians and non-musicians are equally adept at perceiving masked speech. *J Acoust Soc Am* 137:378–387

- Brown L, Sherbenou R, Johnsen SK (1988) Test of nonverbal intelligence. Pro-Ed, Austin
- Butkovic A, Ullén F, Mosing MA (2015) Personality related traits as predictors of music practice: underlying environmental and genetic influences. *Personal Individ Differ* 74:133–138
- Cameron DJ, Grahn JA (2014) Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Front Human Neurosci* 8:1003
- Chen JL, Penhune VB, Zatorre RJ (2008) Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex* 18:2844–2854. doi:[10.1093/cercor/bhn042](https://doi.org/10.1093/cercor/bhn042)
- Cicchini GM, Arrighi R, Cecchetti L, Giusti M, Burr DC (2012) Optimal encoding of interval timing in expert percussionists. *J Neurosci* 32:1056–1060
- Cummins F (2013) Joint speech: the missing link between speech and music? *PERCEPTA-Revista de Cognição Musical* 1:17–32
- Davis MH, Johnsrude IS (2007) Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hear Res* 229:132–147
- Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J Neurosci* 33:5728–5735
- Fuller CD, Galvin III JJ, Maat B, Free RH, Başkent D (2014) The musician effect: does it persist under degraded pitch conditions of cochlear implant simulations? *Front Neurosci* 8:179
- Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517
- Gordon RL, Shivers CM, Wieland EA, Kotz SA, Yoder PJ, Devin McAuley J (2015) Musical rhythm discrimination explains individual differences in grammar skills in children. *Dev Sci* 18:635–644
- Grahn JA (2012) Neural mechanisms of rhythm perception: current findings and future perspectives. *Top Cogn Sci* 4:585–606. doi:[10.1111/j.1756-8765.2012.01213.x](https://doi.org/10.1111/j.1756-8765.2012.01213.x)
- Harle M, Rockstroh B, Keil A, Wienbruch C, Elbert T (2004) Mapping the brain's orchestration during speech comprehension: task-specific facilitation of regional synchrony in neural networks. *BMC Neurosci* 5:40
- 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. *Cereb Cortex* 24:836–843
- Huss M, Verney JP, Fosker T, Mead N, Goswami U (2011) Music, rhythm, rise time perception and developmental dyslexia: perception of musical meter predicts reading and phonology. *Cortex J Devot Study Nerv Syst Behav* 47:674–689. doi:[10.1016/j.cortex.2010.07.010](https://doi.org/10.1016/j.cortex.2010.07.010)
- Jeon JY, Fricke FR (1997) Duration of perceived and performed sounds. *Psychol Music* 25:70–83
- Jones JL, Lucker J, Zalewski C, Brewer C, Drayna D (2009) Phonological processing in adults with deficits in musical pitch recognition. *J Commun Disord* 42:226–234
- Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, 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* 116:2395–2405
- Kotz SA, 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 J Devot Study Nerv Syst Behav* 45:982–990. doi:[10.1016/j.cortex.2009.02.010](https://doi.org/10.1016/j.cortex.2009.02.010)
- Kraus N, White-Schwoch T (in press) Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cogn Sci*
- Kraus N, Strait DL, Parbery-Clark A (2012) Cognitive factors shape brain networks for auditory skills: spotlight on auditory working memory. *Ann NY Acad Sci* 1252:100–107
- Large E, Jones M (1999) The dynamics of attending: How people track time-varying events. *Psychol Rev* 106:119–159
- Large EW, Snyder JS (2009) Pulse and meter as neural resonance. *Ann N Y Acad Sci* 1169:46–57. doi:[10.1111/j.1749-6632.2009.04550.x](https://doi.org/10.1111/j.1749-6632.2009.04550.x)
- Liu F, Jiang C, Wang B, Xu Y, Patel AD (2015) A music perception disorder (congenital amusia) influences speech comprehension. *Neuropsychologia* 66:111–118
- Martin JG (1972) Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychol Rev* 79:487–509
- Mosing MA, Madison G, Pedersen NL, Kuja-Halkola R, Ullén F (2014) Practice does not make perfect no causal effect of music practice on music ability. *Psychol Sci* 0956797614541990
- Nozaradan S, Peretz I, Mouraux A (2012) Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J Neurosci Off J Soc Neurosci* 32:17572–17581. doi:[10.1523/JNEUROSCI.3203-12.2012](https://doi.org/10.1523/JNEUROSCI.3203-12.2012)
- Overy K (2000) Dyslexia, temporal processing and music: the potential of music as an early learning aid for dyslexic children. *Psychol Music* 28:218–229
- Overy K (2003) Dyslexia and music. From timing deficits to musical intervention. *Ann NY Acad Sci* 999:497–505
- Pantev C, Roberts LE, Schulz M, Engelen A, Ross B (2001) Timbre-specific enhancement of auditory cortical representations in musicians. *NeuroReport* 12:169–174
- Parbery-Clark A, Skoe E, Lam C, Kraus N (2009) Musician enhancement for speech-in-noise. *Ear Hear* 30:653–661. doi:[10.1097/AUD.0b013e3181b412e9](https://doi.org/10.1097/AUD.0b013e3181b412e9)
- Parbery-Clark A, Strait DL, Anderson S, Hittner E, Kraus N (2011) Musical experience and the aging auditory system: implications for cognitive abilities and hearing speech in noise. *PLoS ONE* 6:e18082. doi:[10.1371/journal.pone.0018082](https://doi.org/10.1371/journal.pone.0018082)
