AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychomusicology: Music, Mind, and Brain

ISSN: 0275-3987

2022, Vol. 32, No. 1-2, 1–6
https://doi.org/10.1037/pmu0000284

Song Properties and Familiarity Affect Speech Recognition in Musical Noise

Jane A. Brown[1, 2] and Gavin M. Bidelman[1, 2, 3]

[1] School of Communication Sciences and Disorders, University of Memphis
[2] Institute for Intelligent Systems, University of Memphis
[3] Department of Anatomy and Neurobiology, University of Tennessee Health Sciences Center

"Cocktail party" speech perception is largely studied using either linguistic or nonspeech noise maskers. Few studies have addressed how listeners understand speech during concurrent music. We used popular songs to probe the effects of familiarity and different inherent properties of background music (i.e., isolated vocals, isolated instruments, or unprocessed song) on speech recognition. Participants performed an open-set sentence recognition task in the presence of familiar and unfamiliar music maskers (−5 dB signal-to-noise ratio [SNR]) composed of the full unprocessed song, only the instrumentals, or only the vocals. We found that full songs negatively affected recognition performance more so than isolated vocals and instrumentals. Surprisingly, there was also an interaction with music familiarity; well-known music impaired performance in the homologous full song and instrumental conditions. Our results show strong effects of song component and familiarity on speech recognition ability, highlighting interactions between both physical and psychological characteristics of musical noise on task performance. Familiarity impairs speech perception when background music features the instrumentals with or without the vocals. Our findings have implications for understanding the possible facilitation (or interference) of background music during concurrent linguistic tasks including academic study in attempts to promote learning.

Keywords: speech in noise, familiarity, music-language transfer effects, cocktail party scenario

Supplemental materials: https://doi.org/10.1037/pmu0000284.supp

In real-life scenarios, listeners constantly face the "cocktail party problem": extracting target speech while suppressing other sounds in a complex auditory scene (Bregman, 1990; Cherry, 1953; Yost, 1997). Many studies investigating the cocktail party effect use multitalker babble (Arons, 2008; Bronkhorst, 2015) to simulate conversational listening in a crowded room. However, naturalistic cocktail party scenarios often include background music (e.g., concerts, restaurants) that can be detrimental to speech intelligibility, especially in older adults (Başkent et al., 2014; Ekström & Borg, 2011). Previous studies probing the cocktail party effect with regard to music have used instrumental melodies (Ekström & Borg, 2011) or synthetic "music-shaped noise" (Eskridge et al., 2012) maskers. The Music-In-Noise Task (Coffey et al., 2019) uses multi-instrumental maskers and melodic targets as an analogue to speech recognition in multitalker babble. Though Music-In-Noise Task performance correlates with speech-in-noise listening abilities, it uses different musical cues (e.g., rhythm, melodic prediction) to measure nonlinguistic auditory stream segregation. Despite the advantages of these assays, they fail to exploit the rich nature of ecologically valid music heard in everyday life.

Background music has well-established effects on behavioral performance in various domains. Listening to music increases the average power exerted by athletes (Atkinson et al., 2004; Chtourou et al., 2012), improves reaction times in driving-related tasks (Beh & Hirst, 1999), influences purchases while shopping (North et al., 1999), and affects the speed of reading (Kallinen, 2002) and drawing (Nittono et al., 2000). The impact of background music on concurrent cognitive tasks is less clear. In a meta-analysis of background music literature, Kämpfe et al. (2011) found that both memory and reading comprehension are impaired as compared with silence. However, other studies have demonstrated beneficial effects of music on concentration (Darrow et al., 2006), general

Editor's Note. Jessica Grahn served as the action editor for this article.—MS

This article was published Online First February 10, 2022.

Jane A. Brown https://orcid.org/0000-0002-2545-3693
Gavin M. Bidelman https://orcid.org/0000-0002-1821-3261

This work was supported by the National Institute on Deafness and Other Communication Disorders of the National Institutes of Health under Award number National Institutes of Health/NIDCD R01DC016267 (Gavin M. Bidelman; gmbdlman@memphis.edu). Gavin M. Bidelman serves on the editorial board of this journal. Online materials can be found at https://bit.ly/3km8bXE.

