WILEY

RESEARCH ARTICLE OPEN ACCESS

Temporal Structure of Music Improves the Cortical Encoding of Speech

Laura Fernández-Merino[1,2] 🄳 | Mikel Lizarazu[1] 🄳 | Nicola Molinaro[1,3] 🄳 | Marina Kalashnikova[1,3]

[1]Basque Center on Cognition, Brain and Language, San Sebastian, Spain | [2]University of the Basque Country (Universidad del País Vasco/Euskal Herriko Unibertsitatea), San Sebastian, Spain | [3]Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Correspondence: Laura Fernández-Merino (laurafmg@hotmail.com)

ABSTRACT

Long- and short-term musical training has been proposed to improve the efficiency of cortical tracking of speech, which refers to the synchronization of brain oscillations and the acoustic temporal structure of external stimuli. Here, we study how musical sequences with different rhythm structures can guide the temporal dynamics of auditory oscillations synchronized with the speech envelope. For this purpose, we investigated the effects of prior exposure to rhythmically structured musical sequences on cortical tracking of speech in Basque–Spanish bilingual adults (Experiment 1; *N* = 33, 22 female, Mean age = 25 years). We presented participants with sentences in Basque and Spanish preceded by musical sequences that differed in their rhythmical structure. The rhythmical structure of the musical sequences was created to (1) reflect and match the syllabic structure of the sentences, (2) reflect a regular rhythm but not match the syllabic structure of the sentences, and (3) follow an irregular rhythm. Participants' brain responses were recorded using electroencephalography, and speech-brain coherence in the delta and theta bands was calculated. Results showed stronger speech-brain coherence in the delta band in the first condition, but only for Spanish stimuli. A follow-up experiment including a subset of the initial sample (Experiment 2; *N* = 20) was conducted to investigate whether language-specific stimuli properties influenced the Basque results. Similar to Experiment 1, we found stronger speech-brain coherence in the delta and theta bands when the sentences were preceded by musical sequences that matched their syllabic structure. These results suggest that not only the regularity in music is crucial for influencing cortical tracking of speech, but so is adjusting this regularity to optimally reflect the rhythmic characteristics of listeners' native language(s). Despite finding some language-specific differences across frequencies, we showed that rhythm, inherent in musical signals, guides the adaptation of brain oscillations, by adapting the temporal dynamics of the oscillatory activity to the rhythmic scaffolding of the musical signal.

1 | Introduction

There is extensive evidence that experienced musicians and even individuals with more limited musical training exhibit cognitive advantages when processing speech (see Neves et al. 2022 for a recent review). This study aims to investigate the neural mechanisms underlying these effects and the precise characteristics of the musical signal that lead to the music-to-speech processing benefits. Specifically, we focus on the mechanism of cortical tracking of speech, which refers to the synchronization between neural oscillations and the acoustic properties of an external stimulus such as speech (see Meyer 2018 for a review). We propose that rhythmic auditory stimulation found in the musical signal acts as a temporal guide that helps to maintain

---

the temporal dynamics of the synchronization with the speech signal over time.

Music and language communication are human cultural universals (Peretz 2006). There is a growing body of literature aiming to establish links between music and language processing in the brain, especially music-to-speech processing benefits (see Besson et al. 2017 for a review). Research involving professional musicians has shown that music training enhances understanding of speech in noise and sensitivity to prosodic cues (Zendel et al. 2019). More specifically, music training improves auditory processing in noisy conditions by strengthening the sensorimotor coupling between premotor and auditory regions (Du and Zatorre 2017; Herholz and Zatorre 2012). This, in turn, is related to the finer analysis of the spectral properties of the auditory signal, possibly due to greater neural temporal precision (Tierney and Kraus 2013; Wong et al. 2007), and has positive effects on audiovisual speech integration (Musacchia et al. 2007).

Short-term effects of music training (i.e., in non-musicians) have also been shown. For instance, 8 weeks of music training were shown to improve pitch discrimination abilities in children (Moreno and Besson 2006), and 6 months of music training enhanced both pitch discrimination in speech and reading skills (Moreno et al. 2009). In adolescents, 2 years of music instruction improved the neural representation of speech in noisy environments (Tierney et al. 2013), and 3 years of training resulted in enhanced subcortical sound processing and accelerated cortical auditory maturation (Tierney et al. 2015). These benefits also extend to broader language skills, including vocabulary (Piro and Ortiz 2009), phonological awareness (Degé and Schwarzer 2011), and auditory processing (Habibi et al. 2016), and have been observed across age groups, from young infants (Gerry et al. 2012; Zhao and Kuhl 2016) to adults (see Neves et al. 2022).

