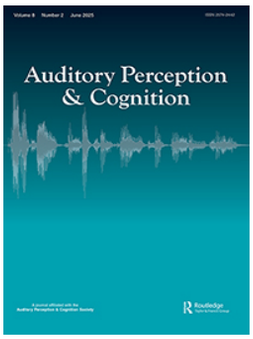

Auditory Perception & Cognition

ISSN: 2574-2442 (Print) 2574-2450 (Online) Journal homepage: www.tandfonline.com/journals/rpac20

Singing in Noise: Can Music-Based Acoustic Features Aid Speech-in-Noise Comprehension?

Benjamin Rich Zendel & Liam Robbins

To cite this article: Benjamin Rich Zendel & Liam Robbins (31 Jul 2025): Singing in Noise: Can Music-Based Acoustic Features Aid Speech-in-Noise Comprehension?, Auditory Perception & Cognition, DOI: 10.1080/25742442.2025.2541861

To link to this article: https://doi.org/10.1080/25742442.2025.2541861

Published online: 31 Jul 2025.

AUDITORY PERCEPTION & COGNITION
https://doi.org/10.1080/25742442.2025.2541861

Routledge
Taylor & Francis Group

Singing in Noise: Can Music-Based Acoustic Features Aid Speech-in-Noise Comprehension?

Benjamin Rich Zendel and Liam Robbins

Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

ABSTRACT

The ability to understand speech-in-noise (SPiN) can be improved with musical training, and music perception is resistant to age-related decline compared to other aspects of auditory cognition. This leads to the possibility that music-based forms of rehabilitation could improve SPiN comprehension. One possible approach to this putative rehabilitation program is to use a scaffolding technique, where a cognitive strength is used to scaffold a cognitive weakness. The first step in this line of research is to determine a source of auditory cognitive strength in SPiN comprehension. Accordingly, the goal of the current study was to determine if adding musical features to target speech could improve SPiN comprehension. Participants were presented with a series of sentences that were either spoken, sung, rapped, or sung with speech-like rhythm. Sentences were presented in noise, and the signal-to-noise ratio [SNR] was adapted based on accuracy. A 50% SNR threshold was determined for each condition. Overall, performance was best when target sentences were sung with speech-like rhythm, and worst when they were sung with musical rhythm. This pattern of results suggests that musical pitch contours can aid in understanding SPiN, and could potentially be used as a cognitive scaffold (i.e. a cognitive strength compared to understanding naturally spoken SPiN) to improve the ability to understand SPiN.

ARTICLE HISTORY
Received 7 January 2025
Accepted 24 July 2025

Keywords
Speech-in-noise; singing; hearing; music

Introduction

One of the most commonly reported hearing difficulties, particularly in older adults, is a difficulty understanding speech when there is loud background noise (i.e., speech-in-noise: SPiN) (Gates & Mills, 2005; Pichora-Fuller et al., 2016). Although hearing aids are commonly used by older adults, in many cases, hearing aids do not help when there is loud background noise (McCormack & Fortnum, 2013). Accordingly, it is of utmost importance to develop alternative forms of auditory rehabilitation that could be used in conjunction with amplification-based hearing assistance to improve peoples "listening" abilities.

An interesting possibility is to use the cognitive scaffolding technique as a foundation for this putative rehabilitation program. A cognitive scaffold is an area of relative cognitive strength that can be used to help support and rehabilitate an area of cognitive

CONTACT Benjamin Rich Zendel bzendel@mun.ca Faculty of Medicine, Memorial University, 300 Prince Phillip Dr, St John's, NL A1B3V6, Canada

weakness. An example of this type of program is Melodic Intonation Therapy (MIT (Albert et al., 1973; Norton et al., 2009).; MIT is a form of treatment that can help some people with aphasia re-learn how to talk by using singing. MIT likely works because some people who have aphasia retain the ability to sing, and MIT uses this area of cognitive strength (i.e., singing) as a scaffold to help rehabilitate an area of cognitive weakness (i.e., speaking). In other words, people with aphasia are taught to speak by "singing" their speech. Over time, the "singing" is slowly removed, and the person with aphasia slowly re-learns to speak. Although there are debates about the effectiveness of MIT in certain situations, the administration of MIT and the underlying neurophysiological mechanisms; the idea of using a cognitive strength to scaffold on a cognitive weakness during rehabilitation is supported (Merrett et al., 2014; Zumbansen et al., 2014).

In the context of understanding SPiN, the first step would be to determine if there is a way to modify target speech that would make it more comprehensible when presented in noise. Accordingly, the goal of the current study is to determine if the acoustic features of singing could be used to aid in understanding SPiN. If the acoustic features of singing aid in understanding SPiN, then it may be possible to develop a rehabilitation program that uses musical acoustic features found in singing as a aural cognitive strength, and then slowly reduce those features so a person with difficulty understanding SPiN can slowly learn to better identify acoustic cues in normal speech. Using features found in singing is a good starting point because natural speech already contains music-like features (Deutsch et al., 2011). Moreover, modifying certain acoustic features of target sounds that aid in their predictability can aid in their comprehension (Bendixen, 2014; Bregman, 1994), and one key feature of music is that it helps guide listener predictions (Koelsch et al., 2019). Accordingly, pairing musical features with speech could improve the predictability and therefore the comprehensibility of that speech. Additionally, understanding SPiN can be improved by music training (Coffey et al., 2017; Hennessy et al., 2022; Maillard et al., 2023; Zendel, 2022), suggesting that musicians may acquire some of these skills intrinsically through their music education. The next three sections will explore how acoustic features of singing could help aid SPiN.