Correspondence concerning this article should be addressed to Jane A. Brown, School of Communication Sciences and Disorders, University of Memphis, 4055 North Park Loop, Memphis, TN 38152, United States. Email: jbrown64@memphis.edu

intelligence test performance (Cockerton et al., 1997), and linguistic processing (Angel et al., 2010).

Presumably, these effects are moderated by acoustic and perceptual components of the background music. For example, reading comprehension is impaired when music is faster and louder (Thompson et al., 2012), and spatial ability is affected by music that is faster or in a major key (Husain et al., 2002). The arousal–mood hypothesis (Husain et al., 2002; Thompson et al., 2001) suggests these components influence arousal level and mood, which in turn affect task performance.

Similarly, many studies (Crawford & Strapp, 1994; Darrow et al., 2006; Martin et al., 1988; Perham & Currie, 2014) showed that the presence of lyrics in background music negatively affects cognitive task performance more than instrumental music. Lyrics introduce meaningful linguistic input, resulting in a switch from energetic masking (just instrumental) to informational masking. Even without lyrics (melodies sung on a "la" syllable), listeners show enhanced memory for vocal melodies (Weiss et al., 2012), perhaps demonstrating stronger processing of vocal sounds in general (Weiss et al., 2017).

Evidence surrounding the role of familiarity in background music is also mixed. Using pupillometry, Weiss et al. (2016) showed increased arousal for familiar versus unfamiliar music, as well as for vocal versus instrumental melodies. This increased physiological arousal for familiar music improves concurrent behavioral tasks requiring concentration (Fontaine & Schwalm, 1979). These results are supported by the arousal–mood hypothesis (Thompson et al., 2001) whereby music's influence on behavioral performance might be moderated by arousal. Familiar background music also improves word identification (Russo & Pichora-Fuller, 2008) and semantic congruity judgements (Feng & Bidelman, 2015) as compared with unfamiliar music. However, de Groot and Smedinga (2014) showed a detrimental effect of familiar background lyrics on foreign language learning. Familiar music also decreases acceptable background noise levels during speech processing (Ahn et al., 2015), indicating listeners are more tolerant of the competing signal. On the contrary, familiar music evokes strong autobiographical memories (Janata et al., 2007) and intense emotions (Ali & Peynircioglu, 2010), which may distract from target speech processing. Altogether, the above findings are somewhat equivocal in describing the effects of background music and its familiarity on concurrent cognitive tasks. One explanation for this may be the varying definitions of "familiarity" used across studies. "Familiar" stimuli have ranged from melodies heard only a single time (Weiss et al., 2016) to music heard countless times in everyday life (Russo & Pichora-Fuller, 2008).

The current study sought to investigate the role of different song components and familiarity in speech recognition. Participants completed a sentence recognition task in the presence of music maskers composed of popular songs that were either familiar or unfamiliar. In addition, the maskers were composed of different song aspects (the full song, the isolated instrumentals, or the isolated vocals) in order to probe their specific roles in masking sentences. We hypothesized that speech recognition would be poorest in the presence of full song maskers and also suffer in the isolated vocals condition due to the introduction of competing linguistic information (Martin et al., 1988) that uses more processing resources (Weiss et al., 2016). We also predicted better performance when the songs were identified as familiar versus unfamiliar (Feng & Bidelman, 2015; Russo & Pichora-Fuller, 2008), supporting the notion that familiarity enhances concurrent linguistic processing.

Method

Participants

The sample included 24 young adults aged 19 to 31 (M = 24.7, SD = 3.1 years; 10 male) recruited via social media advertising to participate in the online experiment. Because all song conditions were counterbalanced across participants, a sample size of 24 allowed for four full counterbalanced rotations. All self-reported normal hearing and English as their native language and had an average of 17.1 ± 2.4 years of formal education. 20 listeners reported having musical training (M = 9.3, SD = 4.6 years). Each was paid for their time and gave written informed consent in compliance with a protocol approved by the Institutional Review Board at the University of Memphis.