Based on this growing evidence, some theories have outlined shared underlying processes for music and speech processing that may explain why there is an improvement in speech processing from music exposure. The Overlap, Precision, Emotion, Repetition, and Attention (OPERA) hypothesis (Patel 2011, 2014) proposes that music training enhances speech processing when music places higher demands than speech on shared cognitive processes and engages these processes in the context of emotion, repetition, and attention. The Precise Auditory Timing Hypothesis (PATH; Tierney and Kraus 2014) explains enhanced linguistic abilities in musicians through cross-domain auditory improvement; in other words, changes in neural processing in one domain are driven by experience or training in another. While both these frameworks suggest potential transfer effects across domains, these effects appear to be primarily associated with sustained engagement or music training. However, brief musical exposure, which consists of isolated experiences with music rather than the cumulative and structured learning characteristic of musical training, may also lead to similar processing benefits.

It has also been proposed that there is a shared nature of structural processing in language and music—including rhythm processing (Fiveash et al. 2021). Language and music are hierarchically organized, and as such, one of the main foci of the last few years has been to establish the neurobiological underpinnings of how the human brain processes these hierarchies. One of the perspectives that has been adopted to study this hypothesis posits that neural activity synchronizes with the rhythmical patterns of external inputs, such as speech and musical signals (Obleser and Kayser 2019). This synchronization mechanism, referred to as *cortical tracking* hereafter (also referred to as neural entrainment in some literature, see Bánki et al. 2022; and Doelling and Assaneo 2021 for a discussion of this terminology), is reflected in the phase and amplitude of the neural oscillations that synchronize with the stimulus (see Doelling and Assaneo 2021 for review). Cortical tracking of auditory stimuli plays a role in processing the linguistic structure of speech, via aligning specific neural oscillators to specific linguistic units: phonological rate (gamma, > 30 Hz) (Hyafil et al., 2015), syllabic rate (theta, 4–8 Hz) (Peelle et al. 2013), and lexical and phrasal rate (delta, 1–3 Hz) (Bourguignon et al. 2013; Kong et al. 2015). The tracking of the incoming signal allows brain oscillations to synchronize with external regularities and predictable cues (e.g., beat in music or stress in speech), and to structure the auditory input. It also helps focus attention on important elements of the auditory stimulus and its presentation over time (Ghitza 2011; Giraud and Poeppel 2012; Peelle and Davis 2012), supporting the segmentation and identification of linguistic units in speech (Meyer 2018). Accurate tracking of acoustic temporal regularities correlates with speech intelligibility (Ahissar et al., 2001; Luo and Poeppel 2007), and it has been suggested to be responsible for accurate syllabic parsing (Doelling et al. 2014). In contrast, diminished cortical tracking has been found in challenging listening conditions (Peelle et al. 2013). While it is well established that cortical tracking underlies the perception of rhythmic regularities in speech and music, whether and how this mechanism may modulate the music-to-speech processing benefits is still under investigation.

Music and speech share a structure and hierarchy with a common underlying denominator—rhythm (Ding et al. 2017; Varnet et al. 2017). A growing body of literature highlights the importance of rhythm as a critical factor in language acquisition, as well as in speech and music processing, emphasizing its role in coordinating temporal patterns and facilitating cognitive and perceptual alignment (Goswami 2022; Ladányi et al. 2020). From the perspective of cortical tracking, for instance, delta brain oscillations phase-align with the beats of the rhythm of speech and music (Lakatos et al. 2013, 2016; Poeppel and Assaneo 2020; Schroeder and Lakatos 2009). There is also evidence for cross-modal effects on cortical tracking. Rhythmic primes from music have been shown to enhance the processing of subsequent speech stimuli, such as when the syllabic structure of pseudo-words matches the metrical structure of a preceding rhythmic musical prime in phonological (Cason and Schön 2012), prosodic (Ríos-López et al. 2017), and syntactic perception (Przybylski et al. 2013).