Natural Speech Contains Music-Like Features

Music and speech are both forms of auditory communication that unfold over time. Patel (2008) highlights the importance of both pitch and rhythm for speech and music perception, but highlights that they are used and perceived differently for speech and music. In music, rhythms tend to be periodic, allowing for the perception of beats, and an organization of "stronger" and "weaker" beats into a metric hierarchy (London, 2002). While speech has a rhythm, it lacks strict periodicity and hierarchy (Patel, 2008). In music, melodies tend to be made up of a series of discrete notes from the chromatic scale (in Western Music) that are organized into tonal scales that allow melodies to have a tonal center (Patel, 2008). In speech, pitch changes are not discrete and do not have a tonal center (Patel, 2008). Accordingly, we will focus on two musical features that occur in speech: melody and rhythm. The connections between speaking and singing is highlighted in the speech-to-song illusion, where natural speech can be made to sound like it is sung by looping a short spoken phrase (Deutsch et al., 2011). This likely happens because the repetition allows for the listener to identify periodicity in the repeating

pattern, and to perceive a tonal center through repeated presentations of the same pitches. Thus, while speech and music are very different, it is not difficult to bridge the gap between them. One important question is if embellishing the pitch, or rhythmic aspects of speech can improve its comprehensibility when presented in noise. Previous work has shown that altering the acoustic features of speech presented in noise can either enhance or degrade speech comprehension, depending on the acoustic cue.

Physical Cues Can Aid in SPiN Comprehension

Predictability can aid in speech comprehension (Bendixen, 2014). For example, when older listeners were presented with SPiN, older adults were better at understanding and remembering sentence-final words that were predictable based on the content of the sentence compared to sentence-final words that were not predictable (Pichora-Fuller et al., 1995). In addition to semantics improving predictability, physical cues in target speech can increase its predictability, as these cues can guide a listeners' attentional focus. In addition, certain physical cues can enhance source separations due to interactions with how acoustic information is transduced, and processed along the auditory pathway. When the fundamental frequency (F0) for the target speech differs from the F0 of masking speech, there is an improvement in speech understanding, highlighting that the pitch of a voice can be used to aid in comprehension (Brokx & Nooteboom, 1982; Song et al., 2011; Stickney et al., 2007; Summers & Leek, 1998). This improvement may be due to enhanced predictability – that is, the target speech is in a different frequency range compared to the masking speech. Alternatively, the improvement may also be due to the speech and speech maskers no longer occupying the same critical band in the auditory pathway, facilitating subsequent perceptual segregation.

When the overall rate of speech is increased, the understanding of speech in background noise is reduced (Gordon-Salant & Fitzgibbons, 2004). Interestingly, when the background babble noise was not sped up to match the target speech, comprehension improved, suggesting that listeners can take advantage of a rate mismatch between the target and the babble noise in order to guide attention to important parts of the target speech that would otherwise be masked if the speech and masker were presented at the same rate (Gordon-Salant & Fitzgibbons, 2004). In an auditory stream segregation task, target tone regularity improved the ability to segregate simultaneous auditory streams (Rimmele et al., 2012). In speech perception, it has been shown that isochronous rhythms, both as a prime (Sidiras et al., 2017, 2020), and in the speech itself, improve the understanding of SPiN (Pearson et al., 2023). These results suggest that temporal predictability should aid in understanding SPiN, and imply that a musical rhythm added to speech should aid in comprehension. In fact, one of the main ideas presented in Bregman's seminal book, *Auditory Scene Analysis* was that the auditory system will use whatever acoustic cues are available to help parse the auditory scene into meaningful auditory streams (Bregman, 1994). Moreover, the brain entrains to isochronous musical rhythms (Nozaradan et al., 2011; Sauvé et al., 2019), which suggests that isochronous speech rhythms could be understood better because rhythmic speech would entrain the brain to oscillate at the rate of the speech (Peelle & Davis, 2012).

Despite these theoretical reasons, there is scant work demonstrating if speech isochrony aids or hinders SPiN. Natural speech is not isochronous, but when listening to

speech, the brain entrains to the rhythm of the speech, and the stronger this entrainment, the better the comprehension (Peelle & Davis, 2012). Isochronous auditory stimulation can evoke neural entrainment (Nozaradan et al., 2011; Sauvé et al., 2019), suggesting that isochronous speech would enhance neural entrainment, and therefore improve speech comprehension. Interestingly, some studies have reported that isochronous speech is more difficult to understand compared to natural speech because speech isochrony is "unnatural" (Aubanel & Schwartz, 2020). These results suggest that interfering with natural speech patterns would hinder comprehension. Given the paucity of literature in the field, it is critical to collect more data to better understand how isochrony, and other musical features impact speech comprehension in noise.

A few studies have explored the idea of using musical features to aid in separating simultaneous auditory inputs or understanding SPiN. In a study using two interleaved tone sequences, it was shown that adding both melody and rhythm aided in the perceptual segregation of the tone sequences (Szalárdy et al., 2014). Since segregation is a necessary pre-cursor to comprehension, this suggests that both melodic and rhythmic information could be used to aid in understanding SPiN. A recent study supports this idea by demonstrating that neural phase-locking to sung speech is better than to spoken speech (Vanden Bosch der Nederlanden et al., 2020). Despite strong theoretical and experimental support, the one study that explored how singing speech impacts the ability to comprehend SPiN in children, found that sung speech was more difficult to understand in noise (Nie et al., 2018). In support of this finding, it has been shown that altering the natural prosody of speech reduces comprehension when there is background noise (Binns & Culling, 2007; Miller et al., 2010). Overall, this complex pattern of results suggests that altering the physical cues of the target speech in a SPiN task could either enhance or reduce the ability to comprehend SPiN. Moreover, it suggests that pitch- and timing-based manipulations may differentially impact SPiN comprehension.

