NeuroImage 252 (2022) 119049

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Familiarity modulates neural tracking of sung and spoken utterances

Christina M. Vanden Bosch der Nederlanden[a], Marc F. Joanisse[a,b], Jessica A. Grahn[a,b], Tineke M. Snijders[c,d], Jan-Mathijs Schoffelen[d,*]

[a] *The Brain and Mind Institute, The University of Western Ontario, London, Ontario, Canada*
[b] *Psychology Department, The University of Western Ontario, London, Ontario, Canada*
[c] *Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands*
[d] *Radboud University, Donders Institute for Brain, Cognition and Behaviour, the Netherlands*

ARTICLE INFO

Keywords:
Neural tracking
Song
Speech
Memory
Familiarity

ABSTRACT

Music is often described in the laboratory and in the classroom as a beneficial tool for memory encoding and retention, with a particularly strong effect when words are sung to familiar compared to unfamiliar melodies. However, the neural mechanisms underlying this memory benefit, especially for benefits related to familiar music are not well understood. The current study examined whether neural tracking of the slow syllable rhythms of speech and song is modulated by melody familiarity. Participants became familiar with twelve novel melodies over four days prior to MEG testing. Neural tracking of the same utterances spoken and sung revealed greater cerebro-acoustic phase coherence for sung compared to spoken utterances, but did not show an effect of familiar melody when stimuli were grouped by their assigned (trained) familiarity. However, when participant's subjective ratings of perceived familiarity were used to group stimuli, a large effect of familiarity was observed. This effect was not specific to song, as it was observed in both sung and spoken utterances. Exploratory analyses revealed some in-session learning of unfamiliar and spoken utterances, with increased neural tracking for untrained stimuli by the end of the MEG testing session. Our results indicate that top-down factors like familiarity are strong modulators of neural tracking for music and language. Participants' neural tracking was related to their perception of familiarity, which was likely driven by a combination of effects from repeated listening, stimulus-specific melodic simplicity, and individual differences. Beyond simply the acoustic features of music, top-down factors built into the music listening experience, like repetition and familiarity, play a large role in the way we attend to and encode information presented in a musical context.

1. Introduction

Language and music are two of the most important means of communication in everyday life. Speaking and singing draw on the specialized knowledge for the domains of language and music to create meaning amongst the elements of spoken (e.g., who is doing what to whom?) and sung (e.g., what pitch will arrive next to resolve melodic tension?) utterances (Patel, 2003; Peretz, 2009; Jackendoff, 2008). Song is a special instance of music that contains semantically meaningful speech sounds. The acoustic form of sung compared to spoken utterances takes on different characteristics, such as rhythmic regularity, discrete pitch movements, metrical structure, and tonal expectancy (Patel, 2003; Kuroyanagi et al., 2019; Tierney et al., 2018). Speech and song comparisons allow for a unique opportunity to examine how the processing of a single dimension—in this case, semantically meaningful lyrics—is altered by whether utterances are heard in a music or language setting. These comparisons are now commonplace in the literature for elucidating domain-specific and domain-general processing using both natural utterances (e.g., Gordon et al., 2010, 2011; Slevc, 2009) and boundary condition stimuli such as the speech-to-song illusion (Deutsch et al, 2011; Tierney et al, 2012; Vanden Bosch der Nederlanden et al., 2015a, 2015b).

A growing body of literature examines the closely related question of how the musical acoustic features of song might be leveraged to benefit the processing of speech (e.g., Falk and Dalla Bella, 2016; Rathcke et al., 2021; Moore et al., 2017; Vanden Bosch der Nederlanden et al., 2020). All of these studies find better processing of speech when it is either rhythmically presented or when utterances are sung rather than spoken. Within this body of literature is the well-established finding of music's mnemonic effect on speech; that is, better memory for words learned from song than speech (Chazin and Neuschatz, 1990; Wallace, 1994; Rainey and Larson, 2002; Purnell-Webb and Speelman, 2008; Good et al., 2015; Tamminen et al., 2017). However, unlike the above work finding a speech processing benefit for highly rhyth-

[*] Corresponding author.
E-mail address: jan.schoffelen@donders.ru.nl (J.-M. Schoffelen).

https://doi.org/10.1016/j.neuroimage.2022.119049.
Received 27 August 2021; Received in revised form 11 February 2022; Accepted 1 March 2022
Available online 4 March 2022.

mic utterances, this body of literature also highlights the importance of familiar music for a pronounced mnemonic effect (Wallace, 1994; Moussard et al., 2012; Tamminen et al., 2017). Previous work from our lab has shown better neural tracking of sung compared to spoken utterances in difficult listening conditions (Vanden Bosch der Nederlanden et al., 2020), but no benefit for neural tracking of normal rate speech. The current study will examine whether, like the pronounced memory benefit for familiar compared to unfamiliar songs and spoken utterances, neural tracking might be greater for words sung to familiar melodies compared to unfamiliar melodies or spoken words.

The memory advantage for song over speech is often discussed in terms of spared memory for music compared to other memory types in people with Alzheimer's disease (Jacobsen et al., 2015; Cuddy and Duffin, 2005). These studies suggest a distinct mechanism for musical memories that undergoes a slower rate of degeneration than other types of memory (Simmons-Stern et al., 2010, 2012; Baird and Samson, 2009). One Alzheimer's case study found better memory for words set to familiar songs and only found an effect for unfamiliar songs after repeated learning sessions, presumably after the melody had become more familiar or predictable (Moussard et al, 2012). The evidence for spared familiar music processing is mixed, with some studies providing evidence of spared long-known or familiar music (e.g., Jacobsen et al., 2015) and others showing only a sparing of implicit musical memory such as how to play the piano (Baird and Samson, 2009). As described above, familiarity seems to play a special role in the encoding of words into memory for healthy adults, as well. Verbatim recall is better for song than speech (Calvert and Tart, 1993; Kilgouret al., 2000; Wallace, 1994), with particularly strong effects for familiar songs, such that some melodies show no verbal memory benefit unless participants are familiar with them (Wallace, 1994; Calvert and Tart, 1993; Tamminen et al., 2017). Familiarity with a melody may draw on the listener's "veridical" knowledge of the music (Bharucha, 1987) enabling them to anticipate both which note will come next, and when in time it will arrive, ultimately benefitting the encoding of words sung to the familiar melody.

The memory benefit for sung over spoken materials is not specific to the laboratory and the clinic. Music is a well-established tool in the classroom for increasing engagement and enjoyment in learning process, with a potential benefit on memory retention. Teachers often put hard-to-remember lists of words together into a song format (e.g., learning to sing states/provinces, foreign language prepositions, or even best mining practices; (Alisaari and Heikkola, 2017; Veiga et al., 2015). Teachers also recycle melodies by putting to-be-remembered words to familiar songs (e.g., Engh, 2013). For example, the alphabet is sung to the same melody as Twinkle, Twinkle Little Star and German dative prepositions [aus, ausser, bei, mit, nach, seit, von, zu] can be sung to the Blue Danube Waltz. The enhanced predictability of familiar melodies could facilitate recall by providing an easy-to-recall unified structure (i.e., the melody) for the to-be-remembered words (i.e., chunking; McElhinney and Annett, 1996), but the predictability of familiar music could also benefit initial neural encoding of the words presented as part of the melody. Although several studies examine the distributed and distinct networks for familiar and unfamiliar musical and linguistic memory (Cuddy and Duffin, 2005; Finke et al., 2012; Saito et al., 2012; Jacobsen et al., 2015; Sternin et al., 2021), the neural mechanisms by which familiar melodies are encoded differently from unfamiliar melodies or speech have not been investigated. One plausible mechanism is neural entrainment (in the "broad sense," see Obleser and Kayser 2019), brought about by tighter alignment of neural activity to the highly predictable rhythms of a familiar melody compared to an unfamiliar melody or irregular speech rhythm.

Over the past decade, there has been considerable interest in how humans neurally track rhythmic information in the environment, particularly the slow rhythms of speech (Luo and Poeppel, 2007). Regardless of whether neural tracking to speech is indicative of the phase resetting of ongoing neural oscillations or to a stimulus-driven response to sound onsets (Meyeret al., 2019; Haegens, 2020), a growing body of evidence suggests that better neural tracking of syllable-level rhythms of speech relates to better comprehension (Peelle et al., 2013; Doelling et al., 2014; Park et al., 2015). The relationship between neural tracking and speech comprehension suggests that speech comprehension may be increased by improving the neural tracking of speech rhythms (Zoefel and Van Rullen, 2015; Zoefel et al., 2020). Since neural tracking is particularly sensitive to rhythmic regularity (e.g., Kayser et al., 2015; Meyer et al., 2019), our previous work examined whether the rhythmic regularity of music improved neural tracking of words set to song compared to spoken words. As mentioned above, sung utterances elicited greater neural tracking in the theta band than spoken utterances, but only when utterances were time-compressed (50%), significantly impairing intelligibility (Vanden Bosch der Nederlanden et al., 2020). These findings suggest that song may aid in the processing of syllable rhythms for difficult listening conditions, but not for normal rate speech. This runs counter to many behavioural accounts of musical enhancement of normal-rate speech processing due to bottom-up factors such as music's rhythmic regularity (Falk and Dalla Bella, 2016; Rathcke et al., 2021; Moore et al., 2017).

