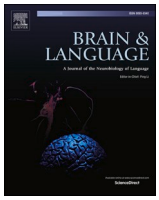
Brain & Language 266 (2025) 105581

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l



Attention, musicality, and familiarity shape cortical speech tracking at the musical cocktail party☆

Jane A. Brown [a,b], Gavin M. Bidelman [c,d,e,*]

[a] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[b] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[c] Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA
[d] Program in Neuroscience, Indiana University, Bloomington, IN, USA
[e] Cognitive Science Program, Indiana University, Bloomington, IN, USA

ARTICLE INFO

Keywords:
Speech-in-noise
Cocktail party
Background music
Familiarity

ABSTRACT

The "cocktail party problem" challenges our ability to understand speech in noisy environments and often includes background music. Here, we explored the role of background music in speech-in-noise listening. Participants listened to an audiobook in familiar and unfamiliar music while tracking keywords in either speech or song lyrics. We used EEG to measure neural tracking of the audiobook. When speech was masked by music, the modeled temporal response function (TRF) peak latency at 50 ms ($P1_{TRF}$) was prolonged compared to unmasked. Additionally, $P1_{TRF}$ amplitude was larger in unfamiliar background music, suggesting improved speech tracking. We observed prolonged latencies at 100 ms ($N1_{TRF}$) when speech was not the attended stimulus, though only in less musical listeners. Our results suggest early neural representations of speech are stronger with both attention and concurrent unfamiliar music, indicating familiar music is more distracting. One's ability to perceptually filter "musical noise" at the cocktail party also depends on objective musical listening abilities.

1. Introduction

Background music is a major part of our everyday listening experiences. Listening to music affects our in-store and online shopping behaviors (Ding & Lin, 2012; Garlin & Owen, 2006; North et al., 1999), driving performance (Beh & Hirst, 1999; Cassidy & MacDonald, 2009; Wang et al., 2015), and athletic performance (Atkinson et al., 2004; Chtourou et al., 2012). Listening to speech in background music, however, presents challenges due to the "cocktail party" phenomenon (Cherry, 1953; Haykin & Chen, 2005), in which the listener must attend to one source of auditory input while ignoring competing noise. Listeners can do this by separating the auditory scene into streams in order to isolate the target from non-target information (Bregman, 1990).

1.1. Effects of background music on speech perception

The impact of background music on concurrent speech or related cognitive tasks is somewhat ambiguous. Background music has been shown to increase listening effort (Du et al., 2020) and performance (Perham & Currie, 2014) on reading comprehension tasks, though other studies have shown no detrimental effect of background music on verbal learning (Jäncke & Sandmann, 2010). Similarly, a meta-analysis (Kämpfe et al., 2011) showed no overall impact of background music on adult listeners across several behavioral domains. While this is in part due to the heterogeneity in experiments investigating background music, it is also worth noting there is significant individual variability. Listeners who prefer not to listen to music while studying showed poorer reading comprehension (Etaugh & Ptasnik, 1982; Johansson et al., 2011) and more susceptibility to tempo changes in background music (Su et al., 2023). Comprehension was impaired with background music in learners with lower working memory capacity (Lehmann & Seufert, 2017). These effects also vary depending on listener personality (Avila et al., 2012; Furnham & Allass, 1999; Furnham & Strbac, 2002), music genre (Angel et al., 2010; Rea et al., 2010), types of musical training (Caldwell & Riby, 2007), and characteristics of the music, such as tempo and volume (Hine et al., 2022; Thompson et al., 2012).

☆ This article is part of a special issue entitled: 'Speech in Noise' published in Brain and Language.

* Corresponding author at: Department of Speech, Language and Hearing Sciences, Indiana University, 2631 East Discovery Parkway, Bloomington IN 47408, USA.
E-mail address: gbidel@iu.edu (G.M. Bidelman).

https://doi.org/10.1016/j.bandl.2025.105581
Received 12 October 2023; Received in revised form 21 December 2024; Accepted 15 April 2025
Available online 25 April 2025

1.2. Acoustic features

Acoustic features account for a wide range of auditory masking effects and therefore may also influence whether background music hinders speech perception. Intelligibility of conversational speech is worse in piano music played in a low octave and at a faster tempo (Ekström & Borg, 2011), consistent with well-known asymmetries in psychoacoustical masking. Similarly, reading comprehension is worse during very high tempo and louder music (Thompson et al., 2011). This is likely due to the arousal-mood hypothesis (Husain et al., 2002; Thompson et al., 2001), where task performance improves when music increases arousal (and thus induces more positive mood) up to a point, but can then oversaturate, creating states of overarousal that impair performance (Unsworth & Robison, 2016; Yerkes & Dodson, 1908). North and Hargreaves (1999) suggested that high-arousal music requires more cognitive resources than less arousing music. Given that the brain is a limited capacity system, more cognitive resources allocated to music listening means there would be fewer resources left to carry out any concurrent tasks (i.e., speech perception). As a result, cognitive task performance should be worse when background music significantly increases listener arousal.

1.3. Vocals

Evidence that background music with vocals impairs concurrent linguistic tasks is clearer. This is likely due, in part, to informational masking, where even unattended sounds in the same domain (e.g., speech on speech) can interfere with target recognition due to lexical interference. Indeed, people tend to listen to instrumental background music while studying or reading, but choose vocal music while driving or performing non-linguistic tasks (Kiss & Linnell, 2022). Music with vocals impaired performance on linguistic tasks (Brown & Bidelman, 2022a, 2022b; Crawford & Strapp, 1994; Scharenborg & Larson, 2018) and immediate recall in learning foreign language tasks (de Groot & Smedinga, 2014). Importantly, this type of informational masking only occurs when the interfering stream is understood by the listener. Brouwer et al. (2021) showed that an English masker impaired speech intelligibility more than "Simlish," a fictional gibberish language that shares phonemic patterns with English but lacks semantic meaning. Collectively, these studies suggest that the linguistic status of the background music and degree to which it carries lexical cues can modulate concurrent speech recognition.

1.4. Familiarity

Evidence for the role of familiarity in background music is also mixed. Several studies report better performance on language and speech tasks in the presence of familiar compared to unfamiliar background music (Brown & Bidelman, 2022a; Feng & Bidelman, 2015; Russo & Pichora-Fuller, 2008). Presumably, if the listener knows the music, the sequence of the song is predictable, which aids in auditory streaming (Bendixen, 2014). Similarly, if the listener already has mental representations of the familiar music, fewer cognitive resources are needed to process the masking stream, and listeners can more easily "tune it out" to prevent interference (Russo & Pichora-Fuller, 2008). Indeed, stream segregation may be easier when the attended and/or the unattended stimuli are more predictable (reviewed in Alain & Arnott, 2000; Jones et al., 1981; Shi & Law, 2010).

However, other studies report more detrimental effects of *familiar* music (Brown & Bidelman, 2022b; de Groot & Smedinga, 2014). Such effects are difficult to explain under the aforementioned arousal hypothesis (Husain et al., 2002) and instead may reflect the redirecting of limited cognitive resources and/or attentional mechanisms (e.g., Lavie, 2005). Familiar music can also provoke autobiographical memories (Belfi et al., 2016; Castro et al., 2020; Janata et al., 2007) and evoke musical imagery (Halpern & Zatorre, 1999; Zatorre & Halpern, 2005), siphoning cognitive resources away from the primary task, and ultimately impairing performance. In support of this notion, we recently demonstrated that speech intelligibility was worse when concurrent background music was familiar to the listener, regardless of whether it contained vocals (Brown & Bidelman, 2022b). The further impairment from vocal music was expected due to informational/linguistic masking.

1.5. Musicianship and speech-in-noise (SIN) processing

Another important factor shown to impact cocktail party and SIN listening is musicianship (Bidelman & Yoo, 2020; Yoo & Bidelman, 2019). Several studies report a so-called "musician advantage" in cognitive processing (c.f. Escobar et al., 2019; Hennessy et al., 2022), whereby individuals with musical training show enhancements across domains like audiovisual integration (Wang et al., 2022) and working memory (Brandler & Rammsayer, 2003; Hansen et al., 2013). Musicians are reported to have enhanced auditory skills (Kraus & Chandrasekaran, 2010) supported by a myriad of neuroplastic changes stemming from cochlea (Bidelman et al., 2016; Bidelman et al., 2017) to cortex (Anderson et al., 2011; Schneider et al., 2002). Musicians are also better at decoding emotion based on speech prosody (Thompson et al., 2004) and have more robust brainstem responses to speech and musical sounds (Bidelman, Gandour, & Krishnan, 2011; Bidelman, Krishnan, & Gandour, 2011; Musacchia et al., 2007). Among the more widely reported musician advantages is the putative enhancement in SIN listening (Coffey et al., 2017; Hennessy et al., 2022). While not all studies have shown SIN advantages (e.g., Boebinger et al., 2015; Madsen et al., 2019; Madsen et al., 2017; Ruggles et al., 2014), several have reported that musicians are more successful at speech segregation in multi-talker scenes (Baskent & Gaudrain, 2016; Bidelman & Yoo, 2020) and show more resilient subcortical encoding of speech sounds in background noise than nonmusicians (Bidelman & Krishnan, 2010; Parbery-Clark et al., 2009). Listeners with music training are also better able to harness executive control in facilitating auditory attention in SIN listening (Strait & Kraus, 2011), and they are less susceptible to interference from informational masking (Oxenham et al., 2003; Swaminathan et al., 2015).