The abovementioned findings suggest that exposure to rhythmic regularities in the musical signal can enhance subsequent speech processing by aligning the phase of neural oscillations and resulting in enhanced cortical tracking of speech. Furthermore, there is evidence that the music-to-speech benefit in cortical tracking is driven by the similarities between the rich and complex temporal structure of music and the prosodic and syllabic rhythms in speech. Falk et al. (2017) demonstrated that regular

musical rhythms, which reflect the temporal and prosodic structure of speech, enhance inter-trial phase coherence during speech processing. In their study, French participants listened to speech sequences preceded by musical sequences with either regular or irregular rhythms. Regular rhythms in music matched the meter (1.65 Hz) and beat (5 Hz) rates in speech, corresponding to its prosodic and syllabic rhythms (delta and theta frequency bands in the EEG signal, respectively). Analyses of electroencephalography (EEG) data revealed stronger phase coherence in the theta band (~5 Hz) in the regular rhythm condition. This was likely due to French being a syllable-timed language with prominent theta-band modulations. Thus, Falk et al. demonstrated a music-to-speech benefit on cortical tracking, but they left unanswered the question of whether a precise matching between the rhythmic pattern of the musical sequences and the speech sequences is required for this benefit to emerge. In this study, we investigate whether cortical tracking can be enhanced when speech is preceded by regular musical rhythms that do not precisely mimic the syllabic structure of speech. If so, it would suggest that the music-to-speech tracking effect is driven by stimulation of the broader prosodic rhythms in speech, rather than by rigid temporal matching.

For this purpose, we used EEG to measure the cortical tracking of speech sequences that were preceded by musical sequences with varying rhythmic properties, belonging to three experimental conditions: Matching Regular sequences, which mirrored the syllabic and temporal structure of the speech; Mis-matching Regular sequences, which followed a regular rhythm but did not align with the syllabic structure of the speech; and Irregular sequences, which followed no consistent rhythmic pattern. This design allowed us to discern the roles of rhythm regularity and temporal alignment in driving cortical tracking of speech. We hypothesized that if it is the case that rhythm regularity in the musical sequences alone is sufficient to elicit the music-to-speech benefit in cortical tracking, then enhanced cortical tracking would be observed in both Matching and Mis-matching regular conditions compared to the Irregular condition. Conversely, if temporal alignment is required, then enhanced cortical tracking of speech would be observed in the Matching Regular condition, followed by the Mis-matching Regular and Irregular conditions.

Our experiment also pursued the secondary aim consisting of assessing how language-specific rhythmic structures influence the music-to-speech cortical tracking effects predicted above. This was possible since all participants in this study were balanced Basque–Spanish bilinguals. Basque and Spanish are classified as syllable-timed languages, but unlike Spanish, Basque also displays stress-timed characteristics due to its syntactic and syllabic structure (Molnar et al. 2014, 2016). Thus, our musical stimuli were manipulated to reflect the prosodic and syllabic rates of each language. We hypothesized that participants would show enhanced cortical tracking of both Spanish and Basque speech sequences following exposure to musical sequences aligned with the temporal structure of the respective language. As explained in greater detail in the following sections, the initial manipulation of the Basque stimuli resulted in low ecological validity, which led us to correct this issue and conduct a second experiment that included a subset of participants from the first experiment tested solely on the Basque half of the task.

This enabled us to test more specifically the extent to which music-to-speech benefits are generalizable across languages.

In view of the robust effects of musical training and expertise on cortical tracking of speech discussed above, we also collected information about participants' musical backgrounds to explore the relation between prolonged musical training and the music-to-speech cortical tracking effects in our sample. Specifically, we employed a questionnaire in which participants reported how many hours per week they spent engaging in musical activities (e.g., playing an instrument, singing, or listening to music) and whether they had received any formal musical training. We hypothesized that participants with higher levels of musical engagement would exhibit enhanced cortical tracking to speech in the Matching and Mis-matching regular conditions.