Top-down factors, such as the direction of attention, are potentially more powerful modulators of neural tracking (Obleser and Kayser, 2019) than the stimulus differences between language and music. There is now a wealth of research showing that when participants are directed to listen to one speech stream over another simultaneously presented speech stream, the attended stream receives greater neural tracking than the unattended stream (Kerlin et al., 2010; Ding and Simon, 2012; Golumbic et al., 2013; O'Sullivan et al., 2015; Fuglsang et al., 2017; Rimmele et al., 2015; Fiedler et al., 2019; Vanthornhout et al., 2019). However, other experience-related factors, such as language background, can also affect neural tracking. Participants who listened to non-native speech exhibited greater neural tracking than native speakers of the language despite less intelligibility for non-native speech (Song and Iverson, 2018; Zou et al., 2019; Reetzke et al., 2021), but still showed greater neural tracking to attended compared to unattended stimuli. To our knowledge, the effect of stimulus familiarity has not been assessed in neural tracking research. Thus, it is an open question whether, like the behavioural memory benefit for familiar sung utterances over unfamiliar sung utterances and spoken words, the neural tracking of utterances sung to a familiar melody would be greater than utterances sung to unfamiliar melodies or for spoken words.

Much of the literature on neural processing of music and language has found evidence for right and left lateralized neural responses, respectively (e.g., Zatorre et al., 1992; Zatorre and Gandour, 2008). However, this work relied primarily on musical and linguistic stimuli that were acoustically quite different, for instance, by comparing speech to melodies played by musical instruments. These acoustic differences – which are not specific to either the music or language domains – drive lateralisation (Joanisse and Gati, 2003; Johnsrude et al., 1997). The differences between left and right hemisphere recruitment have been attributed to neural oscillatory properties (Morillon et al., 2010; Giraud et al., 2007), preferred timescale (Poeppel, 2003), and spectrotemporal characteristics (Albouy et al., 2020). All these theoretical approaches suggest that the left hemisphere preferentially processes fast temporal information unfolding on the 10 s of milliseconds (e.g., VOTs) and the right hemisphere preferentially processes slower spectral information on the 100 s of milliseconds (e.g., pitch). A growing number of studies compare music and language while attempting to control for acoustic differences. These studies show either no lateralisation (e.g., Gordon et al., 2010; Rogalsky et al., 2011) or canonical left asymmetries for speech and right for music (Tierney et al., 2012; Albouy et al., 2020). Previous work in neural tracking of music and language does not directly compare language to music (Peelle et al., 2013; Doelling and Poeppel, 2015) thus lateralisation differences are unknown. One recent study directly compared speech to song and found right lateralisation for song but no lateralisation for speech (Vanden Bosch der Nederlanden et al., 2020), however this was based on a limited range of stimuli, which makes it unclear how this pattern would generalize to a larger corpus of spoken and sung utterances.

The current study examined the effect of melody familiarity on the neural tracking and subsequent recall of lyrics by training participants on a set of novel melodies. In the training phase, participants learned piano melodies, with no associated lyrics, by listening to and counting the notes in each melody to ensure active listening for four days prior to the testing session. In the testing phase, each melody had two assigned text settings (i.e., lyrics). Each text setting was presented in a sung and spoken format. It was hypothesized that participants would show greater neural tracking to words associated with the trained (familiar) melodies than words associated with unfamiliar melodies, or when those same words were spoken. To characterize participants' subjective perception of familiarity separately from the familiarity we hoped to achieve by training participants through repeated listening (Bradley, 1971; Madison and Schiölde, 2017), we asked participants to rate their familiarity for each spoken and sung utterance during the testing portion of the study. We did this because an individual's perceived familiarity with music is notoriously entangled in enjoyment and stimulus complexity (Krugman, 1943; Wallace and Rubin, 1991; Fung, 1996; Serra et al., 2012; van den Bosch, et al., 2013). This allowed us to examine how neural tracking related to *assigned* familiarity–based on the stimuli heard during the training session—and *perceived* familiarity—based on participants' ratings of familiarity. Perceived familiarity allows us to examine the influence of participants' training and other acoustic or individual factors, such as enjoyment and stimulus complexity.

We additionally hypothesized that neural tracking would be greater in the right hemisphere for sung utterances and the left for spoken utterances, given previous neural tracking data (Peelle et al., 2013; Doelling and Poeppel, 2015). We also hypothesize greater neural tracking over the right than left hemispheres for familiar sung utterances, as the familiarity is pitch-based (Doelling and Poeppel, 2015; Zatorre et al., 1994). We also predicted that words sung to familiar melodies would be better encoded and retained in memory, such that a test of memory for lyrics put to familiar utterances would have greater accuracy than words put to unfamiliar melodies. As better memory for words set to music is particularly successful for improving verbatim recall (Calvert and Tart, 1993; Kilgour et al., 2000; Wallace, 1994), we examined whether familiarity would be related to verbatim recall (via the hard change condition with a single word changed from the original lyric) or whether the musical facilitation of memory would be more related to the gist of the lyric (via the easy change condition with the entire sentence changed to a different semantic message).

2. Materials and methods

2.1. Participants

Thirty-two adults (20 females) were recruited from the Institute's online community participation portal, which recruits from the greater Nijmegen, Netherlands area. Participants were 23.13 years old on average (range 19–30 years of age) and were all right-handed. All participants were advanced or fluent in English (English language bilingual: N = 17; reported advanced or fluent, but not bilingual: N = 2, English language of instruction at University: N = 5; reported learning English before age 12: N=3; Native English speakers: N = 5). On average, participants learned English at age 6.72 (range: 3–12 years of age) and their first languages were Dutch (N=7), Italian (N= 3), Spanish (N=2), Greek (N=2), and one speaker of each of the following languages: Bosnian, Bengali, Shona, Hindi, Russian, Indonesian, Portuguese, Slovenian, Polish, German, and Latvian. Seventeen individuals self-identified as bilingual. Overall participants had an average of 9.94 years of musical training (range: 3–22 years) beginning at age 9.94 years on average (1–21 years). Based on a musicianship criterion of 5 or more years of musical training that began before age 10, there were 13 musicians and 19 non-musicians. Six of the musicians reported playing piano, 7 guitar, 2 each of voice, trumpet, and drums, 1 each of recorder, clarinet, and accordion. No participants reported any hearing impairments or neurological disorders, and all reported normal or corrected-to-normal vision. All study materials were approved by the local ethics committee (CMO, Committee on Research Involving Human Participants in the Arnhem-Nijmegen region) and followed the guidelines of the Helsinki declaration. All participants provided informed consent to participate in the study and received monetary compensation for their participation.

2.2. Materials

Stimuli consisted of 2 sets of matched spoken and sung stimuli. The first set of stimuli was a total of 48 utterances consisting of 24 English texts (Harvard Sentences, IEEE Subcommittee, 1969; see Appendix) that were spoken and sung. The second set of stimuli had the same melodies and similar sentence prosody as Set 1, but had alternate lyrics. This allowed for multiple presentations of familiar and unfamiliar melodies with different text settings. Set 1 was obtained from previous studies (including Vanden Bosch der Nederlanden et al., 2020). To create alternate texts, we recruited two male speakers (Canadian and British English) and recorded their speaking and singing to a new set of Harvard Utterances with the same number of syllables as the original stimulus set.[1] All stimuli were created to be similar in average pitch (F0) and duration (see Table 1). Despite this matching, sung utterances were statistically longer than spoken in total duration and average syllable length, however this difference was on the order of milliseconds (70 ms for total utterance duration and 15 ms for syllable duration). There were no differences between speech and song for overall variability of pairwise syllable durations and no differences for the average time to peak amplitude (rise time) of spoken and sung syllables. Sung and spoken utterances were not different in F0 (perceived pitch), but the F0 of each syllable was more variable for speech than song (F0 instability), and the harmonicity (ratio of periodic information to noise in the signal) was greater for song than speech, both differences likely due to the held notes for the vowel portion of the sung syllables. Song and speech both had peaks in the frequency spectrum in the delta (1-4 Hz) and theta (4–8 Hz) bands, corresponding to phrasal and syllable durations of speech (Fig. 1), respectively.

Training stimuli were piano melodies created from the sung utterances using MuseScore (https://musescore.org) but were first converted to MIDI using Melodyne 5. In MuseScore, melodies were represented using grand piano instrument and were manually edited for musical dynamics (e.g., loud/soft/accent) and missed notes to reflect the original sung utterance. These piano melodies were used as the training melodies for the 4 days of melody training before participants came into the lab. Post-test survey materials were developed to test participants' memory for lyrics uttered to familiar and unfamiliar melodies. There were 8 "easy change" trials and 16 "hard change" trials. If the correct lyric was "Glue the sheet to the dark blue background" then a hard change trial could be "Glue the sheet to the *light* blue background" which would assess verbatim lyric recall, while the easy change lyric would be "Many hands help get the job done" which would assess gist recall of the paired melody and lyric.

2.3. Procedure

Participants completed four consecutive days of online melody training (Qualtrics) on 12 of the 24 melodies prior to the day of the MEG

[1] Note that for one sentence, the alternate lyric had 10 syllables instead of 12 as the original did. Alt lyric: "A round hole was drilled through the **thin thin** board" vs. Orig lyric: "They are men who walk in the **middle of the** road". These stimuli were kept in the current analyses since the stimulus spectrum of the sentences did not significantly differ and the contours were so closely matched.

Table 1
Acoustic characteristics for spoken and sung utterances.

Acoustic Features	Song M (SD)	Song Range	Speech M (SD)	Speech Range	*p*
Total Duration (ms)	2,497 (368)	1678–3855	2,427 (330)	1616–3403	**.005**
Syllable Duration (ms)	273 (43)	210–386	261 (40)	188–356	**< .001**
Onset-to-Onset Variability (nPVI; a.u.)	71 (15)	41–107	67 (19)	24–112	.051
Syllable Onset Duration (ms)	95 (19)	5–170	95 (14)	7–132	.862
F0 (Hz)	138.3 (11)	104.4–157.1	138.4 (20)	107.3–184.5	.939
F0 Instability (St)	.7 (.1)	.3–1.0	1.4 (.5)	.7–2.6	**< .001**
Harmonicity (dB)	14.2 (2.4)	10.0–21.3	10.4 (2.1)	7.1–16.3	**< .001**

Abbreviations: M = mean, SD = standard deviation, ms = milliseconds, Hz = Hertz, St = semitones, dB = decibels, a.u. = arbitrary units.