Stimuli

Speech and Music Materials

Speech recognition performance was measured using the Harvard IEEE sentences (Rothauser, 1969) presented in various forms of background music. Music maskers were selected from popular songs identified in pilot testing as having a high/low degree of familiarity. A pilot sample of N = 15 listeners (age: 22–36 years; M = 27.1, SD = 4.8) heard 20 songs and rated each song on a 5-point scale (1 = *not familiar*, 5 = *extremely familiar*). The two songs with the lowest and highest familiarity scores were chosen for our "Unfamiliar" and "Familiar" conditions, respectively. The Unfamiliar songs were "Play With Fire" by Hilary Duff (rating M = 1.13, SD = .34) and "Warmer in the Winter by Lindsey Stirling (M = 1.13, SD = .50), and the Familiar songs were "Rockin' Around the Christmas Tree" by Brenda Lee (M = 4.73, SD = 1) and "Just Dance" by Lady Gaga (M = 4.13, SD = 1.36). Conditions were externally validated using data from the Billboard Hot 100 chart history (billboard.com). "Rockin' Around the Christmas Tree" peaked at number two on the chart in 2019, and "Just Dance" peaked at number one in 2009; neither of the unfamiliar songs have been listed on the Hot 100 chart (www.billboard.com/charts/hot-100).

Vocal/Instrumental Isolation

We processed full song files using a machine learning algorithm (LALAL.AI; available at www.lalal.ai) trained to separate instrumental and vocal stems in audio files (sample stimuli available in online supplemental materials). This allowed us to "unmix" the studio-recorded music clips and isolate the vocals and instrumentals from the aggregate recording. All music files (four songs each with the full song, isolated vocals, and isolated instrumentals) were sampled at 44,100 Hz, converted to mono for diotic presentation, and root mean square amplitude normalized after silences at the beginning and the end of each sound clip were deleted. Though not critical given the suprathreshold nature of our task, exact presentation level was necessarily dependent on listeners' at-home audio configuration (e.g., soundcard, headphones) and was thus unknown. Nevertheless, estimated output level based on laboratory calibrations was ~70 dB sound pressure level (Sennheiser HD280 Pro headphones). Participants were instructed to adjust their computer volume to 75% full scale for the duration of the experiment. The stimulus presentation was piloted at 0dB SNR, but participants reached ceiling results at this level for 90% of trials across conditions. Consequently, stimuli were presented at −5dB SNR for the experiment proper, which decreased the ceiling trials across conditions to a range of 55% to 75%.[1]

Task

The experiment was completed remotely. Listeners downloaded and ran the behavioral task (described in the following text) on their personal computer (17 MAC, 7 PC). The paradigm was coded in MATLAB 2020 and compiled into a standalone executable application for local runtime deployment using the MATLAB Compiler (The MathWorks, Natick, MA). Participants were instructed to wear headphones (14 in-ear buds, nine circumaural, one not reported).

Stimuli were presented in three blocks of 120 sentences (30 sentences per masker). Each block featured the concurrent music under one of the three song properties (i.e., instrumentals alone, vocals alone, full song = vocals + instrumentals). Condition order and song order within each block were counterbalanced across participants. On each trial, participants were asked to recall each sentence by typing their response on the screen. They pressed ENTER to advance to the next trial. The experiment paused every 30 trials. Each stimulus began at the first trial in the block and played continuously for the duration of the 30 sentences.

Following the experimental task, participants completed a survey to rate their familiarity with each of the four songs as in the pilot study. They were also prompted to name the song and/or artist if they were able.

Statistical Analysis

We logged the number of correctly recalled keywords (out of five) per sentence. Percent correct scores were rationalized arcsine transformed (Studebaker, 1985) to account for correlations of the variances and mean in proportionate scores. Data were analyzed using a maximal linear mixed-effects analysis of variance (ANOVA) model (Barr et al., 2013) via the lme4 package in *R* (R Core-Team, 2020). Fixed effects were song condition (three levels: full song, instrumentals, vocals) and familiarity (two levels: familiar vs. unfamiliar). Subjects, sentence item, and condition order served as random effects. For nonsignificant effects determined via frequentist statistics, we also conducted Bayes Factor analysis using the BayesFactor toolbox (Morey & Rouder, 2018) to assess the likelihood of evidence in support of or against the null hypothesis.

We used the MIR toolbox (Lartillot & Toiviainen, 2007) to characterize various acoustic measures of the background music maskers: beats per minute (BPM), beat strength (relative importance of metric pulse clarity), spectral centroid (average frequency of the power spectrum), event density (average number of auditory events, or local temporal envelope peaks, per second), and spectral spread (standard deviation of spectral power). These analyses allowed us to investigate the explanatory power of salient acoustic features on task performance, as well as any interactions with variables of interest (song condition and familiarity).