SPiN Comprehension Can Be Improved

While certain acoustic features have the potential to aid in the ability to understand SPiN, it is difficult to apply these findings in naturalistic environments because it can difficult, socially uncomfortable and in some cases physically impossible for a speaker to alter their speaking patterns. For example, it would be challenging, if not impossible, to speak with a perfectly isochronous rhythm or to speak louder in a social setting where only whispers are acceptable. Accordingly, it makes more sense to help the listener, rather than prescribe that speakers alter their speech. This leads to an important question: can the ability to understand SPiN be improved? The answer is very likely, yes. For example, a number of studies have shown that lifelong musicians have an enhanced ability to understand SPiN compared to non-musicians (e.g., Parbery-Clark et al., 2011; Zendel & Alain, 2012). Zendel and Alain (2012) tested participants across the lifespan, and found that the musician benefit for understanding SPiN was larger for older adults compared to younger adults. Although there have been some studies that have shown no difference between musicians and non-musicians on the ability to understand SPiN, they consistently show that on average, performance on SPiN tasks is similar or slightly better (but not statistically significant) in musicians compared to non-musicians (Boebinger et al., 2015; Madsen et al., 2019; Zendel & Alexander, 2020). To address these issues, recent

reviews and meta-analyses revealed that the musician benefit for SPiN comprehension was likely small, but real (Coffey et al., 2017; Hennessy et al., 2022; Maillard et al., 2023). One major issue with these studies is that musicians may become musicians because of preexisting advantages in hearing abilities; however, longitudinal studies tend to support the idea that music training improves the ability to understand SPiN (Dubinsky et al., 2019; Zendel, 2022; Zendel et al., 2019)

Music Perception and Aging

Given that the ability to understand SPiN can be modified by music training, it may be possible to develop more optimized forms of rehabilitation. One interesting pattern of results that has recently emerged is that music perception seems to be relatively preserved in older adults (Halpern et al., 2017; Lagrois et al., 2018; Sauvé, Bolt, et al., 2022; Sauvé et al., 2019; Sauvé, Marozeau, et al., 2022). In contradistinction, the ability to understand SPiN declines in older adults, and this decline is nearly universal (Gates & Mills, 2005; Pichora-Fuller et al., 2016). This pattern of results suggests that by connecting speech perception with music perception, it may be possible to preserve or enhance speech comprehension in older adults. Moreover, given the connection between music training and SPiN comprehension in older adults (Coffey et al., 2017; Dubinsky et al., 2019; Hennessy et al., 2022; Maillard et al., 2023; Zendel, 2022; Zendel & Alain, 2012; Zendel et al., 2019), focusing on musical features may engender resilience to further decline in speech comprehension. Accordingly, it may be possible to develop adaptive forms of auditory rehabilitation that rely on intact music perception (i.e., a cognitive strength) on which SPiN comprehension (i.e., a cognitive weakness) can be scaffolded, and enhanced with music-based training.

Summary

Putting it all together, previous research suggests that speech contains acoustic features, that when exaggerated, could be considered musical; that altering acoustic features of target speech can aid in understanding SPiN; that the ability to understand SPiN can be improved with training; and that perception of music is relatively preserved in older adults, despite a nearly universal decline in SPiN abilities in older adults. This suggests that musical acoustic features could be a source of cognitive strength for older adults, and that some of these features could be exaggerated in natural speech to aid in understanding SPiN. Over time, slowly removing the musical features from speech and working with listeners to better focus attention toward those features in natural speech could help people learn to better understand SPiN. One question that remains poorly understood is if musical acoustic features aid in understanding SPiN. Accordingly, the goal of the current study was to determine if musical acoustic features can aid in understanding SPiN.

Methods

Participants

A total of 15 participants (9 women, 6 men) completed the study and provided written informed consent in accordance with the Heath Research Ethics Authority of

Newfoundland and Labrador. They ranged in age from 22 to 29 ($M = 24.73$), and had normal hearing (i.e., pure-tone average < 25 dB HL). Participants were all self-reported non-musicians (i.e., <1 year of music training or not having played a musical instrument for more than 10 years). Finally, participants all reported that English was their first language.

Procedure

After obtaining written informed consent, participants completed a short demographic questionnaire, and a pure-tone threshold assessment in a double-walled sound-attenuating booth using an Interacoustics AC40 Clinical Audiometer. Next, participants completed the experimental portion of the study.

All experimental stimuli (described below) were presented to the participant through ER3A insert earphones, while participants were seated in a double walled sound-attenuating booth. Participants were familiarized with the task before they began. In all conditions, participants were asked to repeat a short sentence aloud after they heard it. They were told to prioritize accuracy, and to repeat exactly what they heard. The accuracy of the repetition was judged online by a native English speaker who had the text of the sentence in front of them. A 1-down, 1-up adaptive procedure was used in order to identify the 50% threshold for speech comprehension (Levitt, 1971). In the current study, the first sentence was presented at +15 dB SNR along with four-talker babble noise presented at a combined amplitude of 65 dB SPL. After the participant correctly repeated the entire sentence, the amplitude of the target sentence was reduced in 4 dB steps until they did not repeat the sentence correctly, at which point the amplitude of the target sentence was increased in 4 dB steps. This procedure was then repeated. After two reversals the step-size was reduced to 2 dB. In each block, a total of 20 sentences were played, and the 50% threshold was determined by averaging the reversal points.

Stimuli

One of the goals of this study was to maximize ecological validity so that the findings would more likely generalize to a rehabilitation program. Accordingly, materials were recorded by a human specifically for this project. The content of the target sentences came from "Harvard sentences," which by design are intended to be phonetically balanced, thereby reducing the subjectivity of a sentence's intelligibility (Rothauser, 1969). Of this collection, the first 240 sentences were selected and recorded. All recordings were made in a double-walled sound attenuating booth, using an Audio-Technica 4040 Condenser microphone, and were normalized to −12 dBfs RMS. This correction was made so that the target stimuli were of similar loudness. We acknowledge that the RMS normalization may not always provide for equal loudness perception (e.g., Zhang & Zeng, 1997); however, it did offer an obvious improvement in loudness equalization compared to the raw recordings. All target sentences were recorded by a professional singer/vocalist (soprano) completing her fourth year of a bachelor of music performance in voice. Each of the 240 sentences was recorded four times. For the *Spoken* condition the singer was instructed to read the sentences as naturally as possible. For the other three conditions, the goal was to "musicialize" the sentences. For the *Rapped* condition an