Importantly, enhanced auditory skills and SIN advantages can be observed in listeners with minimal or no musical training but high levels of innate musicality (e.g., Mankel & Bidelman, 2018; Zhu et al., 2021). This suggests that putative benefits in cocktail party listening reported among musicians might not be due to musical training/experience, *per se*, but rather inherent listening skills. Mankel and Bidelman (2018) showed that nonmusicians who scored high on objective measures of musicality had more resilient neural encoding of speech-in-noise than less musical listeners. Similarly, listeners with lower musicality were more affected by the presence of vocals during a speech comprehension task (Brown & Bidelman, 2022a), but only when background music was unfamiliar to them. In contrast, high musicality listeners showed less susceptibility to this informational masking effect, indicating that they were more resilient in difficult listening conditions.

1.6. Selective attention in cocktail party speech perception

Successful "cocktail party" listening requires successful selective attention (Oberfeld & Klöckner-Nowotny, 2016). Attention also plays a role in auditory stream segregation (Bregman, 1990), although there is some debate whether these streams are created pre-attentively or as a result of attention (Fritz et al., 2007). Such attentional modulation is reflected in the brain as increased activity in auditory cortical areas (Elhilali et al., 2009) with a leftward hemispheric lateralization in cases of speech stimuli (Hugdahl et al., 2003). Neural tracking of the target speech signal is stronger for attended sounds, but the brain still maintains representations of the unattended/background sounds whether they are speech or music (Alain & Woods, 1993; Ding & Simon, 2012; Maidhof & Koelsch, 2011).

Attentional effects can be observed even in the early auditory cortical potentials (ERPs) at sensory stages of speech processing. There is a long-established attentional enhancement of the auditory N1, a negative peak around 100 ms in the canonical auditory ERPs (Ding & Simon, 2012; Hillyard et al., 1973; Woldorff et al., 1993). However, attentional modulation of cortical responses has also been observed as early as 40 ms (Teder et al., 1993; Woldorff et al., 1993; Woldorff & Hillyard, 1991) and 75 ms (Bidet-Caulet et al., 2007). These findings suggest attention exerts early influences on auditory sensory coding which may improve SIN analysis by bolstering and/or attenuating target from non-target streams in a cocktail party scenario.

In a study by Ding and Simon (2012), listeners were instructed to attend to one of two speakers. Neural representations of both the attended and unattended talkers were preserved but heavily modulated by attention; that is, cortical encoding of the attended speaker was much larger. However, this study used speech-on-speech stimuli, and the role of selective attention in speech-on-music tracking has not yet been investigated. Our previous study (Brown & Bidelman, 2022a) similarly measured neural tracking to continuous speech, but that study manipulated attention to target speech by changing the level of masker distraction (i.e., background music familiarity); there was no attend-music condition. The current paradigm allows us to further probe listeners' attention by directing their attention to either the music or speech while using ecologically valid stimuli (as in Brown & Bidelman, 2022a; Brown & Bidelman, 2022b).

The current experiment aims to elucidate speech perception in background music and how it is modulated by (i) directed attention, (ii) familiarity of the music, and (iii) listeners' musicality. Participants listened to a speech audiobook and concurrent familiar/unfamiliar music while completing a keyword identification task that directed attention to either the continuous speech or the song lyrics. We measured neural activity using multichannel EEG and extracted the brain's tracking of the continuous amplitude envelope of the audiobook and song vocals using temporal response function (TRF) analysis. We hypothesized that (1) keyword identification and neural tracking would be worse for speech presented in background music compared to in silence (i.e., expected masking effect); (2) neural speech tracking would be weaker in unfamiliar background music (Brown & Bidelman, 2022a); (3) speech tracking would be enhanced when speech was the attended condition versus music as the attended condition; and (4) less musical listeners would show poorer attentional juggling between the speech and music attention conditions, suggesting worse attentional allocation of cognitive resources.

2. Materials and methods

2.1. Participants

The sample included 31 young adults ages 21–33 ($M = 24$, $SD = 3.3$ years, 13 male). All participants showed normal audiometric thresholds < 15 dB HL at octave frequencies 250–8000 Hz, as well as normal SIN perception (QuickSIN scores < 3 dB SNR loss; Killion et al., 2004). All reported English as their native language. Participants were primarily right-handed (mean 70 % laterality using the Edindurgh Handedness Inventory; Oldfield, 1971). Participants also self-reported years of formal music training, which ranged from 0 to 16 years ($M = 4.9$ years, $SD = 4.92$). Each was paid for their time and gave written consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis.

2.2. Stimuli

2.2.1. Music

We used unfamiliar and familiar pop song music selections as music stimuli. To qualify as "familiar," the song had to appear on the Billboard Hot 100 list (https://www.billboard.com/charts/hot-100/) at least once. Each song was sung by a female singer and matched in genre (pop music). All songs were performed at a tempo from 110 to 130 beats per minute. Thompson et al. (2011) showed an effect of faster musical tempi on concurrent reading comprehension, so the tempo range here falls in the "slow" to "intermediate" range of their experiment to avoid tempo effects. Using the above criteria, four songs were used in the current experiment: two familiar ("Girls" by Beyonce; "Stronger (What Doesn't Kill You)" by Kelly Clarkson) and two unfamiliar ("Joan of Arc on the Dance Floor" by Aly & AJ; "OMG What's Happening" by Ava Max). Familiarity categories were determined using a pilot study (N = 37, 15 males, 22 females; age M = 26, SD = 2.95), where participants were asked to rate several songs on a 5-point Likert scale from "Not familiar at all" to "Extremely familiar." The songs used in the current EEG experiment were the two most and least familiar songs from those pilot results.

Songs were converted from stereo to mono, sampled at 44100 Hz, and truncated from onset to 2 min. To maximize data available for analysis, instrumental introductions were cut so that vocals began withing 2 sec of the start of the clip. Clips were RMS-normalized to equate overall levels. However, amplitude fluctuations in the music (i.e., short instrumental segments, chorus) were allowed to vary within 10 dB of the overall RMS to maintain the natural amplitude envelope of the original music.

2.2.2. Speech

The speech stimulus was a public domain audiobook from LibriVox (https://librivox.org/). The selected audiobook was "The Forgotten Planet" by Murray Leinster read by a male speaker; importantly, this story was unfamiliar to all participants. The story was separated into 36 2-min segments. Silences longer than 300 ms were shortened to avoid long gaps in the speech (Brown & Bidelman, 2022a; Ding & Simon, 2012).

2.3. Task

During EEG recording (described below), each audiobook story clip was presented concurrently with one of the four songs in a random order at a signal-to-noise ratio (SNR) of 0 dB or in silence. The story clips were presented in sequence but were broken up into 8 blocks to allow breaks during the task. For half the experiment, the participant was instructed to attend to the audiobook and listen for a keyword; they were instructed to quickly press the space bar every time they heard the keyword. The other half of the experiment was identical, but listeners were cued to listen for a keyword in the music song vocals. All stimuli, both audiobook and song, were 2 min clips. Songs were repeated throughout the duration of the task, but each audiobook segment was presented concurrently with one song. To combat any effects of learning due to song repetition throughout the experiment, several variables (including song presentation order and attention condition order) were counterbalanced across participants. After completing the experiment, participants indicated their familiarity with each song on a sliding scale from 0 (not familiar) to 10 (extremely familiar). They were also asked how much they liked each song (0 to 10 scale).

Participants also completed the shortened version of the Profile of Musical Perception Skills (PROMS-S; Zentner & Strauss, 2017) to assess music-related listening skills. The PROMS is broken up into several subtests that assess different perceptual functions related to music (e.g., rhythm, tuning, melody recognition). In each subtest, two tokens (e.g., rhythms or tones) are presented, and the listener must judge whether the tokens are the same or different using a five-point Likert scale (1 = "definitely different", 5 = "definitely same").

2.4. EEG recording and preprocessing

Participants sat in an electrically shielded, sound-attenuated booth for the duration of the experiment. Continuous EEG recordings were obtained from 64 channels with electrode position according to the

10–10 system (Oostenveld & Praamstra, 2001). Neural signals were digitized at a 500 Hz sample rate using SynAmps RT amplifiers (Compumedics Neuroscan, Charlotte, NC, USA) and online passband of DC-200 Hz. Data were referenced to an electrode placed 1 cm posterior to Cz on the midline (midway between Cz and CPz) during online recording. Contact impedances were maintained below 10kΩ. Music and speech stimuli were each presented diotically at 70 dB SPL (0 dB SNR) via E-A-RTone 2A insert headphones (E-A-R Auditory Systems, 3 M, St. Paul, MN, USA). Presentation of the speech alone served as a control condition to assess speech tracking without music. Stimuli were presented using a custom MATLAB program (v. 2021a; MathWorks, Natick, MA, USA) and routed through at TDT RP2 signal processor (Tucker-Davis Technologies, Alachua, FL, USA).

EEGs were re-referenced to the average mastoids for analysis. We visually inspected the power spectrum for each participant's recording via EEGLAB (Delorme & Makeig, 2004), and paroxysmal channels were spline interpolated with the six nearest neighbor electrodes. The cleaned continuous data were then segmented into 2-minute epochs. Data from 0 to 1000 ms after the onset of each epoch were discarded in order to avoid transient onset responses in later analyses (Crosse et al., 2021). Epochs were then concatenated per condition, resulting in 16 min of EEG in each attention condition for each familiarity condition.

2.5. Data analysis

2.5.1. Familiarity validation

After the experiment, participants were asked to rate their familiarity of each song on a continuous scale to validate our pre-defined familiar and unfamiliar song categories. With 0 being "not familiar at all" and 10 being "extremely familiar," ratings between the familiar ($M = 8.08$, $SD = 1.95$) and unfamiliar ($M = 1.13$, $SD = 1.94$) were significantly different ($t(30) = 11.35$, $p < 0.001$).