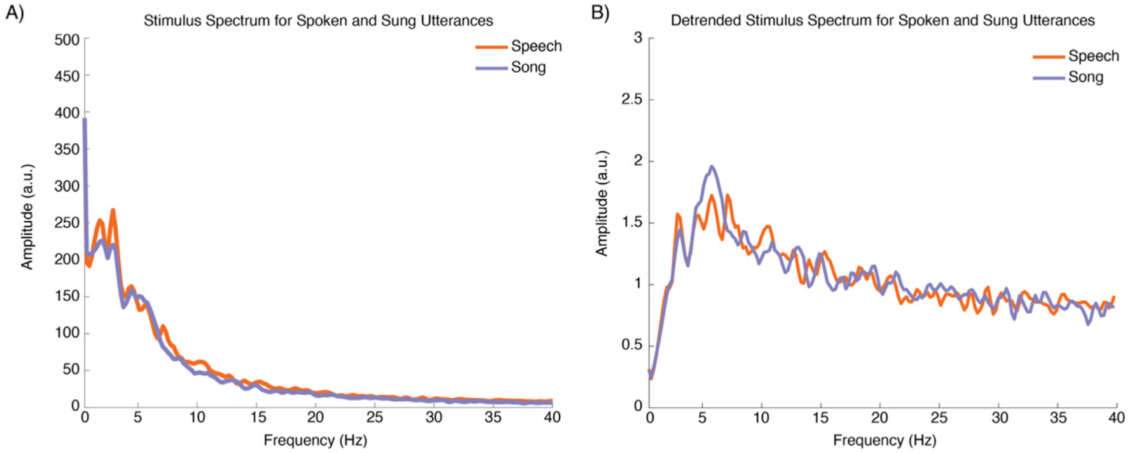

Fig. 1. Average acoustic (A) amplitude spectrum for spoken and sung stimuli and (B) log linear detrend performed separately for spoken and sung utterances to take out 1/f noise.

session. Training occurred on the 4 days immediately before the scheduled testing. Four days was thought to be sufficient given that participants show some effect of familiarity during one listening session (Wallace, 1994), but training on multiple days before testing shows greater effects of familiarity (Moussard et al., 2012; Tamminen et al., 2017). Participants heard one of four pseudo-random training stimulus Orders to allow all 24 utterances to be trained and untrained across participants, each day participants were trained on the same 12 melodies. This training (four exposures of four presentations per session with as many repeated presentations of each stimulus as the participant desired over 4 days) familiarized participants with half of the melodies that would be presented with texts during the experimental testing session. For each training session, participants heard 4 presentations of each melody and had to count the number of notes in each melody. They were quizzed about the number of notes in the melody after listening to each stimulus ("How many notes were in that melody?") using a multiple-choice selection (six choices of seven to twelve notes in total). As motivation for completing training, participants were simply told that they had to learn 12 melodies for the study and if they did not complete each day of training, they would not be able to complete the MEG portion of the study (reducing potential compensation). Participation in each of the four surveys was required before the time of in lab testing.

In the lab MEG session, participants changed into the provided sweatpants and sweatshirt/t-shirt and were instructed to remove all metal. Participants were fitted with a coil at the nasion and were instructed to sit in the MEG chair to be fitted with insert earphones containing coils in each earpiece. Participants were given pillows and blankets to make them comfortable and to minimize movement during testing. One response box was placed under the participants' right hand and they were told to rate utterances after they heard a beep (a 400 ms, 1000 Hz tone with 10 ms onset and offset ramps). according to how familiar it was to them from their melody training ("Did you recognize that last melody from your training? 1 = 'I did not learn that melody' through 4 = 'I learned that melody'."

All 96 utterances ((24 [melodies] x 2 [text settings]) x 2 [spoken versions]) were presented to participants four times during the testing session (384 trials). Half (24 of 48) sung utterances' melodies were familiar (from training) and half were unfamiliar. None of the 48 spoken utterances were familiar, but, since all the sung texts had a matched spoken counterpart, these matching spoken counterparts were labeled as familiar and unfamiliar based on the familiarity of the matching sung utterance. This yoking also allowed us to examine the effect of familiarity on neural tracking while phonological features of these matched sung and spoken utterances were identical. Participants provided familiarity ratings on 25% of the trials; on the other 75% of the trials there was an 800–1300 ms jittered silent interval and then the next sentence was presented. Across 4 blocks of 96 trials, all 96 spoken and sung utterances were rated once by each participant. After participants completed the MEG study, they completed a demographic questionnaire to self-report language, musical, and neurological background. One week after completion of the study, participants completed an online post-test questionnaire (Qualtrics) about their memory for lyrics that were uttered to familiar and unfamiliar melodies with hard and easy lyric change types.

MEG recording, preprocessing, and coherence analyses

MEG was recorded with a 275 axial gradiometer system (CTF), analog low-pass filtered at 300 Hz and digitized at a sampling frequency of 1200 Hz. Three coils on the nasion, and the left and right ear canals were used to register the participants' head to the MEG-sensor array. Their head position was continuously monitored through the entire experiment using custom software (Stolk et al., 2013) and could be repositioned to the starting position using this software. Three Ag/AgCl electrode pairs were used to measure horizontal and vertical eye movements and heartbeat.

Offline analyses were carried out using a custom Matlab (2018b) script developed with FieldTrip (version 20190402; Oostenveld et al., 2011). Participants' data was first epoched into 4.5 s epochs aligned to the onset of the stimulus, which included a 500 ms pre-stimulus-onset baseline. Epochs were shortened when necessary to make sure that epochs were not overlapping with other stimulus presentations. Epochs were high pass filtered at 1 Hz and low pass filtered at 40 Hz using a forward and backward window sinc Finite Impulse Response (FIR) filter implemented with Matlab's fftfilt function. Epochs were also baseline corrected using the 500 ms pre-stimulus period and downsampled to 200 Hz. Epochs were manually inspected using the variance summary visual inspection function in FieldTrip for all trials and channels. Outlier channels and trials with significantly greater variance compared to the rest of the trials for a given participant were removed from analyses (dropping an average of .81 channels, range 0–7 channels, and .59 trials, range 0–4 across all participants). The epochs were submitted to independent component analysis (fastica algorithm), and components corresponding to vertical/horizontal eye movements or heartbeat were removed after they were confirmed by both spatial topographies and their time courses.

Cleaned data were converted to synthetic planar gradients using the 'sincos' method of the *ft_megplanar* function and recombined using the 'svd' method of *ft_combineplanar*. This latter step combines the vertical and horizontal components of the channels based on a singular value decomposition (svd), which rotates the components in a way that maximizes the variance of the signal along the first singular vector. The first singular vector was retained for further analysis. We ensured that stimulus onset times were accurately defined by estimating any delay between the audio trigger sent by the stimulus presentation script and the actual onset of the stimulus to the participant. This delay varied from trial to trial and was caused by the configuration of the stimulus presentation hardware. The delay was estimated for each trial by estimating the slope of the phase difference spectrum between the stimulus audio signal, and its recorded version on one of the analog channels in the MEG data. This correction resulted in a time delay correction for each trial of 10 ms on average. Acoustic envelopes were obtained from the stimulus wav-files (44100 Hz sampling rate) and were then resampled to the timing of the down-sampled MEG data. Zero-padded (5 s) epochs of MEG and acoustic envelopes were converted to the frequency domain using a multi-tapered (1 Hz smoothing parameter) Fast Fourier Transformation, resulting in single-trial power and cross-spectral density for all MEG-envelope channel pairs.

Cerebro-acoustic phase coherence (referred to hereafter as coherence) was estimated using the *ft_connectivityanalysis* function in FieldTrip, to calculate the consistency of the phase alignment between each MEG channel and the amplitude envelope of the stimulus across all trials. This measure indexes of the consistency with which the phase of neural oscillations was aligned to the stimulus, to track the syllable information for spoken and sung utterances. Coherence was computed separately for each condition. The 10 sensors (5 left and 5 right) with the most coherence in the theta band (4-8 Hz) across all conditions were selected for further analyses. The coherence bias was estimated empirically for each participant by randomly shuffling the auditory envelopes across epochs, and re-calculating coherence in 100 permutations. Coherence data for the 10 selected sensors were averaged together and then z-score transformed using the mean and standard deviation from the 100 random MEG-audio pairings for the ten selected sensors. Z-score transformations were calculated for each condition using the condition-specific mean and standard deviation from the random pairing dataset and with the same number of trials as the true MEG-audio pairing dataset.

3. Analyses

3.1. Behavioural data

The proportion of correct responses for each participant was submitted to a 4 Training Day (Days 1-4) x 4 Order (1-4) repeated measures Analysis of Variance (ANOVA). We did not intend to include musicianship as a factor, but since our participants had significant musical training, we added this as a between-subjects factor as additional exploratory analyses. These results are reported using the Huynh-Feldt correction for Sphericity and post hoc tests are reported using the Bonferroni correction for multiple comparisons. Although participants had to listen to the whole sound file before advancing, some participants did not give a response for every training trial, resulting in a null response on 1-2 trials for seven participants across all four training sessions. Averages were always reported from the total number of trials with responses. Participant's familiarity ratings during the MEG session were submitted to a 2 Utterance (Speech, Song) x 2 Familiarity (familiar, unfamiliar) repeated-measures ANOVA. Post-test proportion of correct behavioural responses was submitted to a 2 Melody Familiarity (unfamiliar, familiar) x 2 Difficulty (hard, easy) x 4 Order (1-4) repeated measures ANOVA. An experimental coding error resulted in the loss of 1 familiar hard question for orders 1 and 2 and 1 familiar hard question for order 3. Averages were tallied based on the total number of valid trials per participant. Skewness and kurtosis were within normal ranges (+/- 3).