Results

Participants rated the songs "Just Dance" and "Rockin' Around the Christmas Tree" as more familiar than "Play With Fire" and "Warmer in the Winter," $t(23) = 24.98$, $p < .0001$, validating our preselected song familiarity categories (Figure 1). Furthermore, participants correctly identified the song title more often for the familiar (92%) than for unfamiliar (2%) songs. A 2×3 ANOVA showed speech recognition was strongly modulated by song condition, $F(2, 42.64) = 31.62$, $p < .0001$, $\eta_p^2 = .60$, and the familiarity of background music, $F(1, 25.04) = 38.94$, $p < .001$, $\eta_p^2 = .61$ (Figure 2). Post hoc contrasts revealed performance in the full song condition was worse than in the isolated instrumentals ($z = -7.28$, $p < .0001$) and isolated vocals ($z = -6.44$, $p < .0001$). There was no difference between the instrumental and vocal conditions ($z = .85$, $p = .671$); a $BF_{H0} = .297$ indicated moderate evidence favoring the null for this contrast (Lee & Wagenmakers, 2014).

More critically, there was Condition × Familiarity interaction, $F(2, 46.39) = 13.32$, $p < .0001$, $\eta_p^2 = .36$. Post hoc Tukey-adjusted contrasts revealed speech recognition was poorer for familiar versus unfamiliar music when the background was the full song ($z = -7.91$, $p < .0001$) and instrumentals alone ($z = -2.46$, $p = .014$). There was no familiarity effect in the isolated vocals condition ($z = -1.49$, $p = .14$; $BF_{H0} = .56$, anecdotal evidence favoring the null).

We used a simple linear regression model to investigate relations between salient acoustic properties of the music maskers and task performance. Neither spectral centroid, $F(1, 8530) = .007$, $p = .932$, nor BPM, $F(1, 8530) = .064$, $p = .801$, impacted speech recognition performance. There was a significant effect of pulse, $F(1, 8530) = 54.85$, $p < .0001$; music with stronger beat periodicity strength decreased performance. However, the effect size was negligibly small ($\eta^2 = .004$). There were also significant effects of event density, $F(1, 8547) = 172.82$, $p < .0001$, $\eta^2 = .012$, and spectral spread, $F(1, 8547) = 4.433$, $p = .035$, $\eta^2 = .019$. Consequently, we reran the omnibus ANOVA above adding density and spread as acoustic covariates. Spectral spread was no longer a significant predictor of task performance once factored into the full model, $F(1, 8481.5) = .026$, $p = .873$. Event density remained significant, $F(1, 8460.7) = 5.118$, $p < .0237$; $\eta^2 = .0006$, suggesting that music selections with more temporal events per unit time hindered speech recognition performance. However, the much larger (by several orders of magnitude) effect sizes of song condition ($\eta^2 = .42$), familiarity ($\eta^2 = .35$), and their interaction ($\eta^2 = .22$) confirmed that those variables were more prominent drivers of listeners' sentence recognition scores than acoustic properties of the music maskers alone.

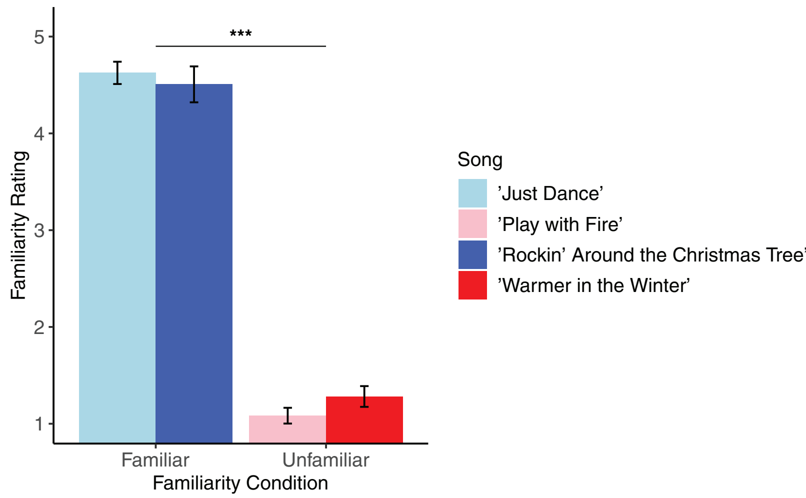
Discussion

We used a speech recognition task in the presence of popular music maskers to probe the effect of background music on speech-language processing. By varying the musical components (vocals, instrumentals, or both) and the familiarity of the background music, we were able to parse the roles that those features play in perceptually interfering with concurrent speech. Speech recognition was more severely impaired by