2.5.2. Behavioral data analysis

Keypresses were logged and compared to the onset of each keyword. A press that fell within 300–1500 ms after the onset of the word was marked a "hit." Responses earlier than 300 ms were discarded as improbably fast guesses (e.g., Bidelman & Walker, 2017). A keyword with no response in the window was marked a "miss," and a response not in a keyword window was marked as a "false alarm." Hits and false alarms were used to calculate *d'* (d-prime) sensitivity. *d'* was calculated by subtracting the z-score of the false alarm rate from the z-score of the hit rate. Because values of 0 or 1 cannot be z-transformed, hit rates or false alarm rates of 0 were changed to 0.001, and rates of 1 were changed to 0.99 to allow for calculation of *d'* (Macmillan & Creelman, 2005).

2.5.3. Temporal response functions (TRFs)

We quantified the neural tracking to the continuous speech signal using the Temporal Response Function toolbox in MATLAB (Crosse et al., 2016). The forward TRF is a linear function that models the deconvolved impulse response to a continuous stimulus. We down-sampled the broadband continuous audiobook speech to 250 Hz, then extracted the temporal envelope via a Hilbert transform. The extracted EEG recording data were also down-sampled to 250 Hz. Using ERPLAB (Lopez-Calderon & Luck, 2014), we removed the DC offset from the data, then filtered between 1 and 30 Hz using a 6th order Butterworth filter to target cortical activity to the low-frequency speech envelope. EEG and stimulus data were both z-score normalized. Due to inherent inter-subject variability, we computed a TRF for each individual (Crosse et al., 2016). We used 6-fold cross-validation to derive TRFs per familiarity and attention condition, then used ridge regression to find the optimal λ smoothing parameter (Crosse et al., 2021). The model was first trained on the neural response to the attended speech-in-quiet condition to find the optimized λ parameter, which was the value that resulted in the best-fitting model with maximum prediction accuracy. That parameter was then used to compute TRFs for the other masking and attention conditions. This approach avoids overfitting while preserving individual response consistency and increasing decoding accuracy across all speech-tracking conditions. The time window for TRF extraction was between −50 ms and 500 ms. We trained the model using EEG recordings from a fronto-central electrode cluster (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) to further optimize fit based on the canonical topography of auditory ERPs.

From TRF waveforms, we measured the amplitude and latency of the "P1" and "N1" waves, which occur within the expected timeframe of auditory attentional effects in the ERPs (~50 ms and ~100 ms, respectively). The potentials in this range are localized to early auditory cortex and occur well before any potential motor response (Picton et al., 1999). Based on grand average waveforms, $P1_{TRF}$ was measured as the positive-going deflection between 30–60 ms and $N1_{TRF}$ as the negative peak between 100 and 150 ms. We measured RMS amplitude and latency for each peak. $P1_{TRF}$ peaks were small and variable across some listeners (see **Fig. S1**). However, initial Bayes factor (BF) (Makowski et al., 2019; Rouder et al., 2009) analyses across all $P1_{TRF}$ responses and using default priors showed extreme evidence (all BFs > 1000) in favor of the alternate hypothesis of non-zero signal amplitude.

2.5.4. Statistical analysis

Statistics were computed in R using the *lme4* (v. 1.1.32; Bates et al., 2015) package. We used mixed models with combinations of familiarity category (familiar/unfamiliar), attention (attending to book or music), and PROMS level as fixed effects as well their interactions. Subjects served as a random factor. We attempted a maximal random effects structure (Barr et al., 2013), but these models resulted in singular fits and thus only a random intercept on subjects was retained in the random effects structure. *F*-statistics and *p*-values were computed using the *lmerTest* package with degrees of freedom (*d.f.*) using Satterthwaite's method (v. 3.1–3; Kuznetsova et al., 2017). Effect sizes are reported as partial eta squared computed from the *effectsize* (v. 0.8.3; Ben-Shachar et al., 2020) package. Multiple comparisons were adjusted using Tukey corrections. An *a priori* significance level was set at $\alpha = 0.05$.

In preliminary analyses we also examined TRFs at two frontal clusters over the right (F2, F4, F6, F8, FC6, FT8) and left (F1, F3, F5, F7, FC5, FT7) scalp to investigate any hemispheric differences. There were no significant interactions between hemisphere and attention for $P1_{TRF}$ (amplitude: $p = 0.94$; latency: $p = 0.34$) or $N1_{TRF}$ (amplitude: $p = 0.89$; latency: $p = 0.86$), as well as no significant interactions between hemisphere and familiarity for $P1_{TRF}$ (amplitude: $p = 0.92$; latency: $p = 0.95$) or $N1_{TRF}$ (amplitude: $p = 0.95$; latency: $p = 0.96$). Thus, subsequent analyses and figures use the frontal central cluster that was used to train the TRF model.

3. Results

3.1. PROMS musicality scores

PROMS scores ranged from 24.5 to 58, ($M = 39.33$, $SD = 9.08$) (Fig. 1**a)** and were positively correlated with listeners' years of formal music training ($r(30) = 0.569$, $p < 0.001$). As in previous studies (Brown & Bidelman, 2022a; Mankel & Bidelman, 2018), we used a median split to create "high PROMS" and "low PROMS" groups. These groups do not necessarily reflect years of musical training ("musicians" vs. "non-musicians"), but rather, an objective measure of listeners' musicality (i.e., music perceptual skills). The groups did differ in years of formal training ($t(22.87) = 2.643$, $p = 0.015$), but the high PROMS group showed more variability in years of training ($M = 6.94$, $SD = 6.94$) than the low PROMS group ($M = 2.73$, $SD = 2.91$) (Fig. 1**b**).

3.2. Masking effect

Fig. 2 shows the main effect of masking on speech processing.

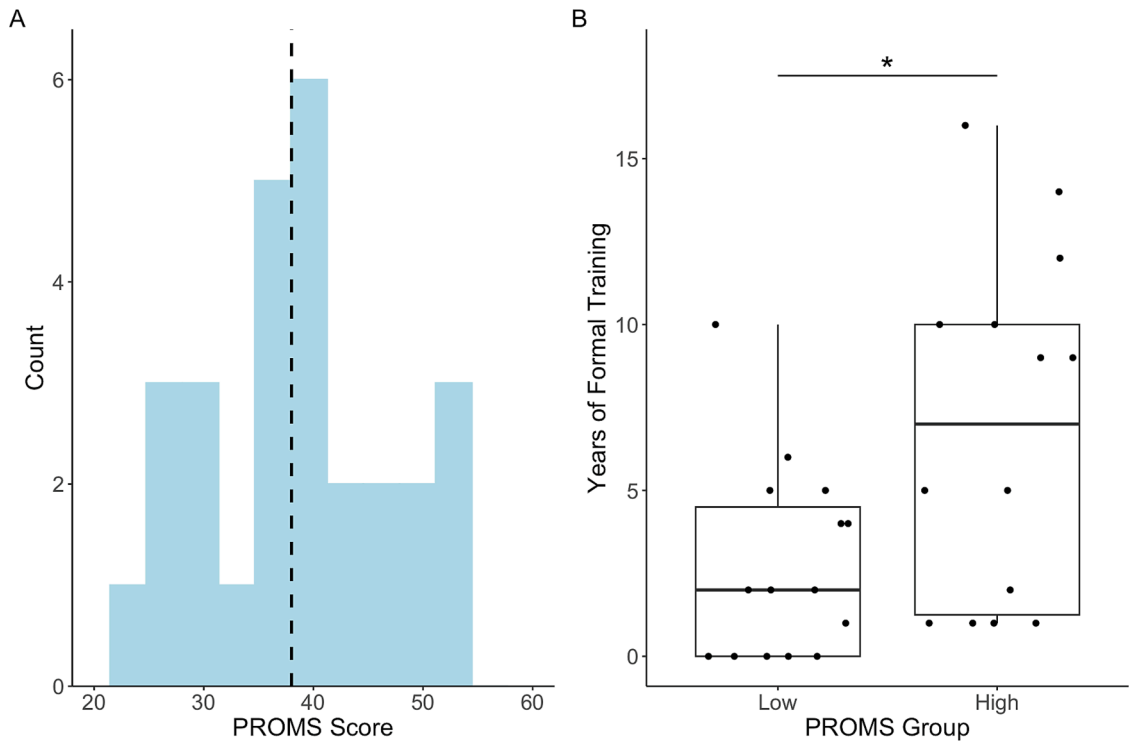


Fig. 1. Results of the PROMS musicality measure. (A) Distribution of PROMS scores, which ranged from 24.5 to 58.0 ($M = 39.33$, $SD = 9.08$). The dashed black line shows the median score (38) used to identify the low and high PROMS groups. (B) Years of formal training (self-reported) for each individual, separated by PROMS group. The high PROMS group had more years of training than the low PROMS group ($t(22.87) = 2.643$, $p = 0.015$). *$p < 0.05$.

Keyword tracking performance (quantified using the *d'* metric) was significantly worse in speech during concurrent music than in silence ($F(1,115) = 52.31$, $p < 0.001$, $\eta_p^2 = 0.31$). Paralleling behavior, the neural TRF $P1_{TRF}$ to speech was slightly longer in latency for masked speech than clean speech ($F(1,115) = 4.78$, $p = 0.003$, $\eta_p^2 = 0.07$), indicating poor encoding of the target speech envelope in noise. The findings confirm that our masking manipulation was successful in weakening the behavioral and neural representation for speech with music as a background noise.

3.3. Familiarity effect

When separating the music maskers by familiarity (for each peak, the main effect of familiarity on latency and amplitude), we found a small effect of familiarity in the strength of the $P1_{TRF}$ evoked by speech (Fig. 3, see also ***Fig. S1***). Amplitude was larger in unfamiliar music than in familiar ($F(2,137) = 3.21$, $p = 0.043$, $\eta_p^2 = 0.04$). There were no amplitude differences between unfamiliar and speech in silence ($p = 0.93$) or between familiar and silence ($p = 0.24$). There was also an effect on latency ($F(2,110) = 4.25$, $p = 0.015$, $\eta_p^2 = 0.07$), which reflected the masking effect. Post-hoc Tukey tests showed that latency of the speech-$P1_{TRF}$ in familiar music was longer than speech in silence ($t(110) = 2.73$, $p = 0.020$). The same prolongation was true for unfamiliar music as compared to speech in silence ($t(110) = 2.67$, $p = 0.024$). There were no differences at $N1_{TRF}$.