3.2. MEG data

Statistical comparisons were made for the theta band (4-8 Hz) based on previous work with entrainment to spoken utterances (Vanden Bosch der Nederlanden et al., 2020) and based on the acoustic analyses of the stimuli presented in our study (see Stimulus section; but see supplemental data for the delta band, 1-4 Hz, results). Two sets of analyses were performed based on A) the *assigned* familiar and unfamiliar melodies for each participant (i.e., the melodies participants received training on) and on B) participants' *perceived* familiarity based on self-reported ratings of familiarity during the MEG session. We dichotomized participant's familiarity ratings into unfamiliar (rating of 1 or 2) and familiar (rating of 3 or 4) groupings for sung utterances. Spoken utterances were categorized according to the familiarity rating given to the matching sung utterance. We did not use speech familiarity ratings because we wanted to compare the matched texts for song and speech directly and because very few spoken utterances were rated as "familiar" as they were not heard during the melody training. The differences in perceived familiarity for sung *and spoken* utterances would have led to a highly imbalanced number of trials for each cell of the study design. Coherence was submitted to a 2 Utterance (speech vs. song) x 2 Familiarity (familiar, unfamiliar) x 2 Hemisphere (left, right) repeated measures ANOVA. We did not intend to include musicianship as a factor, but since our participants had significant musical training, we added this as a between-subjects factor as additional exploratory analyses. Skewness and kurtosis were within normal range for assigned familiarity (i.e., +/- 3), but one outlier for unfamiliar speech in the left sensors affected the normality in the perceived familiarity grouping. The presence or absence of this single cell did not change the outcomes of the perceived familiarity analysis. The same participant was removed from exploratory by block analyses based on perceived familiarity. After outlier removal, skewness and kurtosis were within normal ranges (i.e., +/- 3). Figures and stats include all participants, but all MEG stats without the outlier are provided in the supplement for completeness.

4. Results

4.1. Behavioural

Accuracy is plotted in Fig. 2 for the pre- and post-MEG behavioural testing. For melody training, participants overall performed well (chance performance was 16%) and improved during the four days of training, $F(3,72) = 7.374$, $p < .001$, $\eta^2 = .173$. As illustrated in Fig. 2A, participants had greater accuracy on Day 4 than Day 1 ($p < .001$), marginally greater for Day 2 ($p = .070$), but not Day 3 ($p = 1.000$; Bonferroni corrections). However, this effect interacted with

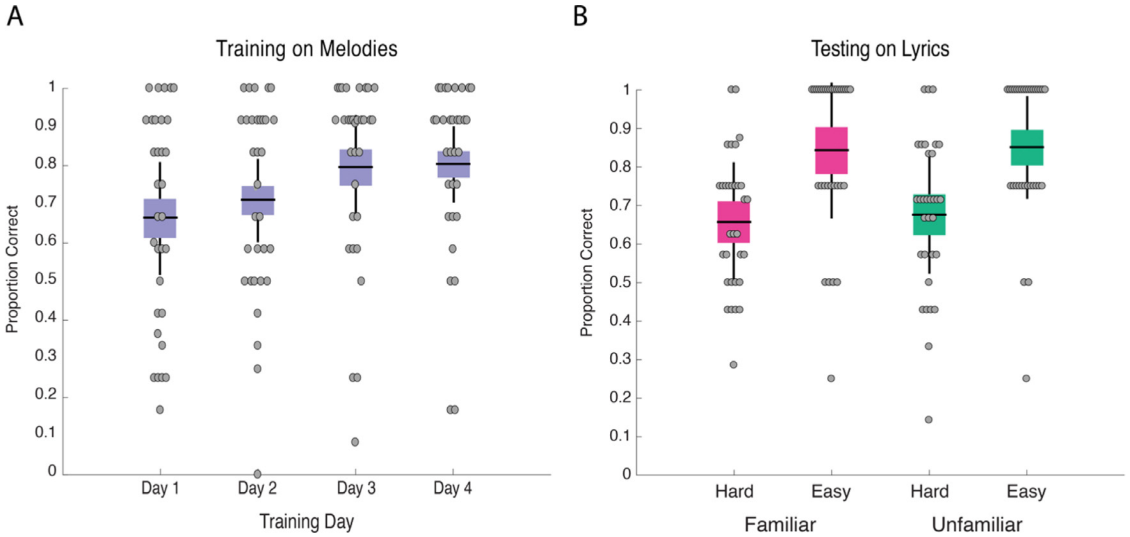

Fig. 2. Melody training performance (A), illustrates that participants improved over the 4 days, suggesting they learned the assigned melodies. Post-test results (B), suggest participants were better at identifying the easy whole lyric substitutions than the difficult specific detail lyric substitutions, for both familiar and unfamiliar lyrics. Means are displayed with shaded areas indicating 1.0 standard deviation and black lines represent 95% confidence intervals.

Table 2
Mean and standard error of familiarity ratings and proportion of trials recoded as familiar per condition.

	Rating M (SE)	Proportion Recoded as Familiar
Familiar Song	3.25 (.08)	.789 (.025)
Familiar Speech	2.30 (.10)	Yoked (same as song)
Unfamiliar Song	2.23 (.08)	.398 (.030)
Unfamiliar Speech	1.9 (.09)	Yoked (same as song)

order, F(1,72) = 3.230, p = .006, η^2 = .228, despite no main effect of order, F(3,24) = 1.228, p = .321, η^2 = .101. Only participants in Order 1 improved significantly in their melody note counting (see Fig. S1). Although training numerically increased across training days for all other orders, it did not reach significance, p = .07. Post-hoc analyses reveal that order is only a significant predictor on day 1 performance, p = .025 (all other p's > .4). Either Order 1 was overall harder and yet participants were able to reach similar accuracy by day 2 of training, or participants randomly assigned to Order 1 had poorer performance on day 1. In general, musicians performed better than non-musicians (Mus: 85.5%; Non-mus: 66.9%), F(1,24) = 5.075, p = .034, η^2 = .140, but that did not interact with order (p=.338). Taken together, training performance indicated that participants did well in their note counting during training (74.5% accuracy).

Average ratings during MEG testing for familiar utterances suggested that participants learned the melodies and recognized them during the MEG testing session (see Table 2). Ratings were significantly higher for song than speech, F(1,30) = 43.984, p <.001, η^2 = .592, and greater for familiar than unfamiliar utterances, F(1,30) = 84.368, p <.001, η^2 = .719. The interaction between familiarity and utterance was significant, F(1,30) = 68.301, p <.001, η^2 = .657, with higher average familiarity ratings for familiar song than familiar speech (p<.001, d=1.608), but less so for unfamiliar spoken and sung utterances (p=.002, d = 0.592). A χ^2 analysis at the stimulus item level showed no significant differences across stimuli based on the proportion of participants rating an item as familiar (dichotomized ratings), χ^2(48, N=517) = 528.0, p = .359, or based on average stimulus rating across participants, χ^2(48, N=940) = 960.0, p = .318. This suggests that certain stimuli were not rated as familiar across all participants, regardless of training. Musicianship interacted with familiarity and utterance ratings during the MEG session but did not interact with utterance alone (p =.607) or with familiarity alone (p = .097). This three-way interaction, F(1,30) = 5.690, p = .024, η^2 = .055, was driven by an interaction that was present in sung utterances but not spoken utterances (familiarity x musicianship for speech alone: p = .559; song alone: p = .034). Non-musicians had greater familiarity ratings for unfamiliar songs, p = .027, d = .837, but not unfamiliar speech, p = .779, d = .102. Musicianship did not alter ratings overall (p = .633), but non-musicians heard unfamiliar songs as more familiar than musicians (see Fig. S2). Together these findings suggest that all participants learned the melodies they were trained to learn during the 4 days preceding MEG testing, but that non-musicians either did not learn as well as musicians or relied more on idiosyncratic or stimulus specific features to provide familiarity ratings.

We also compared how training performance was related to MEG session familiarity ratings by correlating training improvement (day4 minus day1 accuracy) with a familiarity difference score for sung utterances (familiar song minus unfamiliar song ratings). Neither training improvement, r = -0.214, p = .240, nor day 4 accuracy, r = 0.060, p = .744, correlated with familiarity difference scores for song. Thus, while our training paradigm was successful in having participants learn melodies, it did not predict the difference in familiarity ratings for familiar versus unfamiliar sung utterances.

As illustrated in Fig. 2B, participants were more accurate with the easy lyric memory questions than the difficult memory questions, *F*(1,24) = 50.311, *p* < .001, η^2 = .582, but they were similarly accurate for familiar and unfamiliar post-test questions, *F*(1,24) = .235, *p* = .632, η^2 = .008. Musicians (82.6%) had greater accuracy than non-musicians (71.9%; p =.005). No other interactions or main effects were significant. Participants showed no memory advantage for texts sung to familiar melodies (in pink) compared to unfamiliar melodies (in green) in the current study. However, these texts were presented to participants in both a sung and a spoken format four times during the study, which may have made all texts familiar by the end of the MEG session.