estimated tempo was determined for each of the recorded sentences in the spoken condition. This estimation was done by determining how many "beats" there were in the sentence, and how long the recording of the spoken sentence was. A metronome was then played to the vocalist through headphones, and she was instructed to rap the sentence so that the syllables lined up with the metronome clicks, while attempting to maintain the same pitch changes as the spoken sentence. The singer was able to listen to her previous recordings in order to make them as similar as possible. For the *Sung* condition, the singer listened to the previous sets of recordings, and approximated the natural pitch changes into discrete musical notes that formed a melody. First, she listened to the *Spoken* version of the target stimuli, then crafted a major key melody that was of similar contour to the *Spoken* sentence. A musical score was created for each sentence. During the recording, the singer heard the same metronome as the *Rapped* condition, but had the musical score in front of her so she could sing the sentences. Finally, for the *No Rhythm* condition, the singer was asked to sing each sentence using the same melody as the *Sung* condition, but to use the same rhythm as the *Spoken* condition. She was able to see the score for each sentence, and was able to listen to the *Spoken* version of the sentence in order to match the timing accents of the *Spoken* stimuli. The authors and singer verified that each sentence matched the condition to which it was assigned (e.g., a rapped sentence didn't have a melodic contour). A sample spectrogram from each of the conditions can be seen in Figure 1.

There were a total of 240 unique sentences recorded in all four conditions: Spoken, Sung, Rapped, and No Rhythm. The 240 sentences were divided into 12 lists of 20 sentences each. This list size was chosen based on the development of the HINT and HINT for French-Canadian populations (Nilsson et al., 1994; Vaillancourt et al., 2005) where they used lists of 10 and 20 sentences, respectively, to determine speech-in-noise thresholds. This yielded 4 versions of each list, Spoken, Sung, Rapped, and No Rhythm. For example, List 1 Sung had the exact same sentences as List 1 Spoken, only the vocal delivery was altered. Participants heard a total of 8 lists, 2 from each listening condition. They were presented in a pseudo-random order, so that participants heard 1 list from each condition once before hearing the same condition again. Importantly, participants never heard the same sentence more than once throughout the study.

Data Analysis

The data were analyzed using a 2×2 factorial ANOVA. For the analysis, we considered two variables: Rhythm (Musical or Spoken), and Pitch (Musical or Spoken). The Spoken condition had both spoken rhythm and spoken pitch, and the Sung condition had a musical rhythm and musical pitch. The Rapped condition had a musical rhythm, with a spoken pitch, while the No Rhythm condition had a musical pitch with a spoken rhythm. The 50% SNR threshold from each condition was used as the dependent measure.

Results

Data was analyzed using a 2 (Rhythm: Music vs Speech) ×2 (Pitch: Music vs Speech) Repeated Measures ANOVA. Overall, there was a Main Effect of Rhythm, F (1, 14), 9.43,

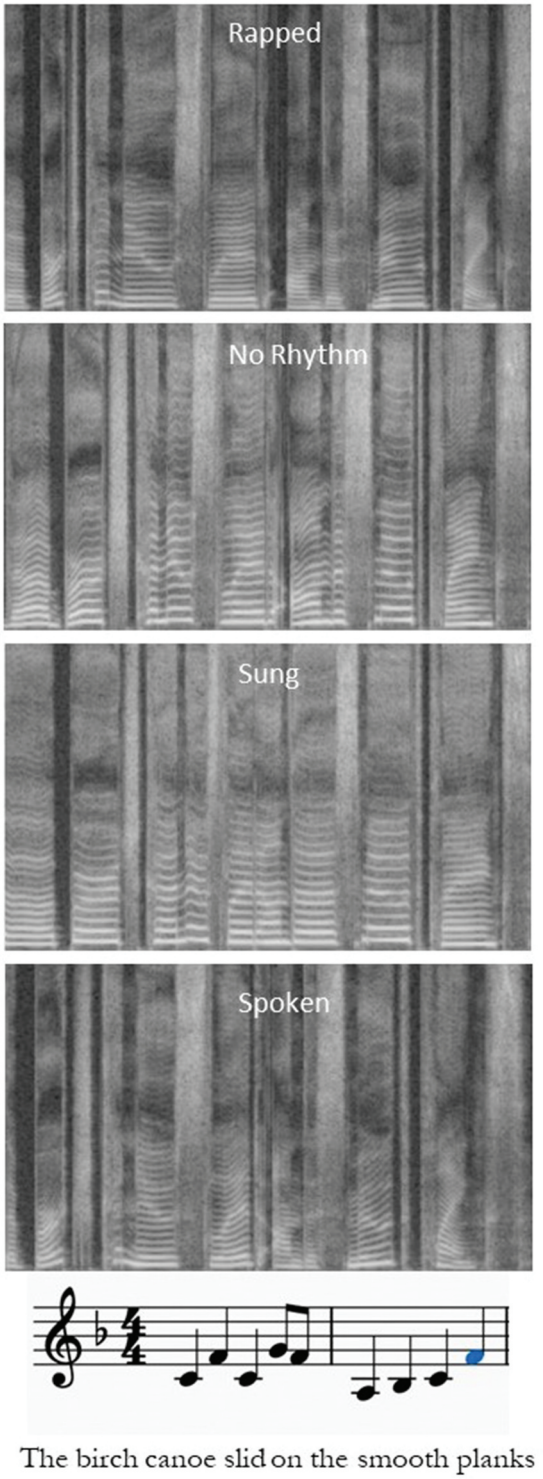

Figure 1. Sample spectrogram from the four conditions. Frequencies up from 1 Hz –10 kHz are shown on each plot.

$p = .008$, $\eta^2 = 0.13$, no Main Effect of Pitch $F(1, 14) = 0.26$, $p = .62$, and most importantly, a significant Pitch by Rhythm Interaction, $F(1, 14) = 35.98$, $p < .001$, $\eta^2 = .36$ (Figure 2, Table 1). Follow-up pairwise comparisons revealed that when the target speech had a Musical Pitch the 50% SNR threshold was lower (i.e., better performance) when it's rhythm was speech-like compared to when its rhythm was musical ($p < .001$). At the same time, when target speech had a Speech-like Pitch, there was no difference in the 50% SNR threshold between Musical Rhythm and Speech Rhythm ($p = .15$). Most critical, the 50% SNR threshold was lower (i.e., better performance) in the No Rhythm condition (i.e., Musical Pitch/Speech Rhythm) compared to the Spoken condition ($p = .005$), while 50% SNR threshold for the Sung speech was higher (i.e., worse performance) compared to Spoken speech ($p = .047$).