2 | Experiment 1

2.1 | Method

2.1.1 | Participants

A total of 33 Basque–Spanish bilinguals (11 males, *M* age: 25 years, SD = 3.31) recruited through the Basque Center on Cognition, Brain, and Language (BCBL) participant database took part in this study. Language proficiency was assessed with the Basque, English, and Spanish Test (BEST) (de Bruin et al. 2017), which measures fluency, lexical resources, grammatical constructions, and pronunciation (a Likert scale with scores ranging from 1 to 5). All participants were highly proficient in Basque and Spanish (Maximum score: 65; Spanish *M* = 64, SD = 1.01; Basque *M* = 58.09, SD = 5.95). Participants' musical background was assessed using the Goldsmiths Musical Sophistication Index (Gold MSI) self-report questionnaire (Müllensiefen et al. 2014), which measures different aspects of musical sophistication, such as active musical engagement, perceptual abilities, musical training, singing abilities, emotional engagement with music, providing a general musical sophistication factor score.

The Ethics and Scientific Committee of the BCBL approved the study protocol (approval number 220321SM), which was developed following the declaration of Helsinki. All participants gave their written informed consent before the study and received monetary compensation.

2.1.2 | Stimuli

Speech stimuli consisted of sixty-four 12-syllable sentences in both Spanish and Basque (128 in total) recorded by a female Basque–Spanish bilingual speaker. We used stimuli recorded in infant-directed speech (IDS) for an identical study with infants because this register provides the ideal context to stress and highlight the temporal modulations of speech (Leong et al. 2017) and ensures natural and highly rhythmical stimuli that highlight the syllable information. Other linguistic properties were considered when constructing the sentences, including natural prosody, grammatical correctness, and lexical frequency to ensure ecological validity.

This experimental design required the creation of three types of musical stimuli for the Matching Regular, Mismatching regular, and Irregular conditions. The musical stimuli for the Matching regular and Mismatching regular conditions were created based on the temporal structure of the sentences. The Matching Regular musical sequences reflected the temporal structure and pitch contour of linguistic stimuli. For this purpose, we extracted the rhythmic structure of the sentences to create the beat structure. This structure was introduced in Musescore 3 (https://musescore.org/es), an open-source software where, using a grand piano timbre, we created the melodic contour from the pitch contour of the sentences. The Mis-matching regular musical sequences preserved the pitch contour of the linguistic stimuli but followed a temporal structure with notes that did not match the syllabic structure of the speech stimuli. To create these sequences, we took the Matching Regular sequences and randomly altered the duration of the notes, preserving a regular meter. Finally, the Irregular musical sequences were created by randomly altering the pitch and note duration of the Matching Regular sequences, thus also altering their regular meter. The melodic sequences that comprised a condition formed a musical structure of A A A A′ (A′ was a slight variation of the melodic contour and temporal structure of A), following the common musical pattern of Basque and Spanish infant-directed songs.

All speech and melodic sequences had a duration of 2500 ms. To ensure highly rhythmical stimuli, each speech sequence was recorded and cued by a metronome at a regular rate of 120 bpm, reflecting the rate of stressed syllables, obtaining a regular meter with an inter-onset interval of 500 ms between stressed syllables. The musical sequences were created based on the temporal patterns of the sentences. There was a constant 500 ms inter-onset interval between stressed notes for the Matching Regular and Mis-matching Regular musical sequences and no constant inter-onset interval for the Irregular sequences.

Given that Spanish and Basque have different rhythmical properties (Molnar et al. 2016; Molnar et al. 2014), spectral analyses of the melodic and speech sequences revealed different frequencies for each language, with averaged peaks at 2 and 4 Hz in Spanish and peaks at 1.6 and 4 Hz in Basque, as shown in Figure 1. Besides, we estimated the syllable rate of the sentences using an automatic algorithm based on the method from de Jong and Wempe (2009) and Praat (Boersma & Weenink, 2021). The overall stressed rate in Spanish was found to be 2.15 Hz (SD = 0.25) and 1.56 Hz (SD = 0.28) in Basque, with the average syllable rate in Spanish being 4.23 Hz (SD = 0.33) and 4.16 Hz (SD = 0.28) in Basque, showing a high degree of similarity between the two languages. Based on the frequency resolution of 0.4 Hz used in our speech-brain coherence analysis, the frequency bin that most closely aligned with the stressed syllable rate was 2 Hz for Spanish and 1.6 Hz for Basque. The syllabic rate in both languages was most strongly represented at 4 Hz.