3.4. Attention

We found a significant main effect of directed attention on TRF *speech tracking* (Fig. 4) dependent on whether listeners were attending to the speech or song vocals. Notably, TRFs were evident in both conditions, suggesting the neural representation of continuous speech was maintained whether or not it was the attended stream. However, $N1_{TRF}$ responses were earlier when attending to the speech compared to song ($F(1,195) = 9.59$, $p = 0.002$, $\eta_p^2 = 0.05$), indicating speech tracking was enhanced by attention. There was no difference in $N1_{TRF}$ amplitude nor latency/amplitude at $P1_{TRF}$. There were also no significant interactions for either peak between attention and familiarity.

3.5. Effects of musicality

To investigate the relationship between attention and musicality (Fig. 5), we split the sample based on a median split of the PROMS musicality scores and examined *a priori* contrasts for the attentional effect in the low PROMS and high PROMS groups. In the low PROMS group, $N1_{TRF}$ latency was longer when attending to the song than when attending to speech ($F(1,14) = 13.37$, $p = 0.003$, $\eta_p^2 = 0.49$). In stark contrast there was no $N1_{TRF}$ latency difference in the high PROMS group ($p = 0.42$), suggesting the neural tracking of speech was equally good whether or not it was the attended stream.

To probe the role of musical experience, we added years of formal training as a covariate to the above model, and the main effect of attention on $N1_{TRF}$ latency remained significant ($F(1,14) = 12.05$, $p = 0.004$, $\eta_p^2 = 0.46$). Additionally, formal training did not significantly correlate with $N1_{TRF}$ latency when attending to speech ($r(26) = 0.108$, $p = 0.59$) or music ($r(26) = -0.069$, $p = 0.31$).

4. Discussion

In this EEG study, participants listened to speech-music cocktail party mixtures (audiobook + pop music) while they selectively attended to either the speech or the song lyrics. We measured neural tracking of the temporal speech envelope of continuous speech using temporal response functions (TRFs). Beyond expected masking effects of concurrent music, we found early cortical responses ($P1_{TRF}$; ~ 50 ms) to attended speech were slightly larger when the background music was unfamiliar to the listener. Neural responses also showed attentional effects, where $N1_{TRF}$ (~100 ms) to speech was later when attending to

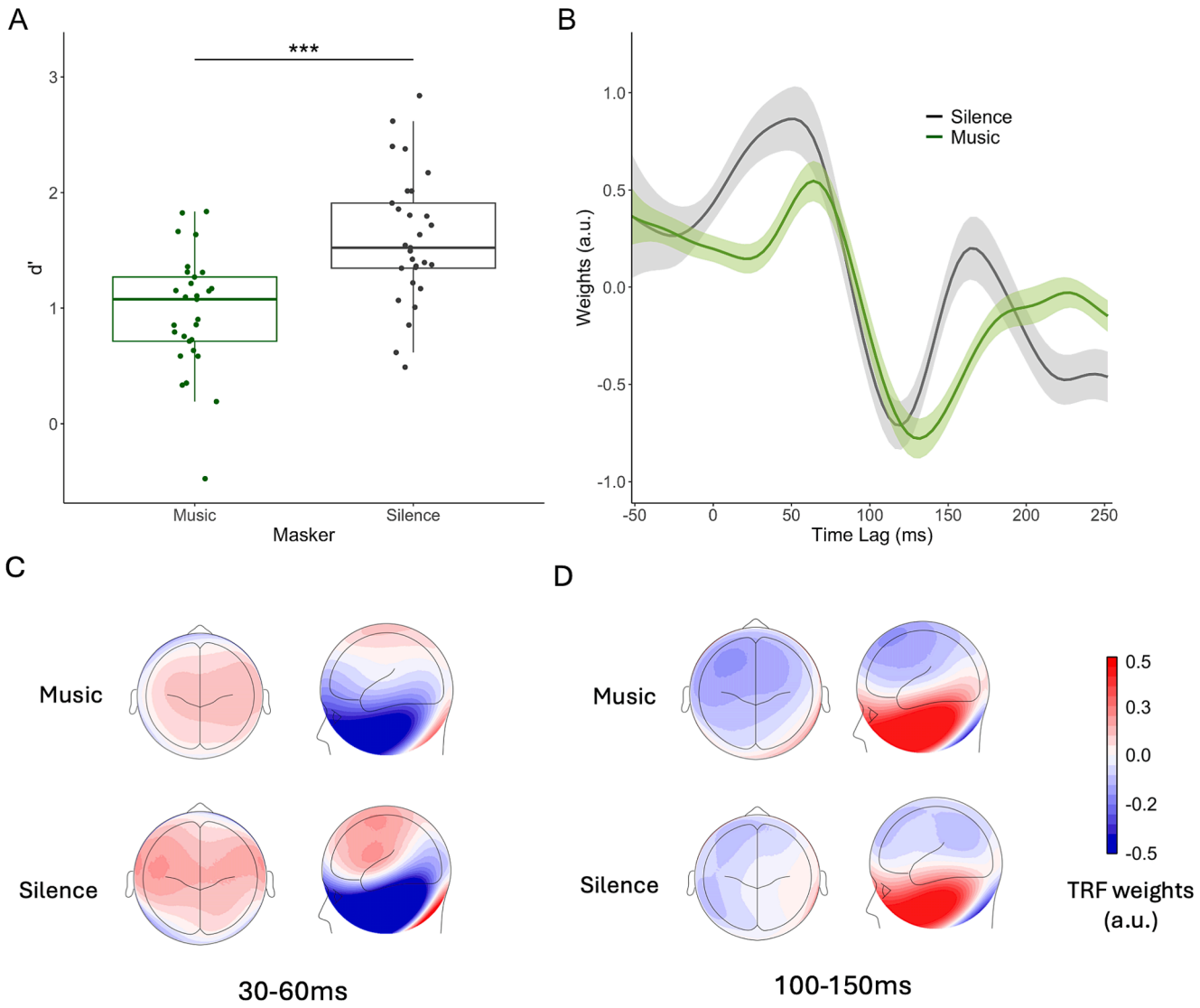


Fig. 2. The presence of music maskers (versus unmasked speech presented in silence) modulates behavior and neural tracking. (A) Behavioral keyword detection performance shown in *d'* (d-prime, a measure of sensitivity) is poorer when masked by music. (B) TRF neural tracking to target speech plotted as the average of the fronto-central electrode cluster (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) shows a prolonged $P1_{TRF}$. Error shading represents ± 1 s.e.m. (C) Topography of $P1_{TRFs}$ in the 30–60 ms analysis window for speech masked by music (top) and silence (bottom). (D) Topography of $N1_{TRFs}$ in the 100–150 ms analysis window for speech masked by music (top) and silence (bottom). ***$p < 0.001$.

song than attending to speech in speech-music mixtures. Interestingly, this attention difference was only prominent in less musical listeners; more musical listeners showed more resilience in tracking speech regardless of whether it was the attended or non-attended stream. Our findings highlight that parsing speech at the cocktail party depends on both the nature of the music backdrop itself as well as the perceptual expertise of the listener.

4.1. Attention enhances neural speech tracking in musical noise

We found a prolonged $N1_{TRF}$ for speech tracking when the audiobook is the attended stream rather than the background (i.e., when attending to the song lyrics). Our far-field EEG data agree with intracranial recordings which show spectrotemporal representations of speech in auditory cortex are heavily modulated by attention (Mesgarani & Chang, 2012). Using spectrotemporal response functions (STRF) applied to far-field MEG, Ding and Simon (2012) showed similar attention effects at 100 ms ($M100_{STRF}$) in a two-talker selective attention task where responses were stronger for the attended speaker versus the unattended speaker. Our similar findings at comparable effect sizes (present study: $\eta_p^2 = 0.05$; Ding and Simon (2012): $\eta_p^2 = 0.06$) show that these attention effects replicate across domains (speech/speech versus speech/music).

Speech intelligibility is easier when the target and interfering speechis spoken by different-sex speakers due to differences in voice fundamental frequency (Brungart, 2001). Bregman (1990) made the distinction between segregation (differentiating different targets or talkers) and streaming (continuously tracking the separated elements). The current study focused on continuous streaming, so segregation was facilitated by having different-sex stimuli (female vocalists, male audiobook reader). The aim of this experiment was not to look at acoustic differences in segregation, but in attentional streaming effects. Future studies may use same-sex stimuli (e.g., a male speaker and a male vocalist) to further investigate speech/music stream segregation when the target and maskers are more similar.