Given the focus on ecological validity in this study, it was difficult to precisely control how the singer produced each sentence in each condition. It is therefore possible that a 2 × 2 design would be inappropriate because the main vocal manipulation (spoken vs. sung melody; spoken vs. sung rhythm), could not be precisely manipulated, and thus may not be properly crossed. Accordingly, the data was also quantified using a 1-way repeated measures ANOVA with each Listening Condition

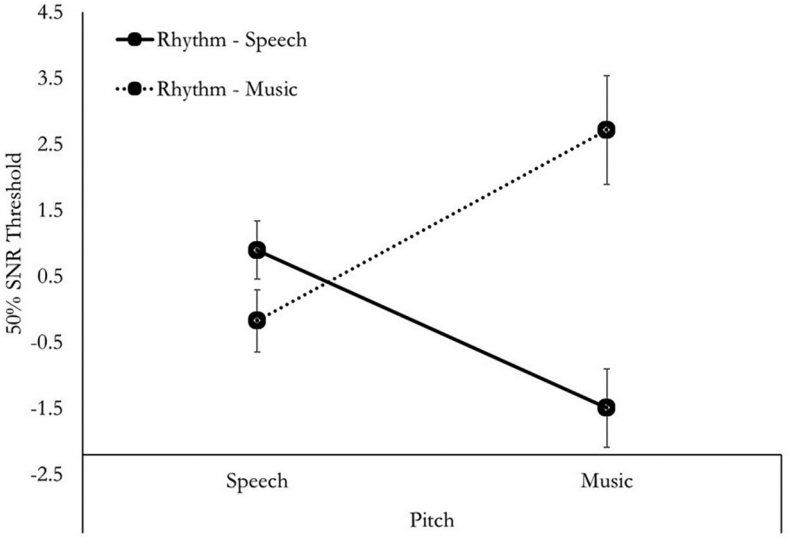

Figure 2. 50% SNR threshold as a function of the properties of the target speech. Thresholds were lowest (i.e., easiest to understand) when target speech had music-like pitch, but speech-like rhythm. Thresholds were highest when target speech was sung, that is, had a music-like pitch and music-like rhythm.

Table 1. 50% SNR threshold: means and standard deviations (SD) for each condition.

Condition	50% SNR threshold	SD
Spoken	0.897	1.714
Sung	2.719	3.186
Rapped	−0.175	1.833
No Rhythm	−1.494	2.296

as a separate factor. This analysis revealed a similar pattern of results. There was a main effect of Listening Condition, F (3,42) = 13.65, $p < .001$, $\eta^2 = 0.49$. Follow-up tests revealed that the "No Rhythm" Condition was significantly better understood compared to both the Sung ($p < .001$), and Spoken ($p = .004$), but not Rapped ($p = .12$). The Rapped condition was significantly better understood compared to the Sung ($p < .001$), but not spoken ($p = .123$). Spoken was significantly better understood than Sung ($p = .032$).

In order to explore if individual variability could predict how well one understands SPiN (i.e., the Spoken condition), a linear regression was calculated that included Pure-tone Average (average of PTT from 500, 1000, 2000 and 4000 Hz), Age, Years of Education, Years of Music Training as predictors. The model was not significant ($p = .96$). To determine if individual differences could account for performance in the other experimental conditions, linear regressions were calculated for the Rapped, No Rhythm and Sung conditions. None of the models were significant ($p = .16$, .72, .17, respectively).

In order to determine how performance in each of the experimental conditions related to each other, a series of bivariate correlations was calculated between each condition. Interestingly, the performance in the Spoken condition was unrelated to any other condition ($ps = .32–.44$). At the same time, all the conditions with musical features were weakly correlated ($ps = .03–.09$). Scatterplots showing the relationship between each condition are presented in Figure 3. As a next step, we calculated a regression analysis with the Sung speech as a dependent, and the other two conditions with musical features (Rapped, No Rhythm) conditions as independent factors. Overall, the regression was significant ($p = .002$); however, neither factor made a significant independent contribution to performance when the target speech was Sung ($p = .36$ & .12, respectively),

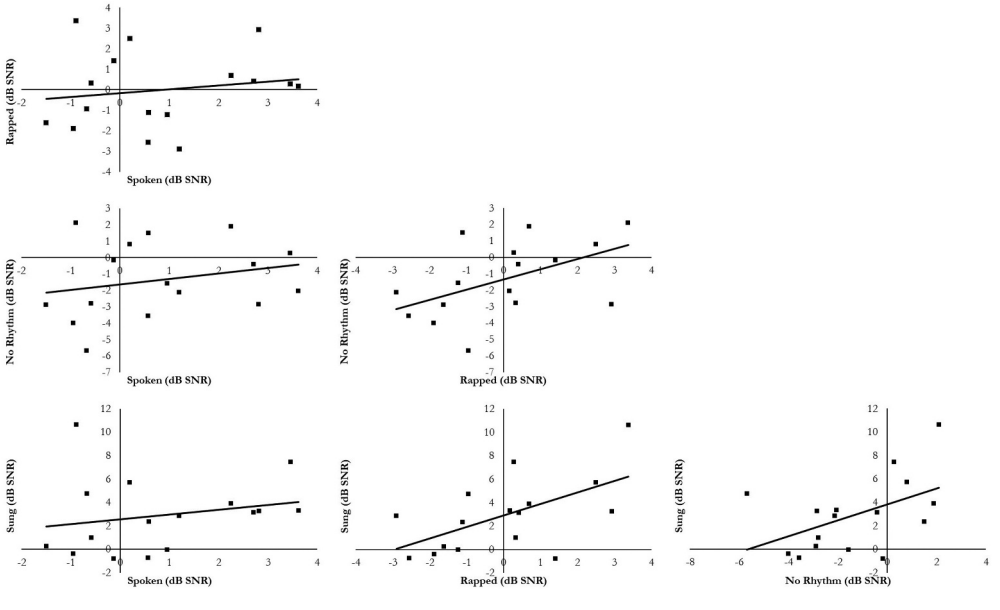

Figure 3. Scatterplots showing the relationship between performance (50% SNR threshold) in each of the listening conditions (spoken, sung, rapped, no rhythm).

suggesting that the impact of adding musical features to target stimuli was similar across participants, and unrelated to their ability to perform in the Spoken condition.