[1] The proportion of ceiling results for each cell of the final design were: familiar song (0.55), familiar instrumental (0.76), familiar vocals (0.73), unfamiliar song (0.65), unfamiliar instrumentals (0.75), and unfamiliar vocals (0.73).

Figure 1

Mean Participant Familiarity Ratings for Songs in Each Familiarity Condition



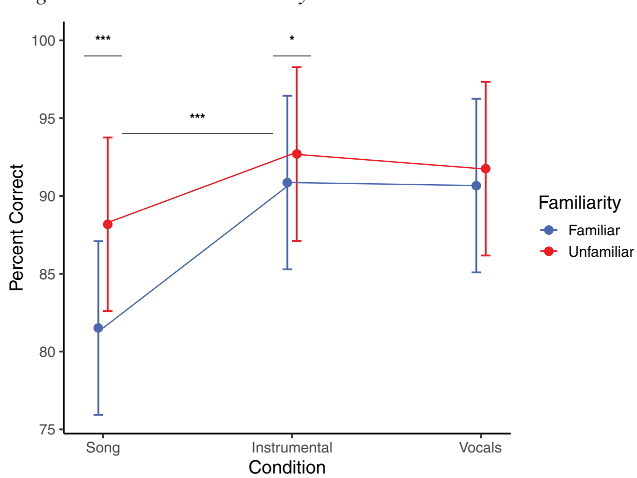
Note. 1 = *not familiar*; 5 = *extremely familiar*. Error bars = ±1 s.e.m. See the online article for the color version of this figure.

full songs than the isolated instrumentals or vocals. Familiar background music inhibited speech recognition more than unfamiliar music, and the effect was seen only when the maskers were full songs or solely instrumentals. In addition, this effect was much larger for the full songs than the instrumentals.

The arousal–mood hypothesis (Thompson et al., 2001) would predict that conditions provoking the higher levels of arousal (familiar music and isolated vocals) would result in enhanced recognition scores. However, this was not what we found in the present study. Poorer task performance in the familiar music condition may be a function of attention resource allocation. The load theory of attention (Lavie et al., 2004) posits that as demand on higher cognitive functions increases, perceptual control (i.e., the ability to suppress distractor interference) is impaired, resulting in poorer selective attention. Here, our paradigm required recognizing sentences in noise while tuning out music distractors. It is possible that familiar music may be more distracting because song recognition draws on cognitive resources. This would account for the overall worse performance in the presence of familiar songs found in this study. Arousal seems to benefit performance by overcoming the detrimental effects of decreased cognitive resources (Thompson et al., 2012). It may be the case here that the conditions showing the poorest performance (familiar, full song, etc.) required enough attentional resources—that is, the background music was so distracting—that any benefits from increased arousal were suppressed.

Figure 2

Performance on Sentence Recognition Task as a Function of Song Condition and Familiarity



Note. Error bars = 95% confidence intervals. The longer horizontal significance line represents the effect of condition. See the online article for the color version of this figure.

* $p < .05$. *** $p < .0001$.

Load theory may also account for the interaction between condition and familiarity, where music familiarity played a detrimental role only in the full song and instrumental conditions. Kraemer et al. (2005) found that when silent gaps were inserted in familiar songs, greater primary auditory cortical activation was induced as compared with unfamiliar songs (true for songs with and without lyrics). Participants also reported hearing a continuation of the familiar songs during these gaps. This continuing auditory imagery suggests listeners might mentally follow (covertly "sing") along with the familiar music, recruiting primary and associated auditory regions (Zatorre & Halpern, 2005). Recognition and distraction of the familiar music, in addition to the concurrent "singing along" to the song, places an increased load on both cognitive and sensory resources (Herholz et al., 2012; Zatorre & Halpern, 2005), thus reducing ability to suppress the background music distractors as observed in our data.