4.2. Early cortical speech processing is weaker in familiar music

We found that $P1_{TRF}$ to continuous speech was slightly smaller when

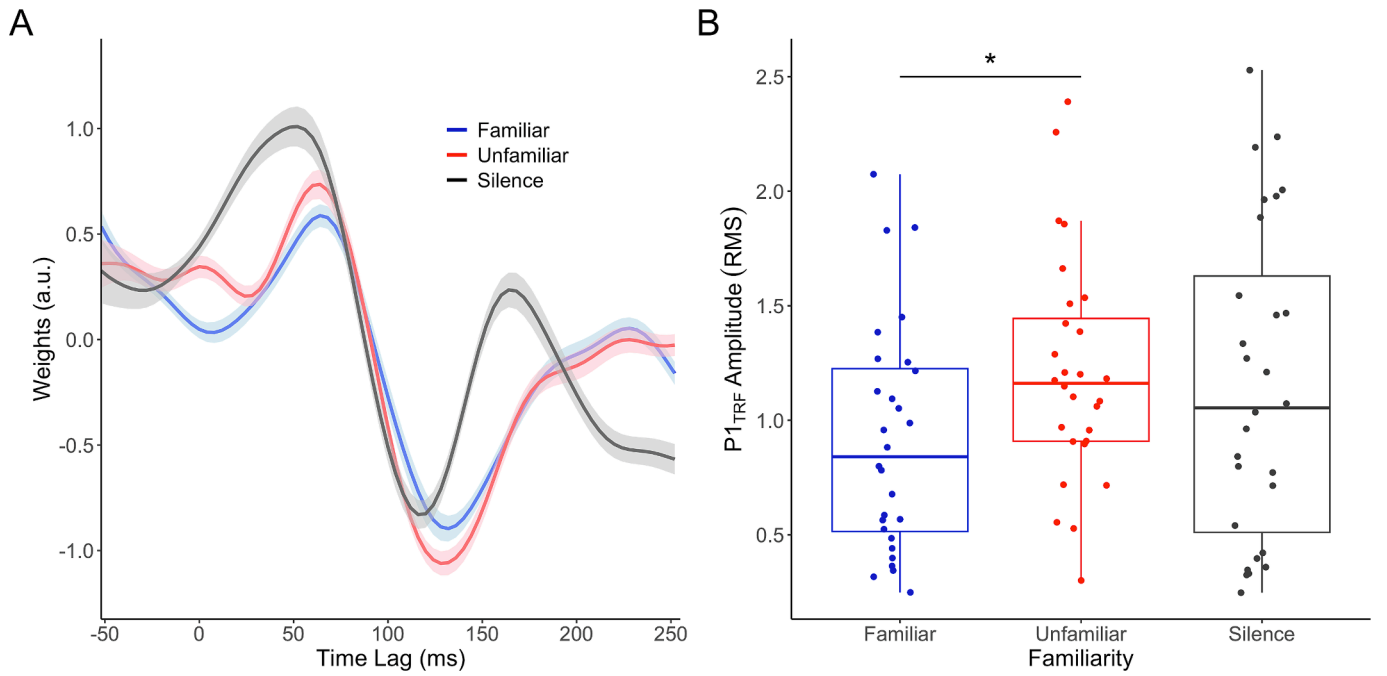


Fig. 3. Speech encoding differs between familiar and unfamiliar music maskers. (A) Grand average TRFs (fronto-central electrodes) representing the neural tracking of speech in familiar and unfamiliar background music. Error shading represents ± 1 s.e.m. (B) $P1_{TRF}$ was larger when presented with unfamiliar music. *$p < 0.05$.

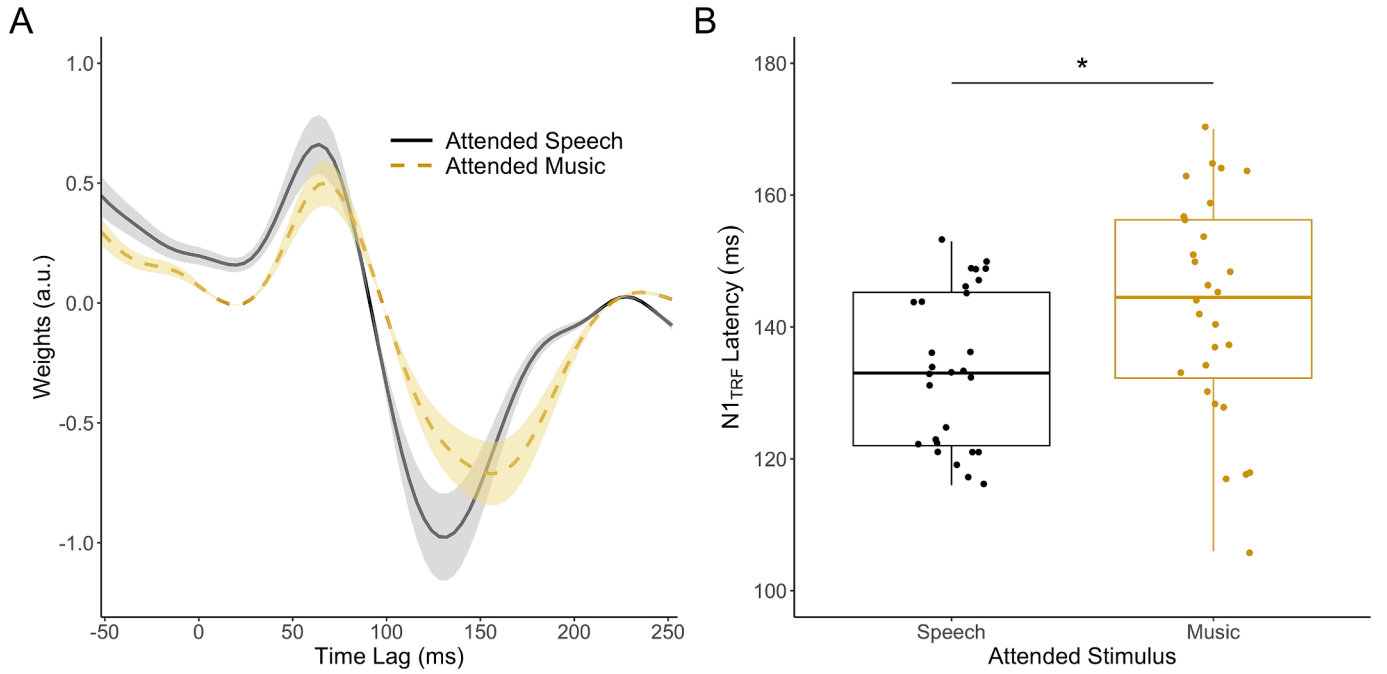


Fig. 4. Selective attention modulates neural speech encoding. (A) Grand average TRFs (plotted at fronto-central electrode cluster) for speech tracking when attention is directed to speech versus song. Error shading represents ± 1 s.e.m. (B) $N1_{TRF}$ for speech encoding is prolonged when attending to the music. *$p < 0.05$.

presented with familiar music. Previous studies from our lab (Brown & Bidelman, 2022a, 2022b) have investigated the role of familiarity in background music on concurrent speech perception using ecological music stimuli like those here. Both studies identified speech processing differences between familiar and unfamiliar music maskers. We previously reasoned that those differences were the result of different allocations of limited cognitive resources needed to facilitate selective attention and inhibit the music maskers (Kahneman, 1973; Lavie et al., 2004). However, prior studies did not direct attention to speech and music (only speech was tracked behaviorally), so such explanations were only speculative. Our data here confirm the impact of background music on speech processing most probably results from subtle changes in the spotlight of attention as familiar music draws attention away from the primary speech signal. These findings agree with other work showing neural synchronization is stronger for familiar than unfamiliar music (Weineck et al., 2022). Stronger synchronization to familiar music would tend to reduce entrainment to other concurrent signal, as observed here for speech.

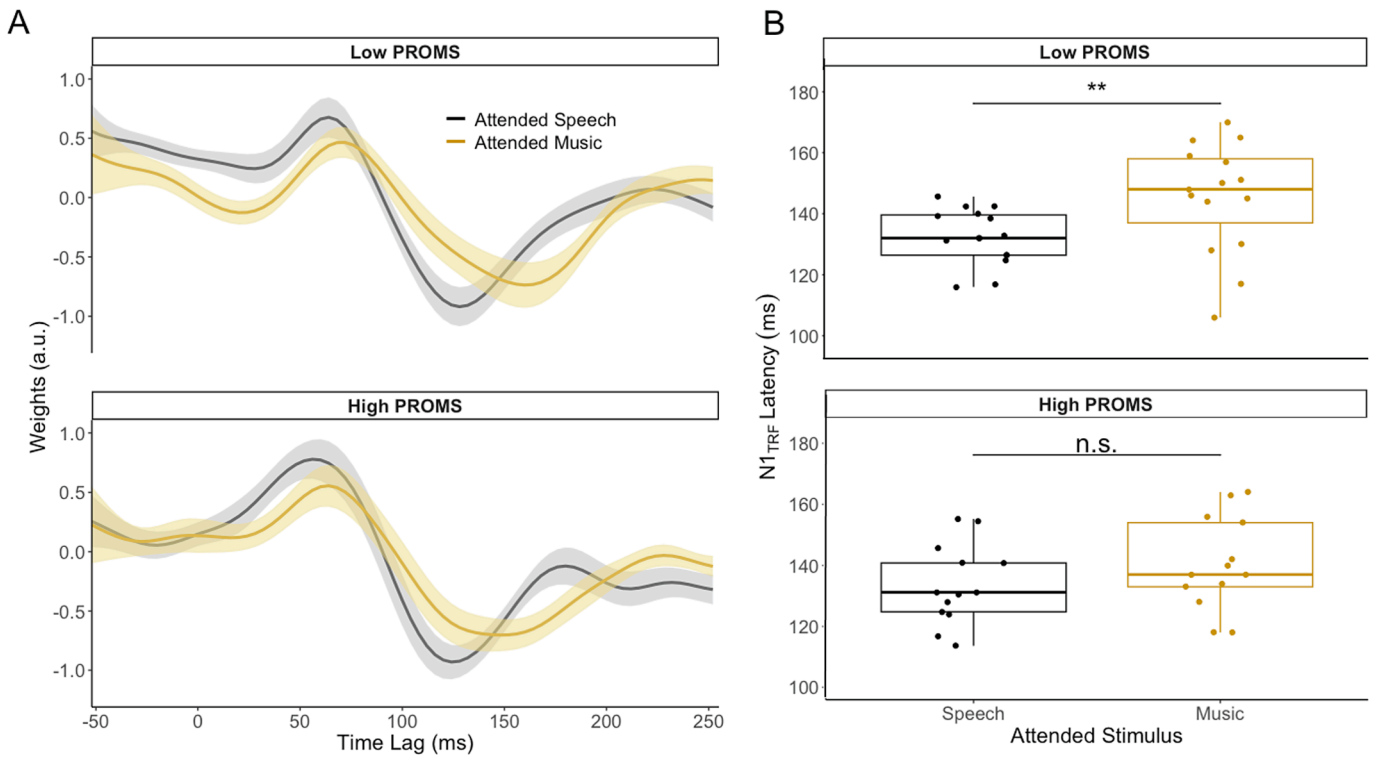


Fig. 5. Attentional allocation at the cocktail party differs between less and more musical listeners. (A) TRF waveforms tracking to speech for low vs. high PROMS listeners. Error shading represents ± 1 s.e.m. (B) $N1_{TRF}$ responses were later than when attending to speech, but only for the less musical listeners. There was no difference in the high PROMS group between music and speech attend conditions. $^{**}p < 0.01$.