To determine if our analyses were appropriately powered, we calculated a power analysis for both the ANOVA and the regression reported above. The power calculation for the ANOVA was .813, while the power for the regression was .526. Accordingly, the study was appropriately powered for the ANOVA analyses, but may have been underpowered for the regression analysis. This could explain why neither factor in the regression made significant contributions to the overall effect.

Discussion

Overall, musical features in speech impacted on the ability to understand speech when presented in multi-talker babble noise. Adding musical pitch information (i.e., a melody) helped listeners understand SPiN when the speech had a natural speech rhythm. At the same time, adding musical pitch information reduced the ability to understand SPiN when the speech was presented with a musical (i.e., isochronous) rhythm. This pattern of results suggests that musical melodic cues can be used to help understand SPiN, but that the combination of multiple musical acoustic cues do not add up, and in the current study, demonstrate that multiple cues can reduce comprehensibility. The next sections will discuss why this might be the case will conclude with a discussion on how this finding could benefit the development of auditory rehabilitation programs for SPiN comprehension.

Overall, musical pitch contours (i.e., melody) aided SPiN comprehension, but only when the speech had a natural speech rhythm. One interesting parallel to this type of speech is "elderspeak" (ES), infant-directed speech (IDS), or Clear Speech (CS) where the speaker makes acoustic accommodations to aid listener comprehension (Krause & Braida, 2004; Shaw et al., 2021; Uchanski, 2005). Importantly, ES is considered inappropriate and infantilizing when used with older adults who have perceived hearing loss or cognitive decline because it sounds like IDS; however, in some cases it does improve comprehension (McGuire et al., 2001; Shaw et al., 2021). For example, in IDS, talkers increase the overall variability of F0, increasing the pitch of F0, and increasing the duration of F0 on each phoneme, which aids in comprehention for infants (Audibert & Falk, 2018; Cox et al., 2022). Moreover, IDS is observed across many cultures, and is thought to be a critical part of language learning (Cox et al., 2022). While ES and IDS are used to help listeners with comprehension issues due to their age, CS is used to help all listeners with hearing impairments (Uchanski, 2005). Some acoustic features of CS include a slowed speaking rate, increased acoustic energy between 1 and 3 kHz, increased low-frequency modulation depth, greater F0 range, increased acoustic energy near the second and third formants and increased voice onset time (VOT); however, CS is usually spoken more naturally compared to both ES and IDS (Krause & Braida, 2004). When compared to normal speech, singing is fundamentally different, with singing generally having longer vowel duration, a higher F0, and more stable F0 contours on individual vowels compared to natural speech (Audibert & Falk, 2018; Cox et al., 2022). Accordingly, it seems like some song-like features are automatically added to speech in order to aid comprehension for listeners with comprehension difficulties (i.e., both infants and people with presumed cognitive or hearing deficits). It is possible that by

adding melodic information to speech, while maintaining a speech-like rhythm, that we generated stimuli that were similar to IDS, ES or CS.

While the addition of melodic information with speech-like rhythm aided in comprehension, the addition of melodic information with a musical rhythm reduced comprehension. This pattern was difficult to predict from previous work as some studies demonstrated both melodic and rhythmic advantages for detecting signals or speech-in-noise (Szalárdy et al., 2014; Vanden Bosch der Nederlanden et al., 2020), while others have shown that singing (Nie et al., 2018), or altering natural speech rhythms reduce the ability to understand SPiN (Binns & Culling, 2007; Miller et al., 2010). The results of the current study potentially bridge the gap between the seemingly contradictory results of previous studies. Adding a melody to speech allows for better neural tracking of F0, as seen in Vanden Bosch der Nederlanden et al. (2020). While Szalárdy et al., (2014) showed benefits for both adding both rhythm and melody to a stream segregation task, the results of Binns and Culling (2007), Miller et al. (2010), and Nie et al. (2018) suggest that when a streaming task involves natural speech, that our auditory systems expect to have a natural speech rhythm, but is more flexible when processing unexpected pitch contours. That is, in the case of speech, the brain cannot use the added acoustic features of rhythmicity to aid in segregation because musical rhythmicity interferes with speech processing. The results from the current study provide empirical support for this interference effect, while demonstrating that melodic information alone is beneficial.

One interesting relationship observed in the data was a correlation in performance between all three of the "musical" conditions (i.e., Sung, Rapped, No Rhythm), but not the Spoken condition, suggesting that listening to musicalized speech is fundamentally different than listening to natural speech. This musical mode of listening may explain why musicians, and particularly older musicians, are better able to understand speech when there is background noise (Coffey et al., 2017; Hennessy et al., 2022; Zendel & Alain, 2012; Zendel et al., 2019). It is possible that musicians are able to engage a musical listening mode when listening to SPiN. Given that the No Rhythm condition was the easiest to comprehend in the current study, this further suggests that tracking pitch would be important when applying this mode of listening to speech. In support of this proposal is evidence that sub-cortical neurons show stronger phase locking to speech stimuli in musicians compared to non-musicians (Dubinsky et al., 2019; Musacchia et al., 2007), suggesting that musical training may help speech perception via a musical mode of listening during natural speech (Vanden Bosch der Nederlanden et al., 2020). The current results suggest that supplemental melodic cues in target speech can intrinsically improve speech understanding. However, one major limitation of this study was that we only used a single target voice for all stimuli, therefore it is possible that the effects observed here were due to features unique to her voice.