Our finding that familiarity negatively affects performance contradicts previous studies suggesting familiar background music is beneficial to speech perception. For example, Russo and Pichora-Fuller (2008) suggested that familiar background music generates stronger expectancies that tax fewer cognitive resources and carry over to the attended stream, thus enhancing speech recognition. Feng and

Bidelman (2015) showed similar results with familiar music causing faster lexical semantic judgements and fewer mind-wandering episodes. However, both studies used solely instrumental music, as well as different tasks (word identification and semantic judgment tasks, respectively). It is possible that the more complex demands of sentence-level processing (current study) elicit stronger familiarity effects. Russo and Pichora-Fuller (2008) used pop songs, but only instrumental versions. In the current study, the familiarity effects were strongest in the full song condition, which is the version of the song listeners would have been exposed to. Although listeners recognize the instrumental version, the full songs may be considered more "familiar" to the listener. These versions may even be associated with strong emotions and/or memories (Janata et al., 2007), which would make them more distracting. Future studies examining a variety of auditory–linguistic tasks are needed to test these possibilities.

It is important to acknowledge that this study was conducted online, and while participants were instructed to attend to the target sentences, it is possible they may have gone off-task and listened more closely to familiar music, which would explain the unexpected negative familiarity effect found here. However, mind wandering occurs more frequently in unfamiliar background music (Feng & Bidelman, 2015). The online format of the experiment also limited playback volume control. Participants were instructed to set their computer volume to 75% for the duration of the experiment, but this does not ensure uniform loudness and sound levels across speaker/headphone systems. Intersubject variability due to loudness may have impacted our ability to detect some of the smaller effects in our study design.

Although all music clips were matched in overall amplitude, their acoustic features change in a time-varying manner. For example, perceived loudness may vary on short time scales, especially between the isolated vocals and the other two conditions. Louder background music negatively affects concurrent cognitive performance (Thompson et al., 2012), which may account for the poor speech recognition in our full song condition if they were indeed perceived as louder than the other two song conditions. In addition, natural gaps in the isolated vocals conditions may provide multiple looks (Viemeister & Wakefield, 1991) to glimpse speech from the target sentences, thus increasing recognition. It is important to note that while the time-varying nature of our stimuli is somewhat captured by frame decomposition in our acoustic calculations, stimulus features were averaged over the full recording, not at the local level.

Better recognition performance in the instrumental condition versus the full song also suggests that the presence of vocals hinders speech recognition. This may be a result of vocals introducing informational masking whereby linguistic cues in both the target and masker conflict and hinder performance. However, arousal is greater in response to vocal melodies than to instrumental melodies (Weiss et al., 2016), which may improve attention to the speech target. Conceivably, this increased arousal may overcome the informational masking, thus preserving speech recognition. In addition to the natural gaps allowing multiple looks (Viemeister & Wakefield, 1991), the AI algorithm separation of the vocal stem from the full song created amplitude artifacts in the isolated vocal stimuli. Though the songs were still easily recognizable, the fluctuations may have continuously interrupted auditory streaming and thus reduced interference on sentence recognition (Fiveash et al., 2018). Increased arousal to vocal melodies, natural silences in the vocals, and the amplitude artifacts may all allow the listener to overcome the informational masking, thus preserving speech recognition. An ideal isolated vocal stimulus would originate from a multitrack (studio) recording of the original song, but these are often not publicly available.

Further supporting our claim of a purely cognitive familiarity effect, neither tempo (BPM) nor spectral centroid (average frequency on power spectrum) was a significant predictor of task performance. Pulse (beat strength) did predict performance but with a negligible effect size. In addition, the isolated vocals condition had a much lower average pulse (.164) than the full song (.448) and instrumental (.414) conditions. Thus, the significant pulse effect is likely confounded with the condition factor, which showed a similar pattern. This was also the case with density, where the full song had a higher event density value (2.48) than the instrumental (1.97) and vocal (1.88) conditions, again resembling the behavioral results.