These peaks were expected because they correspond to the prosodic (reflected in the delta frequency band, ~2 Hz) and syllabic (reflected in the theta frequency band, ~4 Hz) rhythms. Thus, the difference in Matching Regular and Mis-matching Regular

musical sequences remained in the matching between notes and syllables. No specific frequency was reflected in the Irregular musical sequences.

2.1.3 | Procedure

Participants completed all three conditions in Spanish and Basque in a single experimental session: speech preceded by Matching Regular musical sequences, speech preceded by Mismatching Regular cues, and speech preceded by Irregular musical cues. Block structure is shown in Figure 2. Only one type of cue, Matching Regular, Mis-matching Regular, and Irregular, and one language were used within a single experimental block. There was a total of 96 blocks, with a total experiment duration of 32 min. There were 16 blocks per language (Basque or Spanish) and per rhythm condition (Matching, Mis-matching, and Irregular); each condition had a duration of ~5′30″. All linguistic stimuli were preceded by Matching Regular musical sequences, Mis-matching regular musical sequences, and Irregular musical sequences. A single block consisted of four trials and had a total duration of 20 s. Each trial comprised one music sequence immediately followed by a simple sentence and had a duration of 5000 ms. The stimulus onset asynchrony (SOA) between the melodic sequence and the sentence was always 500 ms.

To ensure participants′ attention during data acquisition, after each block, a word appeared on the screen and participants had to indicate, by pressing a key, whether they had or had not heard that word in the sentences in that block. Half of the words appeared in the sentences, and half of the words did not. The next block started immediately after a key was pressed.

The order of condition and language was counterbalanced across participants. Stimuli were presented in PsychoPy (v.1.80.04) via two speakers positioned about 1 m in front of the participants at 65 dB.

For all participants, we also recorded resting state EEG data as participants looked at a fixation point for 5 min at the end of the experimental session.

2.1.4 | EEG Data Acquisition and Preprocessing

EEG data were acquired using a BrainAmp amplifier and BrainVision Recorder software (Brain Products, Germany). EEG was recorded using 32 electrodes that were positioned according to the international 10–20 system (Jasper 1958). Scalp-electrode impedance was kept below 5 kΩ for scalp electrodes and under 10 kΩ for reference and EOG electrodes to ensure high-quality EEG recordings. EEG data was sampled at a rate of 1000 Hz and band-pass filtered online from 0.1 to 1000 Hz. The recording was referenced to electrode FCz online. Electrode AFz was used as the ground. Additionally, two electrodes at the outer canthi of the eyes recorded horizontal eye movements, while electrodes above and below the right eye recorded vertical eye movements.