One of the most important aspects of the current finding is that melodic cues are already present in natural speech, as demonstrated by the speech-to-song illusion (Deutsch et al., 2011). Thus, a rehabilitation program could develop materials with longer vowel durations, and more stable F0 contours on individual vowels as starting points for auditory rehabilitation. If listeners were trained to focus on these features in exaggerated speech stimuli, and then have the features slowly removed, it is possible that the listener would learn to better identify those features in natural speech, which would then improve their ability to understand SPiN.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

The study was supported by the Canada Research Chair program (BRZ), and the NSERC Discovery program (BRZ).

References

Albert, M. L., Sparks, R. W., & Helm, N. A. (1973). Melodic intonation therapy for aphasia. *Archives of Neurology*, *29*(2), 130–131. https://doi.org/10.1001/archneur.1973.00490260074018

Aubanel, V., & Schwartz, J.-L. (2020). The role of isochrony in speech perception in noise. *Scientific Reports*, *10*(1), 19580. https://doi.org/10.1038/s41598-020-76594-1

Audibert, N., & Falk, S. (2018). Vowel space and F0 characteristics of infant-directed singing and speech. *Speech Prosody*, *2018*, 153–157. https://doi.org/10.21437/SpeechProsody.2018-31

Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00060

Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America*, *122*(3), 1765–1776. https://doi.org/10.1121/1.2751394

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. https://doi.org/10.1121/1.4904537

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. MIT press.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*(1), 23–36. https://doi.org/10.1016/S0095-4470(19)30909-X

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research*, *352*, 49–69. https://doi.org/10.1016/j.heares.2017.02.006

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G., & Fusaroli, R. (2022). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behaviour*, *7*(1), 114–133. https://doi.org/10.1038/s41562-022-01452-1

Deutsch, D., Henthorn, T., & Lapidis, R. (2011). Illusory transformation from speech to song. *Journal of the Acoustical Society of America*, *129*(4), 2245–2252. https://doi.org/10.1121/1.3562174

Dubinsky, E., Wood, E. A., Nespoli, G., & Russo, F. A. (2019). Short-term choir singing supports speech-in-noise perception and neural pitch strength in older adults with age-related hearing loss. *Frontiers in Neuroscience*, *13*(1153), 1–18. https://doi.org/10.3389/fnins.2019.01153

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet*, *366*(9491), 1111–1120. https://doi.org/10.1016/S0140-6736(05)67423-5

Gordon-Salant, S., & Fitzgibbons, P. J. (2004). Effects of stimulus and noise rate variability on speech perception by younger and older adults. *The Journal of the Acoustical Society of America*, *115*(4), 1808–1817. https://doi.org/10.1121/1.1645249

Halpern, A. R., Zioga, I., Shankleman, M., Lindsen, J., Pearce, M. T., & Bhattarcharya, J. (2017). That note sounds wrong! Age-related effects in processing of musical expectation. *Brain & Cognition*, *113*, 1–9. https://doi.org/10.1016/j.bandc.2016.12.006

Hennessy, S., Mack, W. J., & Habibi, A. (2022). Speech-in-noise perception in musicians and non-musicians: A multi-level meta-analysis. *Hearing Research*, *416*, 108442. https://doi.org/10.1016/j.heares.2022.108442

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Krause, J. C., & Braida, L. D. (2004). Acoustic properties of naturally produced clear speech at normal speaking rates. *The Journal of the Acoustical Society of America*, *115*(1), 362–378. https://doi.org/10.1121/1.1635842

Lagrois, M.-É., Peretz, I., & Zendel, B. R. (2018). Neurophysiological and behavioral differences between older and younger adults when processing violations of tonal structure in music. *Frontiers in Neuroscience*, *12*(54), 1–15. https://doi.org/10.3389/fnins.2018.00054

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*(2B), 467–477. https://doi.org/10.1121/1.1912375

London, J. (2002). Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception*, *19*(4), 529–550. https://doi.org/10.1525/mp.2002.19.4.529

Madsen, S. M. K., Marschall, M., Dau, T., & Oxenham, A. J. (2019). Speech perception is similar for musicians and non-musicians across a wide range of conditions. *Scientific Reports*, *9*(1), 10404. https://doi.org/10.1038/s41598-019-46728-1

Maillard, E., Joyal, M., Murray, M. M., & Tremblay, P. (2023). Are musical activities associated with enhanced speech perception in noise in adults? A systematic review and meta-analysis. *Current Research in Neurobiology*, *4*, 100083. https://doi.org/10.1016/j.crneur.2023.100083

McCormack, A., & Fortnum, H. (2013). Why do people fitted with hearing aids not wear them? *International Journal of Audiology*, *52*(5), 360–368. https://doi.org/10.3109/14992027.2013.769066

McGuire, L. C., Morian, A., Codding, R., & Smyer, M. A. (2001). Older adults' memory for medical information: Influence of elderspeak and note taking. *International Journal of Rehabilitation and Health*, *12*. https://doi.org/10.1023/A:1012906222395

Merrett, D. L., Peretz, I., & Wilson, S. J. (2014). Neurobiological, cognitive, and emotional mechanisms in melodic intonation therapy. *Frontiers in Human Neuroscience*, *8*, 401. https://doi.org/10.3389/fnhum.2014.00401

Miller, S. E., Schlauch, R. S., & Watson, P. J. (2010). The effects of fundamental frequency contour manipulations on speech intelligibility in background noisea. *Journal of the Acoustical Society of America*, *128*(1), 435–443. https://doi.org/10.1121/1.3397384

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(40), 15894–15898. https://doi.org/10.1073/pnas.0701498104

Nie, Y., Galvin, J. J., Morikawa, M., André, V., Wheeler, H., & Fu, Q.-J. (2018). Music and speech perception in children using sung speech. *Trends in Hearing*, *22*, 233121651876681. https://doi.org/10.1177/2331216518766810

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Norton, A., Zipse, L., Marchina, S., & Schlaug, G. (2009). Melodic intonation therapy: Shared insights on how it is done and why it might help. *Annals of the New York Academy of Sciences*, *1169*(1), 431–436. https://doi.org/10.1111/j.1749-6632.2009.04859.x

Nozaradan, S., Peretz, I., Missal, M., & Mouraux, A. (2011). Tagging the neuronal entrainment to beat and meter. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(28), 10234–10240. https://doi.org/10.1523/JNEUROSCI.0411-11.2011

Parbery-Clark, A., Strait, D. L., 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*(5), e18082. https://doi.org/10.1371/journal.pone.0018082

Patel, A. D. (2008). *Music, languange, and the brain*. Oxford University Press.