4.2. *MEG cerebro-acoustic phase coherence*

Coherence for the top ten sensors per participant in the theta band showed a robust effect of utterance type (Fig. 3a), with greater coherence when participants listened to sung compared to spoken versions of the same lyrics. The significance of this effect did not change between the assigned versus perceived familiarity groupings described below. Similarly, a significant effect of hemisphere (see Fig. 3b), suggested

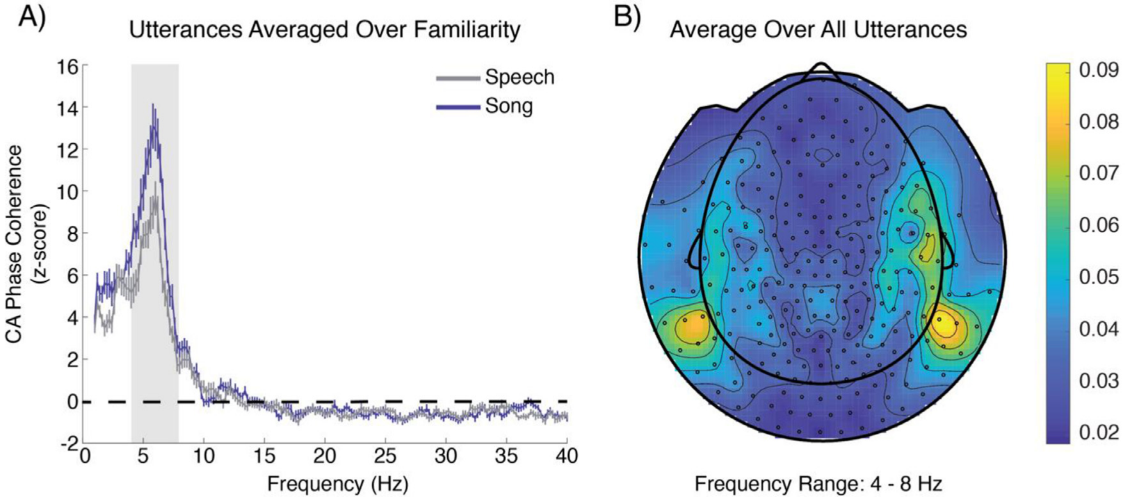

Fig. 3. Coherence (A) and topographies of coherence (B) showed canonical responses to utterances across familiarity type. Panel A demonstrates greater coherence when the same sentence is sung compared to spoken in the theta band (4-8 Hz, grey shading). Panel B illustrates raw (non-z-score transformed) coherence in the theta band, with greater coherence over right than left sensors.

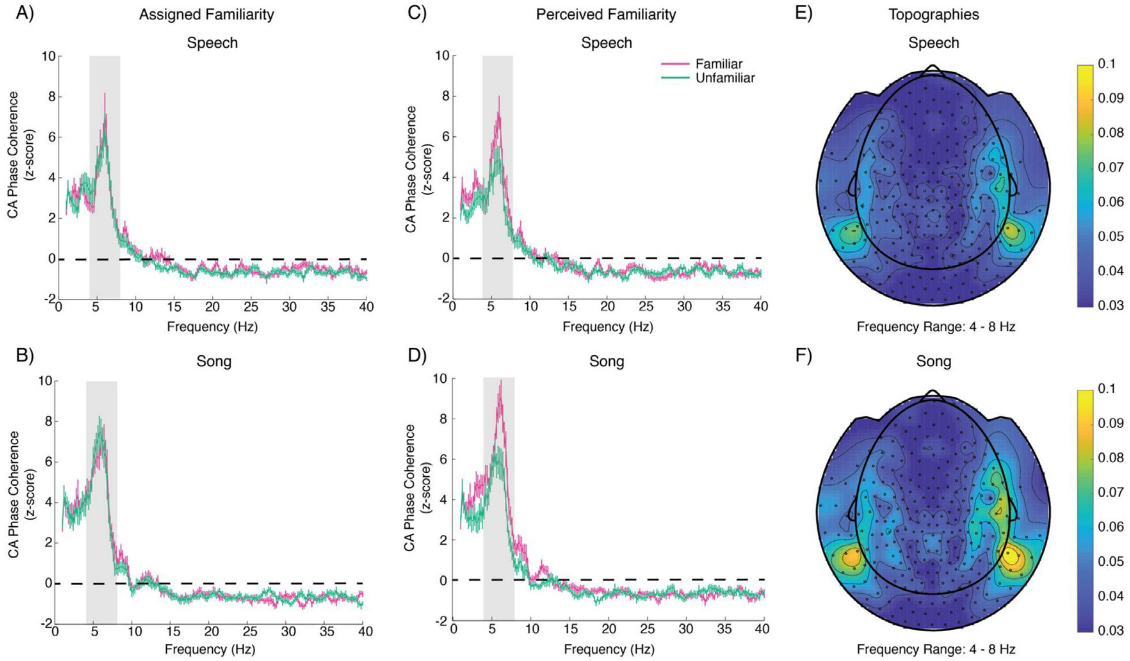

Fig. 4. Phase coherence plots for assigned familiarity (A & B) and perceived familiarity (C & D), illustrating significant coherence in the 4-8 Hz bands corresponding to syllable durations. Panels E and F show auditory generators of phase coherence (raw coherence, not z-scored) for both speech and song, with greater coherence over the right than left hemispheres.

greater coherence overall in the right than left hemispheres, regardless of the utterance type and familiarity grouping (but see below for marginal interactions between utterance and hemisphere for both assigned and perceived familiarity groupings). This same pattern was also observed in the delta band (see Supplement).

In the *assigned* familiarity analyses (Fig. 4a and b), coherence was greater for sung compared to spoken utterances, $F(1, 31) = 33.303$, $p < .001$, $\eta^2 = .518$, as illustrated above (Fig. 3a). There was no effect of familiarity, $p = .995$ and utterance type did not significantly interact with familiarity $p = .929$, as illustrated in the spoken (Fig. 4a) and sung (Fig. 4b) utterances. As previously discussed, there was greater coherence over the right sensors than the left (see Fig. 3b), $F(1, 31) = 8.654$, $p = .006$, $\eta^2 = .218$. Utterance type appeared to interact with hemisphere (greater right lateralisation for song than speech), but the interaction did not reach statistical significance, $F(1, 31) = 3.889$, $p < .058$, $\eta^2 = .111$, (see Fig. 4e and f). Familiarity significantly interacted with hemisphere, $F(1, 31) = 4.461$, $p < .043$, $\eta^2 = .126$, with greater activation for the right than left hemisphere in the unfamiliar condition (right: 6.13 (3.07), left: 4.70 (2.68); p = .005), and no significant difference in the familiar condition (right: 5.62 (2.80), left: 5.12 (2.37); p = .270).

In contrast, when coherence was grouped according to participants' *perceived* familiarity as indexed by their ratings for each sentence during the MEG session (Fig. 4c & d), there was greater neural tracking for familiar compared to unfamiliar utterances. Statistically, there was a main effect of familiarity, $F(1, 31) = 10.818$, $p = .003$, $\eta^2 = .259$, but no significant interaction with spoken or sung utterance type, $F(1, 31) = 0.189$, $p = .667$, $\eta^2 = .006$, which ran counter to the predicted results. As observed in the assigned familiarity analyses, the

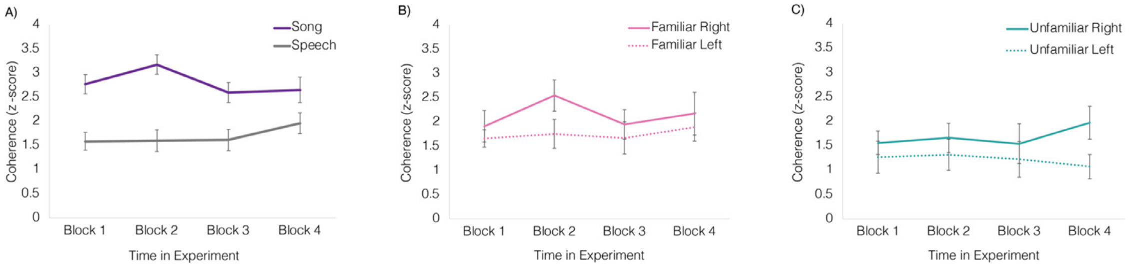

Fig. 5. Average coherence in the theta band (4-8 Hz) illustrating block interactions. (A) illustrates greater coherence for song than speech, an effect size that increased from block 1 to block 2, but decreased in blocks 3 and 4. The second two panels illustrate (B) Familiar and (C) Unfamiliar trials, showing the 3-way interaction between hemisphere, familiarity, and block, with significantly greater coherence for familiar utterances in the right hemisphere compared to the left in block 2, with the same effect for unfamiliar melodies, only it was evident later in block 4.

perceived familiarity grouping also exhibited greater coherence for sung utterances than spoken utterances, $F(1, 31) = 59.225$, $p < .001$, $\eta^2 = .656$, and more coherence over right than left hemisphere sensors, $F(1, 31) = 6.927$, $p = .013$, $\eta^2 = .183$. Again, there was a marginal interaction between utterance and hemisphere, which suggested a trend toward greater coherence in the right than left hemispheres for song but not speech, $F(1, 31) = 3.342$, $p = .077$, $\eta^2 = .097$ (see Fig. 4e and f). There was no interaction between familiarity and hemisphere ($p = .871$). The same pattern of no effect of familiarity for assigned familiarity, but an effect for perceived familiarity groupings was also evident in the delta band, as well as greater neural tracking for right than left sensors overall (see Supplement).

4.3. Exploratory analyses

To further investigate the strong effect of familiarity for both sung and spoken materials—a surprising finding since training only occurred for melodies—we separated trials into the four testing blocks that were presented sequentially to participants during the MEG portion of the study. Each block presented a complete set of all 96 stimuli. We added block (4 levels) to the perceived familiarity analysis described above and found the same main effects as above, with the addition of an interaction, as illustrated in Fig. 5A, between block and utterance, $F(3, 93) = 3.338$, $p = .023$, $\eta^2 = .097$. Initially the difference in coherence for song and speech grew larger, but then began to shrink with increasing time in the experiment (Effect sizes [song>speech]: Block 1, d = .879; Block 2, d = 1.080; Block 3, d = .851; Block 4, d = .482). This pattern suggests an effect of melody familiarity at the beginning of the study, followed by an effect of learning spoken materials during the study. A marginal interaction between hemisphere and utterance, $F(1,31) = 3.910$, $p = .057$, $\eta^2 = .112$, suggests greater right than left hemisphere coherence for song, but not speech. Similarly, a marginal 3-way interaction between block, familiarity, and hemisphere, $F(3, 93) = 2.636$, $p = .054$, $\eta^2 = .078$, lends to an initial effect of familiar utterances and a later effect for learning unfamiliar utterances. Specifically, there was greater coherence for the right than the left hemisphere in the *second* block for *familiar* utterances (Fig. 5B), but during the *fourth* block for *unfamiliar* utterances (Fig. 5C), while no other blocks had significant differences between right and left hemispheres. These patterns suggest an effect of learning for the unfamiliar utterances that was especially evident in the right hemisphere toward the end of the listening session.