All the following processing and analysis steps were implemented in Matlab (The MathWorks Inc., Natick, US) using FieldTrip (Oostenveld et al. 2011) and available custom code
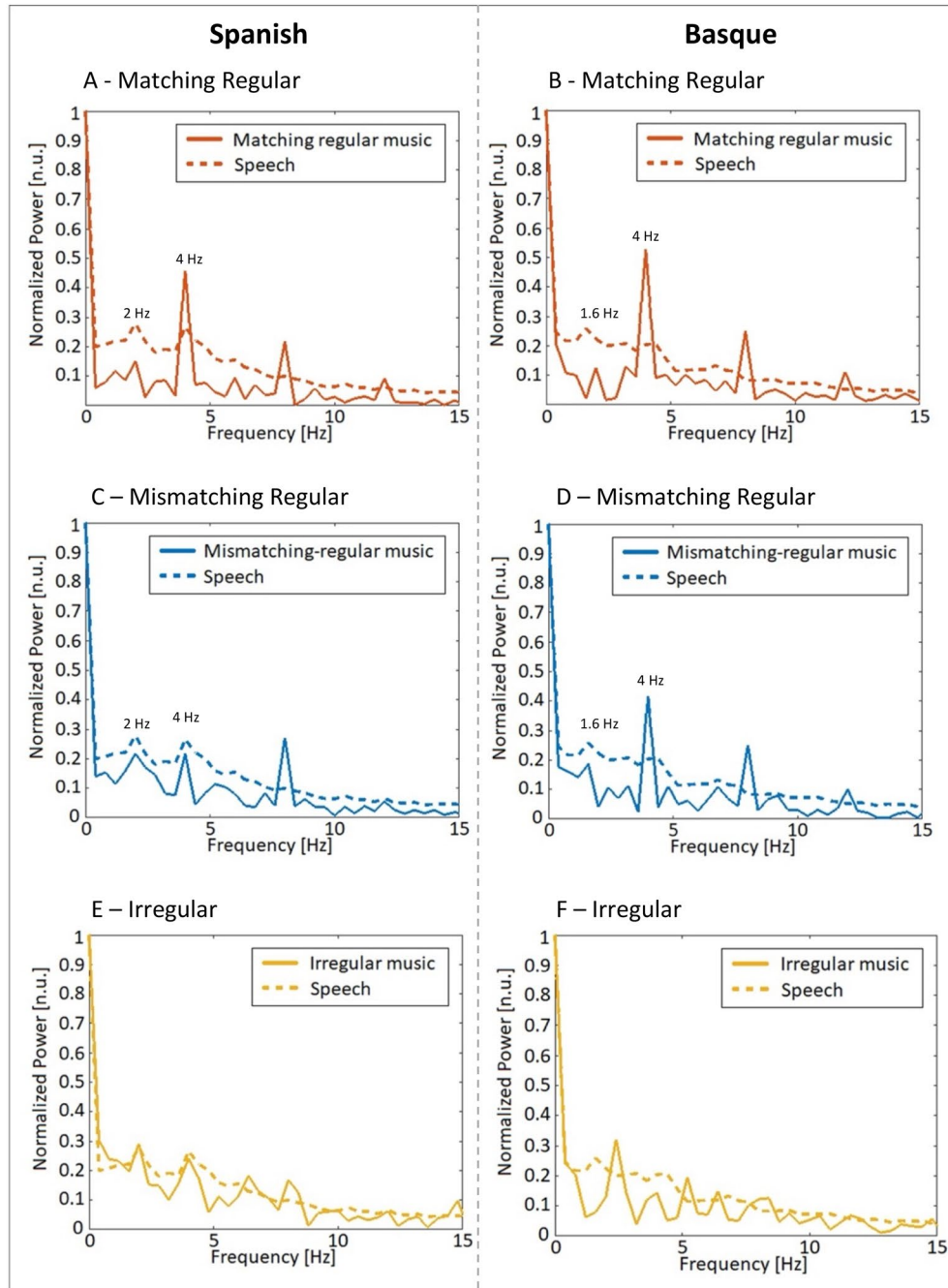
FIGURE 1 | Acoustic properties of Spanish and Basque musical and speech sequences in Experiment 1. Here, we show the spectral information of speech and musical sequences in normalized power (scaled power of signal). Spectra on the left (A, C, E) correspond to Spanish stimuli, showing peaks at 2 and 4 Hz for the Matching and Mis-matching Regular condition. Spectra on the right (B, D, F) correspond to Basque stimuli, showing peaks at 1.6 and 4 Hz.

(Fernández-Merino and Lizarazu 2023). Data were segmented from 0 to 2.5 s relative to stimulus onset, which gave a total of 768 trials. These trials were re-referenced off-line to the left mastoid and low-pass filtered below 30 Hz, since we did not expect any coherence effect above this threshold (see Gross et al. 2013). Filtered data were then resampled to 200 Hz. Electrooculogram artifacts were detected using Independent Component Analysis (ICA) and linearly subtracted from recordings (*fastICA* algorithm implemented in FieldTrip). Artifact rejection was also carried out by excluding all trials with a *z*-score above the variance threshold of 3. A minimum of 75% artifact-free trials per participant was required for inclusion in subsequent analyses (M = 12, SD = 4.5).