Pearson, D. V., Shen, Y., McAuley, J. D., & Kidd, G. R. (2023). The effect of rhythm on selective listening in multiple-source environments for young and older adults. *Hearing Research*, *435*, 108789. https://doi.org/10.1016/j.heares.2023.108789

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. https://doi.org/10.3389/fpsyg.2012.00320

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear and Hearing*, *37*(1), 5S–27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, *97*(1), 593–608. https://doi.org/10.1121/1.412282

Rimmele, J., Schröger, E., & Bendixen, A. (2012). Age-related changes in the use of regular patterns for auditory scene analysis. *Hearing Research*, *289*(1–2), 98–107. https://doi.org/10.1016/j.heares.2012.04.006

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246.

Sauvé, S. A., Bolt, E. L. W., Fleming, D., & Zendel, B. R. (2019). The impact of aging on neurophysiological entrainment to a metronome. *NeuroReport*, *30*(10), 730–734. https://doi.org/10.1097/WNR.0000000000001267

Sauvé, S. A., Bolt, E. L. W., Nozaradan, S., & Zendel, B. R. (2022). Aging effects on neural processing of rhythm and meter. *Frontiers in Aging Neuroscience*, *14*, 848608. https://doi.org/10.3389/fnagi.2022.848608

Sauvé, S. A., Marozeau, J., & Rich Zendel, B. (2022). The effects of aging and musicianship on the use of auditory streaming cues. *PLOS ONE*, *17*(9), e0274631. https://doi.org/10.1371/journal.pone.0274631

Shaw, C. A., Gordon, J. K., & Lee, M.-A. (2021). Understanding elderspeak: An evolutionary concept analysis. *Innovation in Aging*, *5*(3), igab023. https://doi.org/10.1093/geroni/igab023

Sidiras, C., Iliadou, V., Nimatoudis, I., Reichenbach, T., & Bamiou, D.-E. (2017). Spoken word recognition enhancement due to preceding synchronized beats compared to unsynchronized or unrhythmic beats. *Frontiers in Neuroscience*, *11*, 415. https://doi.org/10.3389/fnins.2017.00415

Sidiras, C., Iliadou, V. V., Nimatoudis, I., & Bamiou, D.-E. (2020). Absence of rhythm benefit on speech in noise recognition in children diagnosed with auditory processing disorder. *Frontiers in Neuroscience*, *14*, 418. https://doi.org/10.3389/fnins.2020.00418

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: Neural correlates. *Journal of Cognitive Neuroscience*, *23*(9), 2268–2279. https://doi.org/10.1162/jocn.2010.21556

Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F.-G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentencesa. *Journal of the Acoustical Society of America*, *122*(2), 1069–1078. https://doi.org/10.1121/1.2750159

Summers, V., & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1294–1306. https://doi.org/10.1044/jslhr.4106.1294

Szalárdy, O., Bendixen, A., Böhm, T. M., Davies, L. A., Denham, S. L., & Winkler, I. (2014). The effects of rhythm and melody on auditory stream segregation. *The Journal of the Acoustical Society of America*, *135*(3), 1392–1405. https://doi.org/10.1121/1.4865196

Uchanski, R. M. (2005). Clear speech. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 207–235). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470757024.ch9

Vaillancourt, V., Laroche, C., Mayer, C., Basque, C., Nali, M., Eriks-Brophy, A., Soli, S. D., & Giguère, C. (2005). Adaptation of the HINT (hearing in noise test) for adult Canadian francophone populations. *International Journal of Audiology*, *44*(6), 358–361. https://doi.org/10.1080/14992020500060875

Vanden Bosch der Nederlanden, C. M., Joanisse, M. F., & Grahn, J. A. (2020). Music as a scaffold for listening to speech: Better neural phase-locking to song than speech. *NeuroImage*, *214*, 116767. https://doi.org/10.1016/j.neuroimage.2020.116767

Zendel, B. R. (2022). The importance of the motor system in the development of music-based forms of auditory rehabilitation. *Annals of the New York Academy of Sciences*, *1515*(1), 10–19. https://doi.org/10.1111/nyas.14810

Zendel, B. R., & Alain, C. (2012). Musicians experience less age-related decline in central auditory processing. *Psychology and Aging*, *27*(2), 410–417. https://doi.org/10.1037/a0024816

Zendel, B. R., & Alexander, E. J. (2020). Autodidacticism and music: Do self-taught musicians exhibit the same auditory processing advantages as formally trained musicians? *Frontiers in Neuroscience*, *14*, 752. https://doi.org/10.3389/fnins.2020.00752

Zendel, B. R., West, G. L., Belleville, S., & Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiology of Aging*, *81*, 102–115. https://doi.org/10.1016/j.neurobiolaging.2019.05.015

Zhang, C., & Zeng, F.-G. (1997). Loudness of dynamic stimuli in acoustic and electric hearinga. *The Journal of the Acoustical Society of America*, *102*(5), 2925–2934. https://doi.org/10.1121/1.420347

Zumbansen, A., Peretz, I., & Hébert, S. (2014). Melodic intonation therapy: Back to basics for future research. *Frontiers in Neurology*, *5*, 1–11. https://doi.org/10.3389/fneur.2014.00007