While we favor explanations based on attention, familiarity effects could instead result from idiosyncratic acoustic differences between music selections. However, we aimed to combat this by using multiple songs per familiarity condition, as well as using several criteria to match the different songs: genre, tempo, gender of vocalist, key, and beat strength (i.e., pulse; Lartillot et al., 2008). Additionally, we have investigated the role of several acoustic factors, including pulse, on similar familiarity findings and found that while there were acoustic drivers of those effects, the effect sizes were several orders of magnitude smaller than those of music familiarity (Brown & Bidelman, 2022b). Future studies using this paradigm could use multivariate TRFs (Crosse et al., 2016) to see which acoustic variables contribute more to perceptual tracking (e.g., amplitude envelope to vocals and to full song, spectral flux of full song, etc.) that may not be captured in our current analyses.

The early P1 effects in our data contrast several MEG studies that have not shown attentional modulation in auditory cortical processing before 100 ms (Akram et al., 2017; Chait et al., 2010; Ding & Simon, 2012; Fujiwara et al., 1998; Miran et al., 2018; Puvvada & Simon, 2017). Several explanations may account for differences between this and previous studies. First, the P1 component at 50 ms is thought to be generated by lateral superior temporal gyrus (Liégeois-Chauvel et al., 1994; Ponton et al., 2000) with a radial oriented current dipole. MEG is relatively insensitive to radial currents (Scherg et al., 2019), which might explain why MEG TRF studies have not observed attentional modulation in the P1. Second, P1 is a small amplitude component of the auditory ERPs that is quite variable at the single-subject level. The earlier familiarity effects observed in this ($P1_{TRF}$) study compared to previous work ($N1_{TRF}$) could be due to the larger sample size of the current study. It is also important to acknowledge the fairly small effect sizes of these differences, which likely reflect the individual variability in $P1_{TRF}$ waveforms (**Fig. S1**). Nevertheless, the presence of any familiarity-attention effects at ~ 50 ms suggests music (and how familiar it is to the listener) exerts an influence on speech coding no later than primary auditory cortex (Picton et al., 1999).

Interestingly, Yang et al. (2016) showed that musicians' performance on cognitive tasks was worse when the background music was played on their trained instrument (e.g., a trained pianist performed more poorly on a verbal fluency test when the background music was played on a piano versus a guitar). If we assume their chosen instrument is more "familiar" to them, then these findings contrast our data. In our previous study (Brown & Bidelman, 2022a), we found more musical listeners were less impacted by familar background music. Here, familiarity was measured by self-report and presumably based on real-world exposure to the songs. The operational definition of "familiar" ranges across studies, from real-life exposure (Russo & Pichora-Fuller, 2008) to in-lab training (Weiss et al., 2016) to real vs. artificial instrument timbre (Van Hedger et al., 2022). Further research in this area should carefully consider these definitions.

4.3. *Musicality impacts attentional allocation*

The $N1_{TRF}$ peak in response to speech was prolonged when attention was directed to the song versus towards the speech. However, we only observed this difference in the low PROMS group (i.e., the less musical listeners), which is likely due to the variance in the high PROMS group (i.e., the more musical individuals). In general, high musicality listeners showed less change between the attend-speech and attend-music conditions, indicating they were more successful in tracking speech regardless of whether or not it was in the attentional spotlight. Similarly, the larger attention-dependent change in TRFs of low PROMS listeners suggests they are more susceptible to changes in background music, possibly resulting from poorer attentional resource allocation and/or increased distractibility by the background (Brown and Bidelman, 2022a). The directed attention manipulations in the current study create new evidence for this explanation. Here, low PROMS listeners showed worse inhibition of the background music, suggesting less musical listeners are poorer at regulating auditory attention. In this vein,

attentional benefits are observed in trained musicians (Strait et al., 2010; Thompson et al., 2017; Yoo & Bidelman, 2019), and improvements in selective attention might also account for individual differences in cocktail party listening (Oberfeld & Klöckner-Nowotny, 2016). Musical training also correlates with better tracking of the to-be-ignored stream, as well as a more balanced representation of the attended and to-be-ignored streams (Puschmann et al., 2019). These studies, along with current data, support the link between musicality, attentional deployment, and cocktail party listening.

Using sentences masked by varying levels of informational content, Boebinger et al. (2015) did not find a difference in perception between musicians and nonmusicians. However, changes in speech perception thresholds were significantly predicted by non-verbal IQ. Indeed, several studies posit that the musician SIN listening advantage may be more attributed to other cognitive (Kraus et al., 2012) or genetic (Drayna et al., 2001; Ukkola et al., 2009) differences. The current study did not collect these cognitive measures, so we are unable to disentangle musical versus non-music domain factors driving the results.

Collectively, our PROMS group differences imply that listeners might approach the speech-music cocktail party with different listening strategies facilitated by different types of musical ability. Unfortunately, our sample is not large enough to further stratify our listeners into instrument-specific subgroups. However, there is evidence that musicians listen and react to music differently (e.g., Mikutta et al., 2014) and show genre-specific tuning of brain activity. For example, classical musicians showing heightened P3 responses when listening to classical music, and rock musicians when listening to rock music (Caldwell & Riby, 2007). Future studies that recruit participants specifically based on primary instrument training would be needed to probe this further.

5. Conclusion

In summary, our results provide novel insight into how we listen to speech in background music. Listening to any music can impair concurrent speech understanding, and familiar music is particularly distracting. These differences may occur as early as 50 ms during speech processing, supporting models of early-attentional control that exert influences on speech coding within the primary auditory cortices. Speech tracking is weaker when attending to background music, but only for less musical individuals. These findings reveal that exogenous properties of acoustic mixtures and endogenous factors of the listener interact when navigating noisy listening environments. Furture research is needed to determine what aspects of musicality or listening strategies cause these differential effects.

CRediT authorship contribution statement

Jane A. Brown: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Gavin M. Bidelman:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Jessica MacLean and Rose Rizzi for comments on the early version of this manuscript. This work was supported by the National Institutes of Health (NIH/NIDCD R01DC016267 to G.M.B.).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2025.105581.

Data availability

Data will be made available on request.

References

Akram, S., Simon, J. Z., & Babadi, B. (2017). Dynamic estimation of the auditory temporal response function from MEG in competing-speaker environments. *IEEE Transactions on Biomedical Engineering, 64*, 1896–1905. https://doi.org/10.1109/TBME.2016.2628884

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Frontiers in Bioscience, 5*, 202–212.

Alain, C., & Woods, D. L. (1993). Distractor clustering enhances detection speed and accuracy during selective listening. *Perception and Psychophysics, 54*(4), 509–514.

Anderson, S., Parbery-Clark, A., Yi, H. G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing, 32*, 750–757. https://doi.org/10.1097/AUD.0b013e31822229d3

Angel, L. A., Polzella, D. J., & Elvers, G. C. (2010). Background music and cognitive performance. *Perceptual and Motor Skills, 110*, 1059–1064. https://doi.org/10.2466/04.11.22.pms.110.c.1059-1064

Atkinson, G., Wilson, D., & Eubank, M. (2004). Effects of music on world-rate distribution during a cycling time trial. *International Journal of Sports Medicine, 25*, 611–615. https://doi.org/10.1055/s-2004-815715

Avila, C., Furnham, A., & McClelland, A. (2012). The influence of distracting familiar vocal music on cognitive performance of introverts and extraverts. *Psychology of Music, 40*, 84–93. https://doi.org/10.1177/0305735611422672

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Baskent, D., & Gaudrain, E. (2016). Musician advantage for speech-on-speech perception. *The Journal of the Acoustical Society of America, 139*(3), Article EL51-56. https://doi.org/10.1121/1.4942628

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*. https://doi.org/10.18637/jss.v067.i01

Beh, H. C., & Hirst, R. (1999). Performance on driving-related tasks during music. *Ergonomics, 42*, 1087–1098. https://doi.org/10.1080/001401399185153

Belfi, A. M., Karlan, B., & Tranel, D. (2016). Music evokes vivid autobiographical memories. *Memory, 24*(7), 979–989. https://doi.org/10.1080/09658211.2015.1061012

Ben-Shachar, M. S., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of Effect Size Indices and Standardized Parameters. *Journal of Open Source Software, 5*(56), 2815. https://doi.org/10.21105/joss.02815

Bendixen, A. (2014). Predictability effects in auditory scene analysis: A review. *Frontiers in Neuroscience, 8*, 1–16. https://doi.org/10.3389/fnins.2014.00060

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience, 23*(2), 425–434. https://doi.org/10.1162/jocn.2009.21362

Bidelman, G. M., & Krishnan, A. (2010). Effects of reverberation on brainstem representation of speech in musicians and non-musicians. *Brain Research, 1355*, 112–125.

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience, 33*(3), 530–538.

Bidelman, G. M., Nelms, C., & Bhagat, S. P. (2016). Musical experience sharpens human cochlear tuning. *Hearing Research, 335*, 40–46.