Although we did not recruit participants based on their musical training, several participants had a significant amount of musical training (see Participants section above). Our task involved learning melodies and, as this ability could be significantly affected by musical training, we examined how musical training affected performance. As described above, musicians had greater performance learning melodies overall and had a larger difference in familiarity ratings for familiar compared to unfamiliar songs. When musicianship was included in our perceived familiarity analyses, there was no main effect of musicianship ($p = .266$) and musicianship did not interact with any other variables (utterance type, familiarity, or hemisphere: all ps > .169).

5. Discussion

The current study examined whether neural tracking of spoken and sung utterances would be affected by melody familiarity. In particular, music may benefit speech processing through improved memory for words set to music compared to speech. However, the consistency of this effect is mixed, with the best memory being observed for sung utterances set to a familiar melody. We hypothesized that familiar melodies may lead to improved neural tracking of the stimulus, providing a mechanism that improves word encoding. To investigate whether familiarity indeed modulated neural tracking of speech and song, we familiarized participants with novel melodies and tested their neural tracking of words sung with familiar and unfamiliar melodies. Despite clear evidence that participants learned the melodies *assigned* in the study, only participants' *perception* of familiarity affected neural tracking of both sung and spoken utterances. Participants were better at tracking syllable- and word-level information for utterances they perceived as familiar compared to unfamiliar, providing the first evidence of the effect of familiarity on neural tracking in theta and delta bands. These results are consistent with additional studies highlighting the importance of top-down factors such as attention (Obleser and Kayser, 2019; Song and Iverson, 2018) and language background (Zou et al., 2019; Reetzke et al., 2021) on neural tracking.

In contrast to previous literature on familiarity, we found that only perceived familiarity, and not assigned familiarity based on the training sessions, related to differences in neural tracking. Although a significant proportion of trained melodies were still part of the recoded familiar groupings based on participant's ratings, many unfamiliar melodies were recoded as sounding familiar. Our melodies were composed by mapping sentence prosody onto the Western diatonic musical scale. This approach, as well as the short duration of utterances, gave the melodies a simple melodic contour that may have elicited some sense of familiarity regardless of training. Some of the reported mnemonic effects of music on speech were most evident when the musical structure or the text setting for an unfamiliar melody was simplified (Gingold and Abravanel, 1987; Wallace and Rubin, 1991; Wallace, 1994; Good et al., 2015), which leaves open the possibility that melody-specific characteristics could also play a role in memory encoding. Therefore, even unfamiliar melodies may have felt somewhat familiar to each individual depending on the simplicity of the melodic contour or the individual's degree of Western musical enculturation. Participants also reported enjoying how sentences fit within our melodies, suggesting that other factors like enjoyment could lead a participant to feel familiar with an unfamiliar melody (e.g., van den Bosch et al., 2013). Finally, although our training resulted in high performance, it is possible that participants did not listen to utterances enough to engender as strong a sense of familiarity as in previous studies that used long-known melodies, such as Beethoven's Ode to Joy (Moussard et al., 2012; Wallace, 1994). Longer melody training sessions with a different method for training than note counting could foster better learning of novel melodies (e.g., melody testing as in Tamminen et al. 2017). Learning a melody may even be more efficient when learned with than without words (cf Weiss et al. 2012., for a vocal encoding benefit for melodies), suggesting a mnemonic effect of words on melody retention.

Contrary to our predictions of an isolated effect of familiarity on sung words, our familiarity effect extended to spoken utterances that were matched to the familiar sung melodies. Exploratory analyses suggested that there was a more robust effect of familiarity during the first and second blocks of the study, with an increase in coherence for song over speech in Block 2 accompanied by a marginal increase in coherence in the right hemisphere for familiar utterances. After the initial effect of familiarity, our listeners began to learn the texts of the unfamiliar songs and spoken utterances over the course of the study, as evidenced by the increase in coherence for spoken utterances and marginally for unfamiliar utterances over the right hemisphere during the final block of the study. There was greater coherence over the right than left hemisphere overall—consistent with the literature surrounding right lateralized processing of pitch (e.g., Zatorre et al., 1994)—which fits well with our finding of neural tracking dynamics were most evident over the right hemisphere. Together these exploratory interactions by block suggest a peak in benefit from familiar melodies by the second block over the right hemisphere, followed by a peak in learning through repetition of unfamiliar melodies by the end of the study, again over the right hemisphere. As the spoken utterances had the same texts as their matched sung counterparts, these findings could also be evidence that, once the words were well-encoded via the melodic text setting, the spoken versions of those same words were also neurally tracked better. The current study was focused on controlling for acoustic differences between speech and song, but future work could more carefully test whether a memory benefit could be elicited by testing participants on different lyrics for sung and spoken materials so that there can be no transfer of learning from lyrics put to song to their identical spoken counterparts.

Right-lateralized coherence for both spoken and sung utterances is surprising given the wealth of evidence for left lateralized responses to speech. Here, each participant's task during the MEG session was to determine whether they had learned the melody of the utterance during their training session. This may have biased their attention toward the prosodic contour of spoken utterances, which is also pitch-based and is right lateralized in speech processing (Friederici and Alter, 2004; Meyer et al., 2002; Zhang et al., 2010). It may not be entirely surprising that we did not see left lateralized responses for speech, given previous findings of bilateral neural tracking of speech in neural tracking (Vanden Bosch der Nederlanden et al., 2018) and fMRI (e.g., Rogalsky et al., 2011). Finally, our findings are specifically based in tracking the slow rhythms of speech in the theta band. In several theories of auditory hemispheric lateralisation , events at 4-8 Hz ought to be preferentially processed in the right hemisphere (Giraud et al., 2007). Therefore, the right-lateralized processing may be more indicative of the right-lateralized bias for slow rhythms (e.g., Albouy et al., 2020).

The learning effects for unfamiliar and spoken stimuli in the current study may be due primarily to the number of presentations or repetitions of the stimulus during the study. After all, we used repetition as a method for increasing the familiarity of our trained melodies. Perhaps repetition alone, which is a feature of music and not of language (Margulis et al., 2012), may bring about increased neural tracking regardless of whether the stimulus was music or speech (Vanden Bosch der Nederlanden 2015a, 2015b). Future work should characterize the time-course of this neural tracking enhancement for familiar and unfamiliar utterances, to understand whether music is related to faster boosts in neural tracking with repetition than speech. Similarly, song may be more robust to repetition suppression because repetition is a feature of song and not speech.

Our study was motivated by understanding the memory benefit of sung over spoken words, but we found no selective enhancement for familiar song in the neural tracking of our stimuli or behavioural recall one week after the MEG session. This null effect should not be taken as evidence for a lack of musical mnemonic benefit. First, the same texts were used for spoken and sung utterances, making it difficult to tease apart the effects of familiarity on word learning and recall. Second, learning likely occurred over the course of the study for unfamiliar and spoken utterances, so a post-test delayed by one week may be indexing learning from the testing session. Finally, some studies find inconsistent connections between increased neural tracking and better speech comprehension (Reetzke, et al., 2021; Zou et al., 2019). Therefore, in our study, the increased neural tracking for song than speech may not lead to better comprehension or memory for sung than spoken utterances. For instance, increased neural tracking could result from more attention to familiar than unfamiliar utterances or ease-of-processing for sung over spoken utterances that is not associated with downstream effects on behaviour. Future studies should more carefully assess the relationships between neural tracking and memory by using different sentences for spoken and sung utterances. This is not trivial, because different sentences for speech and song must be phonetically matched so that acoustic differences in the shape of the amplitude envelope due to different speech sounds do not drive neural tracking.

One of the strongest effects in the current study is that of greater coherence for song compared to speech. This effect is notable given that previous work found no difference in phase coherence to song and speech under normal listening conditions, which is the listening setting presented here, but only under difficult listening conditions (Vanden Bosch der Nederlanden et al., 2020). The current study differs from past work for several reasons. First, the current task required active participation in listening to each utterance, monitoring utterances for familiarity and providing ratings on 25% of trials. Greater neural tracking for song than speech may be related to that active task. However, if this were the case, then all sung utterances would have yielded greater coherence and not just those perceived as familiar. Second, the current study used a wide variety of spoken and sung utterances, whereas previous work relied on multiple presentations of very few sentences, which may have made those sentences over-learned and easier to suppress. Third, the current study utilized MEG instead of EEG, which may be more sensitive to auditory activity (Coffey et al., 2016). Finally, although our participants were fluent in English, they were not native speakers. Thus, the easy listening condition may have been more like a difficult listening condition, similar to the time-compressed difficult condition from previous work (Vanden Bosch der Nederlanden et al., 2020). Indeed, previous work has shown that non-native utterances elicit greater coherence than native speech, despite poorer comprehension for non-native than native utterances (Zou et al., 2019). If our participants perceived the English spoken and sung utterances as more difficult because they were non-native English speakers, this would likely have resulted in both spoken and sung utterances receiving greater coherence, and not an effect specific to song. More research is needed to fully characterize the effect of native language background, stimulus processing difficulty, and their effects on neural tracking. Taken together, our results suggest that adults are better at neurally tracking the same utterance when it is sung compared to spoken during active listening, even when utterances are spoken at a normal listening rate.