- Patel AD (2010) *Music, language, and the brain*. Oxford University Press, Oxford
- Patel AD (2011) Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Front Psychol* 2:142
- Patel AD, Iversen JR (2014) The evolutionary neuroscience of musical beat perception: the action simulation for auditory prediction (ASAP) hypothesis. *Front Syst Neurosci*. doi:[10.3389/fnsys.2014.00057](https://doi.org/10.3389/fnsys.2014.00057)
- Peelle JE, Davis MH (2012) Neural oscillations carry speech rhythm through to comprehension. *Front Psychol* 3:320
- Port RF (2003) Meter and speech. *J Phon* 31:599–611
- Putkinen V, Tervaniemi M, Huotilainen M (2013) Informal musical activities are linked to auditory discrimination and attention in 2–3-year-old children: an event-related potential study. *Eur J Neurosci* 37:654–661. doi:[10.1111/ejn.12049](https://doi.org/10.1111/ejn.12049)
- Quene H, Port RF (2005) Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62:1–13. doi:[10.1159/000087222](https://doi.org/10.1159/000087222)
- Rammsayer T, Altenmüller E (2006) Temporal information processing in musicians and nonmusicians. *Music Percept* 24:37–48
- Roncaglia-Denissen MP, Schmidt-Kassow M, Kotz SA (2013) Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS ONE* 8:e56000. doi:[10.1371/journal.pone.0056000](https://doi.org/10.1371/journal.pone.0056000)
- Rothermich K, Schmidt-Kassow M, Kotz SA (2012) Rhythm's gonna get you: regular meter facilitates semantic sentence processing. *Neuropsychologia* 50:232–244
- Ruggles DR, Freyham RL, Oxenham AJ (2014) Influence of musical training on understanding voiced and whispered speech in noise. *PLoS ONE* 9:e86980

- Schellenberg EG (2015) Music training and speech perception: a gene–environment interaction. *Ann NY Acad Sci* 1337:170–177
- Schmidt-Kassow M, Kotz SA (2009) Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J Cogn Neurosci* 21:1693–1708
- Shahin AJ (2011) Neurophysiological influence of musical training on speech perception. *Front Psychol* 2:126
- Shamma SA, Elhilali M, Micheyl C (2011) Temporal coherence and attention in auditory scene analysis. *Trends Neurosci* 34:114–123
- Slater J, Skoe E, Strait DL, O’Connell S, Thompson E, Kraus N (2015) Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behav Brain Res* 291:244–252
- Smith MR, Cutler A, Butterfield S, Nimmo-Smith I (1989) The perception of rhythm and word boundaries in noise-masked speech. *J Speech Hear Res* 32:912–920
- Song JH, Skoe E, Banai K, Kraus N (2011) Training to improve hearing speech in noise: biological mechanisms. *Cereb Cortex* 22:1180–1190. doi:10.1093/cercor/bhr196
- Stahl B, Kotz SA, Henseler I, Turner R, Geyer S (2011) Rhythm in disguise: why singing may not hold the key to recovery from aphasia. *Brain* awr240
- Strait DL, Hornickel J, Kraus N (2011) Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behav Brain Funct* 7:44
- Strait DL, Chan K, Ashley R, Kraus N (2012) Specialization among the specialized: auditory brainstem function is tuned into timbre. *Cortex J Devot Study Nerv Syst Behav* 48:360–362. doi:10.1016/j.cortex.2011.03.015
- Swaminathan J, Mason CR, Streeter TM, Best V, Kidd G Jr, Patel AD (2015) Musical training, individual differences and the cocktail party problem. *Sci Rep* 5:11628. doi:10.1038/srep11628
- Thomson JM, Goswami U (2008) Rhythmic processing in children with developmental dyslexia: auditory and motor rhythms link to reading and spelling. *J Physiol Paris* 102:120–129. doi:10.1016/j.jphysparis.2008.03.007
- Tierney AT, Kraus N (2013) The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain Lang* 124:225–231. doi:10.1016/j.bandl.2012.12.014
- Vuust P, Pallesen KJ, Bailey C, van Zuijlen TL, 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* 24:560–564
- Vuust P, Brattico E, Seppänen M, Näätänen R, Tervaniemi M (2012) Practiced musical style shapes auditory skills. *Ann N Y Acad Sci* 1252:139–146
- Wallentin M, Nielsen AH, Friis-Olivarius M, Vuust C, Vuust P (2010) The Musical Ear Test, a new reliable test for measuring musical competence. *Learn Individ Differ* 20:188–196
- Wilson RH, McArdle RA, Smith SL (2007) An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. *J Speech Lang Hear Res* 50:844–856. doi:10.1044/1092-4388(2007)059
- Woodcock RW, McGrew K, Mather N (2001) Woodcock-Johnson tests of achievement. Riverside Publishing, Itasca
- Woodruff Carr K, White-Schwoch T, Tierney AT, Strait DL, Kraus N (2014) Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. In: *Proceedings of the National Academy of Sciences*, 201406219
- Zendel BR, Alain C (2012) Musicians experience less age-related decline in central auditory processing. *Psychol Aging* 27:410
- Zendel BR, Tremblay CD, Belleville S, Peretz I (2015) The impact of musicianship on the cortical mechanisms related to separating speech from background noise. *J Cogn Neurosci* 27:1044–1059. doi:10.1162/jocn\_a\_00758