In summary, our results demonstrate that familiar music negatively affects sentence recognition but only when it contains the instrumental element of the music (i.e., vocals alone fail to impact speech understanding). In addition, sentence recognition is worse when listening to a full song than when listening to its instrumentals or isolated vocals. To our knowledge, this is the first study using AI audio separation to examine how the vocal and instrumental components of popular songs independently affect concurrent speech perception. Using natural stimuli derived from popular songs increases the ecological validity of this work, and may inform future studies aiming to utilize more realistic "cocktail party" scenarios.

References

Ahn, H. J., Bahng, J., & Lee, J. H. (2015). Measurement of acceptable noise level with background music. *Journal of Audiology and Otology*, *19*(2), 79–84. https://doi.org/10.7874/jao.2015.19.2.79

Ali, S. O., & Peynircioglu, Z. F. (2010). Intensity of emotions conveyed and elicited by familiar and unfamiliar music. *Music Perception*, *27*(3), 177–182. https://doi.org/10.1525/mp.2010.27.3.177

Angel, L. A., Polzella, D. J., & Elvers, G. C. (2010). Background music and cognitive performance. *Perceptual and Motor Skills*, *110*(3_suppl), 1059–1064. https://doi.org/10.2466/pms.110.C.1059-1064

Arons, B. (2008). A review of the cocktail party effect. *Journal of American Voice I/O Society*, *12*(7), 35–50.

Atkinson, G., Wilson, D., & Eubank, M. (2004). Effects of music on work-rate distribution during a cycling time trial. *International Journal of Sports Medicine*, *25*(8), 611–615. https://doi.org/10.1055/s-2004-815715

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Başkent, D., van Engelshoven, S., & Galvin, J. J., III. (2014). Susceptibility to interference by music and speech maskers in middle-aged adults. *The Journal of the Acoustical Society of America*, *135*(3), EL147–EL153. https://doi.org/10.1121/1.4865261

Beh, H. C., & Hirst, R. (1999). Performance on driving-related tasks during music. *Ergonomics*, *42*(8), 1087–1098. https://doi.org/10.1080/001401399185153

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press. https://doi.org/10.7551/mitpress/1486.001.0001

Bronkhorst, A. W. (2015). The cocktail-party problem revisited: Early processing and selection of multi-talker speech. *Attention, Perception and Psychophysics*, *77*(5), 1465–1487. https://doi.org/10.3758/s13414-015-0882-9

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*(5), 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*(6), e85–e88. https://doi.org/10.1016/j.scispo.2012.02.006

Cockerton, T., Moore, S., & Norman, D. (1997). Cognitive test performance and background music. *Perceptual and Motor Skills*, *85*(3_suppl), 1435–1438. https://doi.org/10.2466/pms.1997.85.3f.1435

Coffey, E. B. J., Arseneau-Bruneau, I., Zhang, X., & Zatorre, R. J. (2019). The Music-In-Noise Task (MINT): A tool for dissecting complex auditory perception. *Frontiers in Neuroscience*, *13*, Article 199. https://doi.org/10.3389/fnins.2019.00199

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*(2), 237–245. https://doi.org/10.1016/0191-8869(94)90162-7

Darrow, A. A., Johnson, C., Agnew, S., Fuller, E. R., & Uchisaka, M. (2006). Effect of preferred music as a distraction on music majors' and nonmusic majors' selective attention. *Bulletin of the Council for Research in Music Education*, *170*, 21–31.

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*(4), 681–707. https://doi.org/10.1017/S0272263114000059

Ekström, S.-R., & Borg, E. (2011). Hearing speech in music. *Noise and Health*, *13*(53), 277–285. https://doi.org/10.4103/1463-1741.82960

Eskridge, E. N., Galvin, J. J., III Aronoff, J. M., Li, T., & Fu, Q. J. (2012). Speech perception with music maskers by cochlear implant users and normal-hearing listeners. *Journal of Speech, Language, and Hearing Research*, *55*(3), 800–810. https://doi.org/10.1044/1092-4388(2011/11-0124)

Feng, S., & Bidelman, G. M. (2015). Music listening and song familiarity modulate mind wandering and behavioral success during lexical processing. In D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio. (Eds.) *Proceedings of the 37th Annual Meeting of the Cognitive Science Society. Pasadena, California, July 22-25.*