The current study highlights the importance of top-down factors such as familiarity on neural tracking of spoken and sung utterances. Moreover, our findings replicate and extend previous work examining how the syllable rhythms of song and speech are processed in the brain, with greater neural tracking for song than speech even in normal listening conditions. These results suggest that song as well as familiar utterances may both be effective at boosting neural tracking of the signal, with the potential for downstream benefits on comprehension.

Author note

The first author has moved to the University of Toronto – Mississauga (c.dernederlanden@utoronto.ca).

Data and code availability statement

All anonymised data, stimuli, and code are available upon request.

Funding

NSERC RGPIN-2016-05834 awarded to J.A.G, James S. McDonnell Foundation Understanding Human Cognition Scholar Award to J.A.G, Donders-BMI Exchange awarded to CMVBdN, and an NWO VIDI grant to J.M.S.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119049.

Credit authorship contribution statement

Christina M. Vanden Bosch der Nederlanden: Conceptualization, Funding acquisition, Methodology, Software, Validation, Formal analysis, Project administration, Writing – original draft, Writing – review & editing. **Marc F. Joanisse:** Conceptualization, Methodology, Resources, Writing – review & editing. **Jessica A. Grahn:** Conceptualization, Methodology, Resources, Writing – review & editing. **Tineke M. Snijders:** Conceptualization, Methodology, Supervision, Resources, Writing – review & editing. **Jan-Mathijs Schoffelen:** Conceptualization, Methodology, Software, Supervision, Resources, Visualization, Writing – review & editing.

References

Albouy, Phillippe, Benjamin, Lucas, Morillon, Benjamin, Zatorre J, Robert, 2020. Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. Science 367, 1043–1047.

Alisaari, J., Heikkola, L.M., 2017. Songs and poems in the language classroom: teachers' beliefs and practices. Teach. Teach. Educ. 63, 231–242. doi:10.1016/j.tate.2016.12.021.

Baird, A., Samson, S., 2009. Memory for music in Alzheimer's disease: unforgettable? Neuropsychol. Rev. 19 (1), 85–101. doi:10.1007/s11065-009-9085-2.

Bharucha, J.J., 1987. Music cognition and perceptual facilitation: a connectionist framework. Music Percept. Interdiscip. J. 5 (1), 1–30. doi:10.2307/40285384.

Bradley, I.L., 1971. Repetition as a factor in the development of musical preference. J. Res. Music Educ. 19, 295–298. doi:10.2307/3343764.

Calvert, S.L., Tart, M., 1993. Song versus verbal forms for very-long-term, long-term, and short-term verbatim recall. J. Appl. Dev. Psychol. 14 (2), 245–260. doi:10.1016/0193-3973(93)90035-T.

Chazin, S., Neuschatz, J.S., 1990. Using a mnemonic to aid in the recall of unfamiliar information. Percept. Mot. Skills 71 (3), 1067–1071. doi:10.2466/PMS.71.8.1067-1071, Pt 2.

Coffey, E.B.J., Herholz, S.C., Chepesiuk, A.M.P., Baillet, S., Zatorre, R.J., 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. Nat. Commun. 7. doi:10.1038/ncomms11070.

Cuddy, L., & Duffin, J. (2005). Music, memory, and Alzheimer's disease: is music recognition spared in dementia, and how can it be assessed?. Med. Hypotheses, 64, 229-235. 10.1016/j.mehy.2004.09.005

Deutsch, D., Henthorn, T., Lapidis, R., 2011. Illusory transformation from speech to song. J. Acoust. Soc. Am. 129 (4), 2245–2252. doi:10.1121/1.3562174.

Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. 109 (29), 11854–11859. doi:10.1073/pnas.1205381109.

Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85 (0 2), 761–768. doi:10.1016/j.neuroimage.2013.06.035, Pt 2.

Doelling B, Keith, Peoppel, David, 2015. Cortical entrainment to music and its modulation by expertise. Proceedings in the national academy of sciences E6233–E6242.

Engh, D., 2013. Why use music in english language learning? A survey of the literature. Engl. Lang. Teach. 6, 113–127. doi:10.5539/elt.v6n2p113.

Falk, S., Dalla Bella, S., 2016. It is better when expected: aligning speech and motor rhythms enhances verbal processing. Lang. Cognit. Neurosci. 31 (5), 699–708. doi:10.1080/23273798.2016.1144892.

Fiedler, L., Wostmann, M., Herbst, S.K., Obleser, J., 2019. Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. Neuroimage 186, 33–42. doi:10.1016/j.neuroimage.2018.10.057.

Finke, C., Esfahani-Bayerl, N., Ploner, C., 2012. Preservation of musical memory in an amnesic professional cellist. Curr. Biol. 22 (15), R591–R592. doi:10.1016/j.cub.2012.05.041.

Friederici D, Angela, Alter, Kai, 2004. Lateralization of auditory language functions: A dynamic dual pathway model. Brain and Language 89, 267–276.

Fuglsang, S.A., Dau, T., Hjortkjær, J., 2017. Noise-robust cortical tracking of attended speech in real-world acoustic scenes. Neuroimage. 156, 435–444. doi:10.1016/j.neuroimage.2017.04.026, Aug 1Epub 2017 Apr 13. PMID: 28412441.

Fung, C.V., 1996. Musicians' and nonmusicians' preferences for world musics: relation to musical characteristics and familiarity. J. Res. Music Educ. 44 (1), 60–83. doi:10.2307/3345414.

Gingold, H., Abravanel, E., 1987. Music as a mnemonic: the effects of good- and bad-music settings on verbatim recall of short passages by young children. Psychomusicol. J. Res. Music Cognit. 7 (1), 25–39. doi:10.1037/h0094188.

Giraud, Anne-Lise, Kleinschmidt, Andreas, Poeppel, David, Lund E, Torben, Frackowiak SJ, Richard, Laufs, Helmut, 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron 56, 1127–1134. doi:10.1016/j.neuron.2007.09.038.

Good, A., Russo, F., Sullivan, J., 2015. The efficacy of singing in foreign-language learning. Psychol. Music 43, 627–640. doi:10.1177/0305735614528833.

Gordon, R.L., Schön, D., Magne, C., Astésano, C., Besson, M., 2010. Words and melody are intertwined in perception of sung words: EEG and behavioral evidence. PLoS One 5. doi:10.1371/journal.pone.0009889.

Gordon, R.L., Magne, C.L., Large, E.W., 2011. EEG correlates of song prosody: a new look at the relationship between linguistic and musical rhythm. Front. Psychol. 2. doi:10.3389/fpsyg.2011.00352.

Haegens, S., 2020. Entrainment revisited: a commentary on Meyer, Sun, and Martin (2020). Lang. Cognit. Neurosci. 35 (9), 1119–1123. doi:10.1080/23273798.2020.1758335.

Jackendoff, R., 2008. Construction after construction and its theoretical challenges. Language 84 (1), 8–28. doi:10.1353/lan.2008.0058.

Jacobsen, J.H., Stelzer, J., Fritz, T.H., Chételat, G., La Joie, R., Turner, R., 2015. Why musical memory can be preserved in advanced Alzheimer's disease. Brain J. Neurol. 138 (Pt 8), 2438–2450. doi:10.1093/brain/awv135.

Joanisse J, Marc, Gati S, Joseph, 2003. Overlapping neural regions for processing rapid temporal cues in speech and nonspeech signals. NeuroImage 19, 64–79. doi:10.1016/S1053-8119(03)00046-6.

Johnsrude S, Ingrid, Zatorre J, Robert, Milner A, Brenda, Evans C, Alan, 1997. Left-hemisphere specialization for the processing of acoustic transients. NeuroReport 8, 1761–1765.

Kayser, S.J., Ince, R.A.A., Gross, J., Kayser, C., 2015. Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. J. Neurosci. 35 (44), 14691–14701. doi:10.1523/JNEUROSCI.2243-15.2015.

Kerlin, J.R., Shahin, A.J., Miller, L.M., 2010. Attentional gain control of ongoing cortical speech representations in a "cocktail party". J. Neurosci. 30 (2), 620–628. doi:10.1523/JNEUROSCI.3631-09.2010.

Kilgour, A.R., Jakobson, L.S., Cuddy, L.L., 2000. Music training and rate of presentation as mediators of text and song recall. Mem. Cognit. 28 (5), 700–710. doi:10.3758/BF03198404.

Krugman, H.E., 1943. Affective response to music as a function of familiarity. J. Abnorm. Soc. Psychol. 38 (3), 388–392. doi:10.1037/h0061528.

Kuroyanagi, J., Sato, S., Ho, M.J., Chiba, G., Six, J., Pfordresher, P., Tierney, A., Fujii, S., Savage, P., 2019. Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales. In: Proceedings of the 9th International Workshop on Folk Music Analysis (FMA), pp. 35–40. doi:10.31234/osf.io/zpv5w.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010. doi:10.1016/j.neuron.2007.06.004.

Madison, G., Schiölde, G., 2017. Repeated listening increases the liking for music regardless of its complexity: Implications for the appreciation and aesthetics of music. Front. Neurosci. 11. doi:10.3389/fnins.2017.00147.

Margulis Hellmuth, Elizabeth, 2012. Musical repetition detection across multiple exposures. Music Perception: An Interdisciplinary Journal 29, 377–385.

McElhinney, M., Annett, J.M., 1996. Pattern of efficacy of a musical mnemonic on recall of familiar words over several presentations. Percept. Mot. Skills 82 (2), 395–400. doi:10.2466/pms.1996.82.2.395.

Meyer, Martin, Alter, Kai, Friederici D, Angela, Lohmann, Gabriele, von Cramon Yves, D, 2002. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Human Brain Mapping 17, 73–88.