Bidelman, G. M., Schneider, A. D., Heitzmann, V. R., & Bhagat, S. P. (2017). Musicianship enhances ipsilateral and contralateral efferent gain control to the cochlea. *Hearing Research, 344*, 275–283. https://doi.org/10.1016/j.heares.2016.12.001

Bidelman, G. M., & Walker, B. (2017). Attentional modulation and domain specificity underlying the neural organization of auditory categorical perception. *European Journal of Neuroscience, 45*(5), 690–699. https://doi.org/10.1111/ejn.13526

Bidelman, G. M., & Yoo, J. (2020). Musicians Show Improved Speech Segregation in Competitive, Multi-Talker Cocktail Party Scenarios. *Frontiers in Psychology, 11*, 1–11. https://doi.org/10.3389/fpsyg.2020.01927

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., & Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *Journal of Neuroscience, 27*, 9252–9261. https://doi.org/10.1523/JNEUROSCI.1402-07.2007

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. *The Journal of the Acoustical Society of America, 137*(1), 378–387. https://doi.org/10.1121/1.4904537

Brandler, S., & Rammsayer, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music, 31*, 123–138. https://doi.org/10.1177/0305735603031002290

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound.*. https://doi.org/10.1121/1.408434
Brouwer, S., Akkermans, N., Hendriks, L., van Uden, H., & Wilms, V. (2021). "Lass frooby noo!" the interference of song lyrics and meaning on speech intelligibility. *Journal of Experimental Psychology: Applied.*. https://doi.org/10.1037/xap0000368
Brown, J. A., & Bidelman, G. M. (2022a). Familiarity of Background Music Modulates the Cortical Tracking of Target Speech at the "Cocktail Party". *Brain Sciences, 12*(10). https://doi.org/10.3390/brainsci12101320
Brown, J. A., & Bidelman, G. M. (2022b). Song properties and familiarity affect speech recognition in musical noise. *Psychomusicology: Music, Mind, and Brain*. doi: 10.1037/pmu0000284.
Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America, 109*, 1101–1109. https://doi.org/10.1121/1.1345696
Caldwell, G. N., & Riby, L. M. (2007). The effects of music exposure and own genre preference on conscious and unconscious cognitive processes: A pilot ERP study. *Consciousness and Cognition, 16*, 992–996. https://doi.org/10.1016/j.concog.2006.06.015
Cassidy, G., & MacDonald, R. (2009). The effects of music choice on task performance: A study of the impact of self-selected and experimenter-selected music on driving game performance and experience. *Musicae Scientiae, 13*, 357–386. https://doi.org/10.1177/102986490901300207
Castro, M., L'Heritier, F., Plailly, J., Saive, A. L., Corneyllie, A., Tillmann, B., & Perrin, F. (2020). Personal familiarity of music and its cerebral effect on subsequent speech processing. *Scientific Reports, 10*(1), 14854. https://doi.org/10.1038/s41598-020-71855-5
Chait, M., de Cheveigne, A., Poeppel, D., & Simon, J. Z. (2010). Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia, 48*(11), 3262–3271. https://doi.org/10.1016/j.neuropsychologia.2010.07.007
Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *Journal of the Acoustical Society of America, 25*, 975–979. https://doi.org/10.1121/1.1907229
Chtourou, H., Jarraya, M., Aloui, A., Hammouda, O., & Souissi, N. (2012). The effects of music during warm-up on anaerobic performances of young sprinters. *Science and Sports, 27*. https://doi.org/10.1016/j.scispo.2012.02.006
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
Crawford, H. J., & Strapp, C. M. (1994). Effects of vocal and instrumental music on visuospatial and verbal performance as moderated by studying preference and personality. *Personality and Individual Differences, 16*, 237–245. https://doi.org/10.1016/0191-8869(94)90162-7
Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00604
Crosse, M. J., Zuk, N. J., Liberto, G. M. D., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Naturalistic Stimuli : Methodological Considerations for Applied Research. *Frontiers in Neuroscience, 15*.
de Groot, A. M. B., & Smedinga, H. E. (2014). Let the music play! : A short-term but no long-term detrimental effect of vocal background music with familiar language lyrics on foreign language vocabulary learning. *Studies in Second Language Acquisition, 36*, 681–707. https://doi.org/10.1017/S0272263114000059
Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009
Ding, C. G., & Lin, C.-H. (2012). How does background music tempo work for online shopping? *Electronic Commerce Research and Applications, 11*(3), 299–307. https://doi.org/10.1016/j.elerap.2011.10.002
Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America, 109*, 11854–11859. https://doi.org/10.1073/pnas.1205381109
Drayna, D., Manichaikul, A., de Lange, M., Snieder, H., & Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science, 291*, 1969–1972.
Du, M., Jiang, J., Li, Z., Man, D., & Jiang, C. (2020). The effects of background music on neural responses during reading comprehension. *Scientific Reports, 10*. https://doi.org/10.1038/s41598-020-75623-3
Ekström, S. R., & Borg, E. (2011). Hearing speech in music. *Noise and Health, 13*, 277–285. https://doi.org/10.4103/1463-1741.82960
Elhilali, M., Xiang, J., Shamma, S. A., & Simon, J. Z. (2009). Interaction between attention and bottom-up saliency mediates the representation of foreground and background in an auditory scene. *PLoS Biology, 7*(6), Article e1000129. https://doi.org/10.1371/journal.pbio.1000129
Escobar, J., Mussoi, B. S., & Silberer, A. B. (2019). The Effect of Musical Training and Working Memory in Adverse Listening Situations. *Ear and Hearing, 41*, 278–288. https://doi.org/10.1097/AUD.0000000000000754
Etaugh, C., & Ptasnik, P. (1982). Effects of studying to music and post-study relaxation on reading comprehension. *Perceptual and Motor Skills, 55*, 141–142.
Feng, S., & Bidelman, G. M. (2015). Music listening and song familiarity modulate mind wandering and behavioral success during lexical processing. *Annual Meeting of the Cognitive science Society (CogSci 2015)*.
Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention - focusing the searchlight on sound. *Current Opinion in Neurobiology, 17*, 437–455. https://doi.org/10.1016/j.conb.2007.07.011
Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., & Shibasaki, H. (1998). Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. *Cognitive Brain Research, 7*, 99–109.
Furnham, A., & Allass, K. (1999). The influence of musical distraction of varying complexity on the cognitive performance of extroverts and introverts. *European Journal of Personality, 13*(1), 27–38. https://doi.org/10.1002/(sici)1099-0984(199901/02)13:1<27::Aid-per318>3.0.Co;2-r
Furnham, A., & Strbac, L. (2002). Music is as distracting as noise: The differential distraction of background music and noise on the cognitive test performance of introverts and extraverts. *Ergonomics, 45*, 203–217. https://doi.org/10.1080/00140130210121932
Garlin, F. V., & Owen, K. (2006). Setting the tone with the tune: A meta-analytic review of the effects of background music in retail settings. *Journal of Business Research, 59*(6), 755–764. https://doi.org/10.1016/j.jbusres.2006.01.013
Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9*, 697–704. https://doi.org/10.1093/cercor/9.7.697
Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music, 41*, 779–793. https://doi.org/10.1177/0305735612452186
Haykin, S., & Chen, Z. (2005). The cocktail party problem. *Neural Computation, 17*, 1875–1902. https://doi.org/10.1162/0899766054322964
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*, Article 108442. https://doi.org/10.1016/j.heares.2022.108442
Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science, 182*, 177–180. https://doi.org/10.1126/SCIENCE.182.4108.177
Hine, K., Abe, K., Kinzuka, Y., Shehata, M., Hatano, K., Matsui, T., & Nakauchi, S. (2022). Spontaneous motor tempo contributes to preferred music tempo regardless of music familiarity. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.952488
Hugdahl, K., Thomsen, T., Ersland, L., Rimol, L. M., & Niemi, J. (2003). The effects of attention on speech perception: An fMRI study. *Brain and Language, 85*(1), 37–48. https://doi.org/10.1016/s0093-934x(02)00500-x
Husain, G., Thompson, W., & Schellenberg, E. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception, 20*, 151–171.
Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterisation of music-evoked autobiographical memories. *Memory, 15*, 845–860. https://doi.org/10.1080/09658210701734593
Jäncke, L., & Sandmann, P. (2010). Music listening while you learn: No influence of background music on verbal learning. *Behavioral and Brain Functions, 6*(3).
Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2011). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of Music, 40*(3), 339–356. https://doi.org/10.1177/0305735610387777
Jones, M., Kidd, G., & Wetzel, R. (1981). Evidence for Rhythmic Attention. *Journal of Experimental Psychology: Human Perception and Performance, 7*(5), 1059–1073.
Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall Inc.
Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music, 39*, 424–448. https://doi.org/10.1177/0305735610376261
Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 116*(4 Pt 1), 2395–2405. http://www.ncbi.nlm.nih.gov/pubmed/15532670.
Kiss, L., & Linnell, K. J. (2022). Making sense of background music listening habits: An arousal and task-complexity account. *Psychology of Music, 51*(1), 89–106. https://doi.org/10.1177/03057356221089017
Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews. Neuroscience, 11*(8), 599–605. https://doi.org/10.1038/nrn2882
Kraus, N., Strait, D. L., & Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Annals of the New York Academy of Sciences, 1252*, 100–107. https://doi.org/10.1111/j.1749-6632.2012.06463.x
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13
Lartillot, O., Eerola, T., Toiviainen, P., & Fornari, J. (2008). Multi-Feature Modeling of Pulse Clarity: Design, Validation. *and Optimization*. International Society for Music Information Retrieval.
Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75–82. https://doi.org/10.1016/j.tics.2004.12.004
Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General, 133*, 339–354. https://doi.org/10.1037/0096-3445.133.3.339
Lehmann, J. A. M., & Seufert, T. (2017). The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity. *Frontiers in Psychology, 8*, 1902. https://doi.org/10.3389/fpsyg.2017.01902
Liégeois-Chauvel, C., Musolino, A., Badier, J., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology, 92*, 204–214.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Macmillan, N., & Creelman, C. (2005). The Yes-No Experiment: Sensitivity. In *Detection Theory: A User's Guide* ((2nd ed.,, pp. 3–26). Lawrence Erlbaum Associates.