Fiveash, A., McArthur, G., & Thompson, W. F. (2018). Syntactic and non-syntactic sources of interference by music on language processing. *Scientific Reports*, *8*(1), Article 17918. https://doi.org/10.1038/s41598-018-36076-x

Fontaine, C. W., & Schwalm, N. D. (1979). Effects of familiarity of music on vigilant performance. *Perceptual and Motor Skills*, *49*(1), 71–74. https://doi.org/10.2466/pms.1979.49.1.71

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience*, *24*(6), 1382–1397. https://doi.org/10.1162/jocn_a_00216

Husain, G., Thompson, W., & Schellenberg, E. (2002). Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Perception*, *20*(2), 151–171. https://doi.org/10.1525/mp.2002.20.2.151

Janata, P., Tomic, S. T., & Rakowski, S. K. (2007). Characterization of music-evoked autobiographical memories. *Memory*, *15*(8), 845–860. https://doi.org/10.1080/09658210701734593

Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: Effects of the tempo of background music. *Computers in Human Behavior*, *18*(5), 537–551. https://doi.org/10.1016/S0747-5632(02)00005-5

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2011). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music*, *39*(4), 424–448. https://doi.org/10.1177/0305735610376261

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: Sound of silence activates auditory cortex. *Nature*, *434*(7030), Article 158. https://doi.org/10.1038/434158a

Lartillot, O., & Toiviainen, P. (2007). A matlab toolbox for musical feature extraction from audio. In *Proceeding of the 10th International Conference on Digital Audio Effects (DAFx-07), Bordeaux, France, September 10-15.*

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*(3), 339–354. https://doi.org/10.1037/0096-3445.133.3.339

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Martin, R. C., Wogalter, M. S., & Forlano, J. G. (1988). Reading comprehension in the presence of unattended speech and music. *Journal of Memory and Language*, *27*(4), 382–398. https://doi.org/10.1016/0749-596X(88)90063-0

Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. *R package version 0.9.12-4.2*. https://CRAN.R-project.org/package=BayesFactor

Nittono, H., Tsuda, A., Akai, S., & Nakajima, Y. (2000). Tempo of background sound and performance speed. *Perceptual and Motor Skills*, *90*(3_suppl), 1122. https://doi.org/10.2466/pms.2000.90.3c.1122

North, A. C., Hargreaves, D. J., & McKendrick, J. (1999). The influence of in-store music on wine selections. *Journal of Applied Psychology*, *84*(2), 271–276. https://doi.org/10.1037/0021-9010.84.2.271

Perham, N., & Currie, H. (2014). Does listening to preferred music improve reading comprehension performance? *Applied Cognitive Psychology*, *28*(2), 279–284. https://doi.org/10.1002/acp.2994

R Core-Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rothauser, E. H. (1969). IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio and Electroacoustics*, *17*(3), 225–246. https://doi.org/10.1109/TAU.1969.1162058

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*(5), 746–760. https://doi.org/10.1097/AUD.0b013e31817bdd1f

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *Journal of Speech, Language, and Hearing Research*, *28*(3), 455–462. https://doi.org/10.1044/jshr.2803.455

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2001). Arousal, mood, and the Mozart effect. *Psychological Science*, *12*(3), 248–251. https://doi.org/10.1111/1467-9280.00345

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*(6), 700–708. https://doi.org/10.1177/0305735611400173

Viemeister, N. F., & Wakefield, G. H. (1991). Temporal integration and multiple looks. *Journal of the Acoustical Society of America*, *90*(2), 858–865. https://doi.org/10.1121/1.401953

Weiss, M., Schellenberg, E., & Trehub, S. (2017). Generality of the memory advantage for vocal melodies. *Music Perception*, *34*(3), 313–318. https://doi.org/10.1525/mp.2017.34.3.313

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science*, *23*(10), 1074–1078. https://doi.org/10.1177/0956797612442552

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*, *42*(8), 1061–1065. https://doi.org/10.1037/xhp0000226

Yost, W. (1997). The cocktail party problem: forty years later. In R. H. Gilkey & T. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments* (pp. 329–347). Erlbaum Inc.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron*, *47*(1), 9–12. https://doi.org/10.1016/j.neuron.2005.06.013

Received May 13, 2021
Revision received October 20, 2021
Accepted November 4, 2021 ■