Meyer, L., Sun, Y., Martin, A.E., 2019. Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. Lang. Cognit. Neurosci. 35 (9), 1089–1099. doi:10.1080/23273798.2019.1693050.

Moore, E., Schaefer, R.S., Bastin, M.E., Roberts, N., Overy, K., 2017. Diffusion tensor MRI tractography reveals increased fractional anisotropy (FA) in arcuate fasciculus following music-cued motor training. Brain Cognit. 116, 40–46.

Morillon, Benjamin, Lehongre, Katia, Frackowiak SJ, Richard, Ducorps, Antoine, Kleinschmidt, Andreas, Poeppel, David, Giraud, Anne-Lise, 2010. Neurophysiological origin of human brain asymmetry for speech and language. Proceedings of the National Academy of Sciences 107 (43), 18688–18693.

Moussard, A., Bigand, E., Belleville, S., Peretz, I., 2012. Music as an aid to learn new verbal information in Alzheimer's disease. Music Percept. Interdiscip. J. 29 (5), 521–531. doi:10.1525/mp.2012.29.5.521.

Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cognit. Sci. 23 (11), 913–926. doi:10.1016/j.tics.2019.08.004.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. doi:10.1155/2011/156869.

O'Sullivan, J., Reilly, R., Lalor, E., 2015. Improved decoding of attentional selection in a cocktail party environment with EEG via automatic selection of relevant independent components. In: Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5740–5743. doi:10.1109/EMBC.2015.7319696.

Park, H., Ince, R.A., Schyns, P.G., Thut, G., Gross, J., 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr. Biol. 25 (12), 1649–1653. doi:10.1016/j.cub.2015.04.049.

Patel, A., 2003. Language, music, syntax and the brain. Nat. Neurosci. 6 (7), 674–681. doi:10.1038/nn1082.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23 (6), 1378–1387. doi:10.1093/cercor/bhs118, (New York, N.Y.: 1991).

Peretz, I., 2009. Music, language and modularity framed in action. Psychol. Belg. 49 (2-3), 157–175. doi:10.5334/pb-49-2-3-157.

Poeppel, David, 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymetric sampling in time'. Speech Communication 41, 245–255.

Purnell-Webb, P., Speelman, C.P., 2008. Effects of music on memory for text. Percept. Mot. Skills 106 (3), 927–957. doi:10.2466/PMS.106.3.927-957.

Rainey, D.W., Larsen, J.D., 2002. The effects of familiar melodies on initial learning and long-term memory for unconnected text. Music Percept. 20 (2), 173–186. doi:10.1525/mp.2002.20.2.173.

Rathcke, T., Falk, S., Dalla Bella, S., 2021. Music to your ears: Sentence sonority and listener background modulate the "speech-to-song illusion. Music Percept. Interdiscip. J. 38 (5), 499–508. doi:10.1525/mp.2021.38.5.499.

Reetzke, R., Gnanateja, G.N., Chandrasekaran, B., 2021. Neural tracking of the speech envelope is differentially modulated by attention and language experience. Brain Lang. 213. doi:10.1016/j.bandl.2020.104891.

Rimmele, J.M., Golumbic, E.Z., Schroger, E., Poeppel, D., 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex 68, 144–154. doi:10.1016/j.cortex.2014.12.014.

Rogalsky, Corianne, Rong, Feng, Saberi, Kourosh, Hickok, Gregory, 2011. Functional anatomy of language and music perception: Temporal and structural factors investigated using functional magnetic resonance imaging. The Journal of Neuroscience 31 (10), 3843–3852.

Saito, Y., Ishii, K., Sakuma, N., Kawasaki, K., Oda, K., Mizusawa, H., 2012. Neural substrates for semantic memory of familiar songs: Is there an interface between lyrics and melodies? PLoS One 7 (9). doi:10.1371/journal.pone.0046354.

Serrà, J., Corral, Á., Boguñá, M., Haro, M., Arcos, J.L., 2012. Measuring the evolution of contemporary western popular music. Sci. Rep. 2. doi:10.1038/srep00521.

Simmons-Stern, N.R., Budson, A.E., Ally, B.A., 2010. Music as a memory enhancer in patients with Alzheimer's disease. Neuropsychologia 48 (10), 3164–3167. doi:10.1016/j.neuropsychologia.2010.04.033.

Simmons-Stern, N.R., Deason, R.G., Brandler, B.J., Frustace, B.S., O'Connor, M.K., Ally, B.A., Budson, A.E., 2012. Music-based memory enhancement in Alzheimer's disease: promise and limitations. Neuropsychologia 50 (14), 3295–3303. doi:10.1016/j.neuropsychologia.2012.09.019.

Slevc, L.R., Rosenberg, J.C., Patel, A.D., 2009. Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. Psychon. Bull. Rev. 16, 374–381. doi:10.3758/16.2.374.

Song, J., Iverson, P., 2018. Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. Cognition 179, 163–170. doi:10.1016/j.cognition.2018.06.001.

Sternin, A., Mcgarry, L., Owen, A., Grahn, J.A., 2021. The effect of familiarity on neural representations of music and language. J. Cognit. Neurosci. 33 (8), 1595–1611. doi:10.1162/jocn_a_01737.

Stolk, A., Verhagen, L., Schoffelen, J-M., Oostenveld, R., Blokpoel, M., Hagoort, P., van Rooij, I., Toni, I., 2013. Neural mechanisms of communicative innovation. Proc. Natl. Acad. Sci. 110 (36), 14574–14579. doi:10.1073/pnas.1303170110.

Tamminen, J., Rastle, K., Darby, J., Lucas, R., Williamson, V., 2017. The impact of music on learning and consolidation of novel words. Memory 25. doi:10.1080/09658211.2015.1130843.

Tierney, A., Dick, F., Deutsch, D., Sereno, M., 2012. Speech versus song: multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. Cereb. Cortex 23, 249–254. doi:10.1093/cercor/bhs003.

Tierney, A., Patel, A.D., Breen, M., 2018. Acoustic foundations of the speech-to-song illusion. J. Exp. Psychol. Gen. 147 (6), 888–904. doi:10.1037/xge0000455.

van den Bosch, I., Salimpoor, V.N., Zatorre, R.J., 2013. Familiarity mediates the relationship between emotional arousal and pleasure during music listening. Front. Hum. Neurosci. 7. doi:10.3389/fnhum.2013.00534.

Vanden Bosch der Nederlanden, C.M., Hannon, E.E., Snyder, J.S, 2015a. Everyday musical experience is sufficient to perceive the speech-to-song illusion. J. Exp. Psychol. Gen. 144 (2), e43–e49. doi:10.1037/xge0000056.

Vanden Bosch der Nederlanden, C.M., Hannon, E.E., Snyder, J.S, 2015b. Finding the music of speech: musical knowledge influences pitch processing in speech. Cognition 143, 135–140. doi:10.1016/j.cognition.2015.06.015.

Vanden Bosch der Nederlanden, C.M., Joanisse, F.M., Grahn, J.A, 2020. Music as a scaffold for listening to speech: better neural phase-locking to song than speech. Neuroimage 214. doi:10.1016/j.neuroimage.2020.116767.

Vanthornhout, J., Decruy, L., Francart, T., 2019. Effect of task and attention on neural tracking of apeech. Front. Neurosci. 13. doi:10.3389/fnins.2019.00977.

Veiga, M., Brandon, D., Holuszko, M., 2015. Teaching cleaner and responsible mining through songs. Extr. Ind. Soc. 2. doi:10.1016/j.exis.2015.02.001.

Wallace, W.T., Rubin, D.C., 1991. Characteristics and constraints in ballads and their effects on memory. Discourse Process. 14 (2), 181–202. doi:10.1080/01638539109544781.

Wallace, W.T., 1994. Memory for music: effect of melody on recall of text. J. Exp. Psychol. Learn. Mem. Cognit. 20 (6), 1471–1485. doi:10.1037/0278-7393.20.6.1471.

Weiss, M.W., Trehub, S.E., Schellenberg, G.E., 2012. Something in the way she sings: enhanced memory for vocal melodies. Psychol. Sci. 23 (10), 1074–1078. doi:10.1177/0956797612442552.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14 (4), 1908–1919. doi:10.1523/JNEUROSCI.14-04-01908.1994.

Zatorre J, Robert, Evans C, Alan, Meyer, Ernst, Gjedde, Albert, 1992. Laterlaization of phonetic and pitch discrimination on speech processing. Science 256 (5058), 846–849.

Zatorre J, Robert, Gandour T, Jackson, 2008. Neural specializations for spech and pitch: moving beyond the dichotomies. Philosophical Transactions of the Royal Society B 363, 1087–1104. doi:10.1098/rstb.2007.2161.

Zhang, Linjun, Shu, Hua, Zhou, Fengying, Wang, Xiaoyi, Li, Ping, 2010. Common and distinct neural substrates for the perception of speech rhythm and intonation. Human Brain Mapping 31, 1106–1116.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E, 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". Neuron 77 (5), 980–991. doi:10.1016/j.neuron.2012.12.037.

Zoefel, B., VanRullen, R., 2015. The role of high-level processes for oscillatory phase entrainment to speech sound. Front. Hum. Neurosci. 9. doi:10.3389/fnhum.2015.00651.

Zoefel, B., Allard, I., Anil, M., Davis, M.H., 2020. Perception of rhythmic speech is modulated by focal bilateral transcranial alternating current stimulation. J. Cognit. Neurosci. 32 (2), 226–240. doi:10.1162/jocn_a_01490.

Zou, J., Feng, J., Xu, T., Jin, P., Luo, C., Zhang, J., Pan, X., Chen, F., Zheng, J., Ding, N., 2019. Auditory and language contributions to neural encoding of speech features in noisy environments. Neuroimage 192, 66–75. doi:10.1016/j.neuroimage.2019.02.047.