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*, 10404. https://doi.org/10.1038/s41598-019-46728-1

Madsen, S. M. K., Whiteford, K. L., & Oxenham, A. J. (2017). Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds. *Scientific Reports, 7*, 12624. https://doi.org/10.1038/s41598-017-12937-9

Maidhof, C., & Koelsch, S. (2011). Effects of Selective Attention on Syntax Processing in Music and Language. *Journal of Cognitive Neuroscience, 23*(9), 2252–2267.

Makowski, D., Ben-Shachar, M., & Lüdecke, D. (2019). bayestestR: Describing Effects and their Uncertainty, Existence and Significance within the Bayesian Framework. *Journal of Open Source Software, 4*(40). https://doi.org/10.21105/joss.01541

Mankel, K., & Bidelman, G. M. (2018). Inherent auditory skills rather than formal music training shape the neural encoding of speech. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 13129–13134. https://doi.org/10.1073/pnas.1811793115

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485*. https://doi.org/10.1038/nature11020

Mikutta, C. A., Maissen, G., Altorfer, A., Strik, W., & Koenig, T. (2014). Professional musicians listen differently to music. In *Neuroscience* (Vol. 268, pp. 102-111): Pergamon.

Miran, S., Akram, S., Sheikhattar, A., Simon, J. Z., Zhang, T., & Babadi, B. (2018). Real-Time Tracking of Selective Auditory Attention From M/EEG: A Bayesian Filtering Approach. *Frontiers in Neuroscience, 12*, 262. https://doi.org/10.3389/fnins.2018.00262

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*, 15894–15898. https://doi.org/10.1073/pnas.0701498104

North, A. C., & Hargreaves, D. J. (1999). Music and driving game performance. *Scandinavian Journal of Psychology, 40*(4), 285–292. https://doi.org/10.1111/1467-9450.404128

North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology, 84*, 271–276. https://doi.org/10.1037//0021-9010.84.2.271

Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife, 5*. https://doi.org/10.7554/eLife.16747

Oldfield, R. (1971). The Assessment and Analysis of Handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97–113. https://doi.org/10.1007/978-0-387-79948-3_6053

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology, 112*, 713–719.

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G., Jr. (2003). Informational masking and musical training. *Journal of the Acoustical Society of America, 114*(3), 1543–1549. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14514207.

Parbery-Clark, A., Skoe, E., & Kraus, N. (2009). Musical experience limits the degradative effects of background noise on the neural processing of sound. *Journal of Neuroscience, 29*, 14100–14107, 29/45/14100 [pii] 10.1523/JNEUROSCI.3256-09.2009.

Perham, N., & Currie, H. (2014). Does listening to preferred music improve reading comprehension performance? In. *Applied Cognitive Psychology, 28*, 279–284.

Picton, T. W., Alain, C., Woods, D. L., John, M. S., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., & Trujillo, N. J. (1999). Intracerebral Sources of Human Auditory-Evoked Potentials. *Audiology and Neuro-Otology, 6*, 64–79.

Ponton, C. W., Eggermont, J. J., Kwong, B., & Don, M. (2000). Maturation of human central auditory system activity: Evidence from multi-channel evoked potentials. *Clinical Neurophysiology, 111*, 220–236.

Puschmann, S., Baillet, S., & Zatorre, R. J. (2019). Musicians at the Cocktail Party: Neural Substrates of Musical Training During Selective Listening in Multispeaker Situations. *Cerebral Cortex, 29*, 3253–3265. https://doi.org/10.1093/cercor/bhy193

Puvvada, K. C., & Simon, J. Z. (2017). Cortical Representations of Speech in a Multitalker Auditory Scene. *The Journal of Neuroscience, 37*(38), 9189–9196. https://doi.org/10.1523/JNEUROSCI.0938-17.2017

Rea, C., MacDonald, P., & Carnes, G. (2010). Listening to classical, pop, and metal music: An investigation of mood. *Emporia State Research Studies, 46*(1), 1–3.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Ruggles, D. R., Freyman, R. L., & Oxenham, A. J. (2014). Influence of musical training on understanding voiced and whispered speech in noise. *PLoS One1, 9*(1), Article e86980. https://doi.org/10.1371/journal.pone.0086980

Russo, F. A., & Pichora-Fuller, M. K. (2008). Tune in or tune out: Age-related differences in listening to speech in music. *Ear and Hearing, 29*, 746–760. https://doi.org/10.1097/AUD.0b013e31817bdd1f

Scharenborg, O., & Larson, M. (2018). *The conversation continues: The effect of lyrics and music complexity of background music on spoken-word recognition*. International Speech Communication Association.

Scherg, M., Berg, P., Nakasato, N., & Beniczky, S. (2019). Taking the EEG Back Into the Brain: The Power of Multiple Discrete Sources. *Frontiers in Neurology, 10*, 855. https://doi.org/10.3389/fneur.2019.00855

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5*(7), 688–694. https://doi.org/10.1038/nn871 nn871 [pii]

Shi, L. F., & Law, Y. (2010). Masking effects of speech and music: Does the masker's hierarchical structure matter? *International Journal of Audiology, 49*, 296–308. https://doi.org/10.3109/14992020903350188

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

Strait, D. L., Kraus, N., Parbery-Clark, A., & Ashley, R. (2010). Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research, 261*(1–2), 22–29. S0378-5955(09)00311-6 [pii] 10.1016/j.heares.2009.12.021.

Su, Y., He, M., & Li, R. (2023). The effects of background music on English reading comprehension for English foreign language learners: Evidence from an eye movement study. *Frontiers in Psychology, 14*, Article 1140959. https://doi.org/10.3389/fpsyg.2023.1140959

Swaminathan, J., Mason, C. R., Streeter, T. M., Best, V., Kidd Jr., G., & Patel, A. D. (2015). Musical training, individual differences and the cocktail party problem. In *Scientific Reports* (2015/06/27 ed., Vol. 5, pp. 11628). Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA. Department of Psychology, Tufts University, Medford, MA.

Teder, W., Kujala, T., & Näätänen, R. (1993). Selection of speech messages in free-field listening. *Neuroreport, 5*, 307–309.

Thompson, E. C., Woodruff Carr, K., White-Schwoch, T., Otto-Meyer, S., & Kraus, N. (2017). Individual differences in speech-in-noise perception parallel neural speech processing and attention in preschoolers. *Hearing Research, 344*, 148–157. https://doi.org/10.1016/j.heares.2016.11.007

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science, 12*, 248–251. https://doi.org/10.1111/1467-9280.00345

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding Speech Prosody: Do Music Lessons Help? In. *Emotion* (Vol. 4,, 46–64.

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2011). Fast and loud background music disrupts reading comprehension. In. *Psychology of Music, 40*, 700–708.

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music, 40*, 700–708. https://doi.org/10.1177/0305735611400173

Ukkola, L. T., Onkamo, P., Raijas, P., Karma, K., & Jarvela, I. (2009). Musical aptitude is associated with AVPR1A-haplotypes. *PLoS One1, 4*(5), e5534.

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience, 16*(4), 601–615. https://doi.org/10.3758/s13415-016-0417-4

Van Hedger, S. C., Johnsrude, I., & Batterink, L. J. (2022). Musical instrument familiarity affects statistical learning of tone sequences. *Cognition, 218*. https://doi.org/10.1016/j.cognition.2021.104949

Wang, D., Jimison, Z., Richard, D., & Chuan, C. (2015). Effect of Listening to Music as a Function of Driving Complexity: A Simulator Study on the Differing Effects of Music on Different Driving Tasks. *Driving Assessment Conference, 8*.

Wang, L., Tang, X., Wang, A., & Zhang, M. (2022). Musical training reduces the Colavita visual effect. *Psychology of Music, 51*(2), 592–607. https://doi.org/10.1177/03057356221108763

Weineck, K., Wen, O. X., & Henry, M. J. (2022). Neural synchronization is strongest to the spectral flux of slow music and depends on familiarity and beat salience. *eLife, 11*. https://doi.org/10.7554/eLife.75515

Weiss, M. W., Trehub, S. E., Schellenberg, E. G., & Habashi, P. (2016). Pupils Dilate for Vocal or Familiar Music. *Journal of experimental psychology. Human perception and performance.*. https://doi.org/10.1037/xhp0000226.supp

Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., & Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences of the United States of America, 90*, 8722–8726.

Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology, 79*, 170–191.

Yang, J., McClelland, A., & Furnham, A. (2016). The effect of background music on the cognitive performance of musicians: A pilot study. *Psychology of Music, 44*, 1202–1208. https://doi.org/10.1177/0305735615592265

Yerkes, R. M., & Dodson, J. D. (1908). The relationship of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.

Yoo, J., & Bidelman, G. M. (2019). Linguistic, perceptual, and cognitive factors underlying musicians' benefits in noise-degraded speech perception. *Hearing Research, 377*, 189–195. https://doi.org/10.1016/j.heares.2019.03.021

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron, 47*, 9–12. https://doi.org/10.1016/j.neuron.2005.06.013

Zentner, M., & Strauss, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-PROMS and the Mini-PROMS. *Annals of the New York Academy of Sciences, 1400*(1), 33–45. https://doi.org/10.1111/nyas.13410

Zhu, J., Chen, X., & Yang, Y. (2021). Effects of Amateur Musical Experience on Categorical Perception of Lexical Tones by Native Chinese Adults: An ERP Study. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.611189