2.1.5 | Coherence Analysis

We quantified phase synchronization, referred to as coherence, between the EEG signal from each channel ($x(t)$) and the envelopes of the corresponding auditory stimulus ($y(t)$). The envelopes were extracted from the Hilbert-transformed broadband stimulus waveforms and resampled from 44,100 to 200 Hz
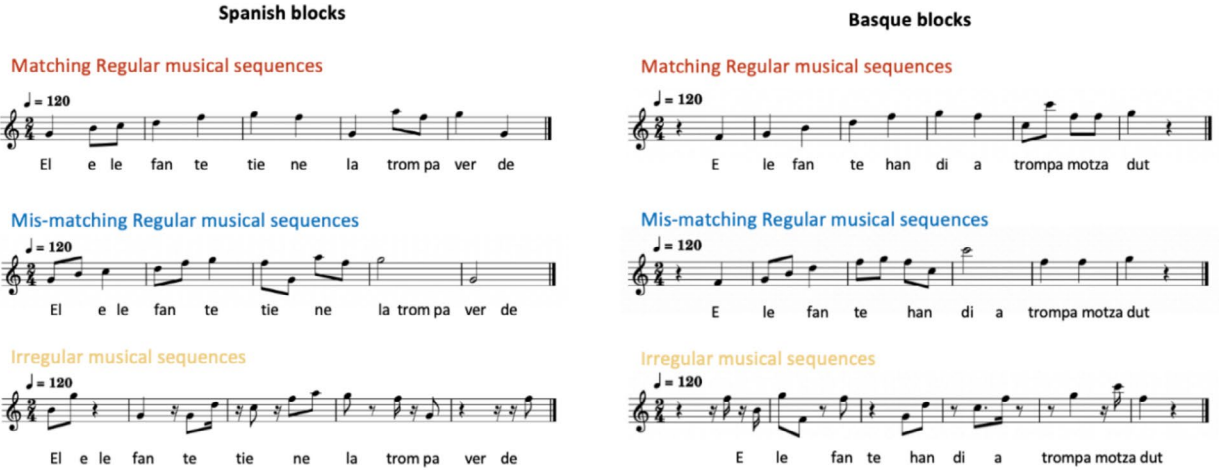
FIGURE 2 | Procedure. (1) Graphical example of an experimental block. The block comprises four trials. Each trial comprises one melodic sequence followed by a sentence. Only one type of cue (Matching, Mismatching, or Irregular) was used within a block. After each block, a word appeared on the screen. (2) Example of the musical sequences for each condition. Below each musical stave, an example of the sentence that would follow the musical sequences (English translation: The elephant's trunk is green). Note that only in the Matching Regular musical sequences, the notes structure of the musical sequences and the syllabic structure of the sentences are matched.

to match the EEG data sampling rate (Molinaro et al. 2016; Molinaro and Lizarazu 2018).

Coherence measures the degree of phase coupling between $x(t)$ and $y(t)$ on a scale from 0 to 1, where 0 indicates no linear relationship and 1 indicates a perfect linear relationship. A trial length of 2.5 s provided a frequency resolution of approximately 0.4 Hz, corresponding to the inverse of the trial duration. Subsequent analyses were conducted separately for each experimental condition. Cross-spectral density (P_{xy}) and power spectral densities (P_{xx} and P_{yy}) were estimated using the Fourier transform. Finally, coherence was calculated according to the formula described by Halliday (1995).

$$\text{Coh}_{xy}(f) = \frac{\left| P_{xy}(f) \right|^2}{P_{xx}(f) P_{yy}(f)'}$$

Using this methodology, we obtained coherence values for each participant, condition, EEG sensor, and frequency bin below 15 Hz.

The coherence bias was estimated empirically for each participant by randomly shuffling the auditory envelopes across trials and recalculating coherence in 500 permutations with resting state data (Molinaro et al. 2016; vanden Bosch der Nederlanden et al. 2022). For each sensor and frequency bin, coherence data were z-score transformed using the mean and standard deviation from the 500 random resting state EEG-audio pairings.

For the subsequent analyses, we extracted the maximum z-scored coherence values of our frequencies of interest from the nearest neighbor bins per band from each channel for each participant. Figures 3 and 4, bottom, show the sensors with the highest coherence values per frequency band and stimuli type. In the Spanish blocks, maximum values were extracted from 1.6 to 2.4 Hz in the case of delta, and 3.6 to 4.4 Hz in the case of theta. In the Basque blocks, maximum values were extracted from 1.2 to 2 Hz in the case of delta, and 3.6 to 4.4 Hz in the case of theta. Trials that deviated by 2 SD or more from the mean were removed from these analyses and are reported in Supporting Information, Appendix.

2.1.6 | Statistical Analysis

To ensure that participants' language proficiency in the two languages was comparable, BEST scores were compared using paired sample t-tests. Participants' answers in the behavioral task were scored as correct when the participant correctly identified that the word had or had not appeared in the preceding test block. Accuracy scores were calculated for each condition in each language task. The total number of correct responses
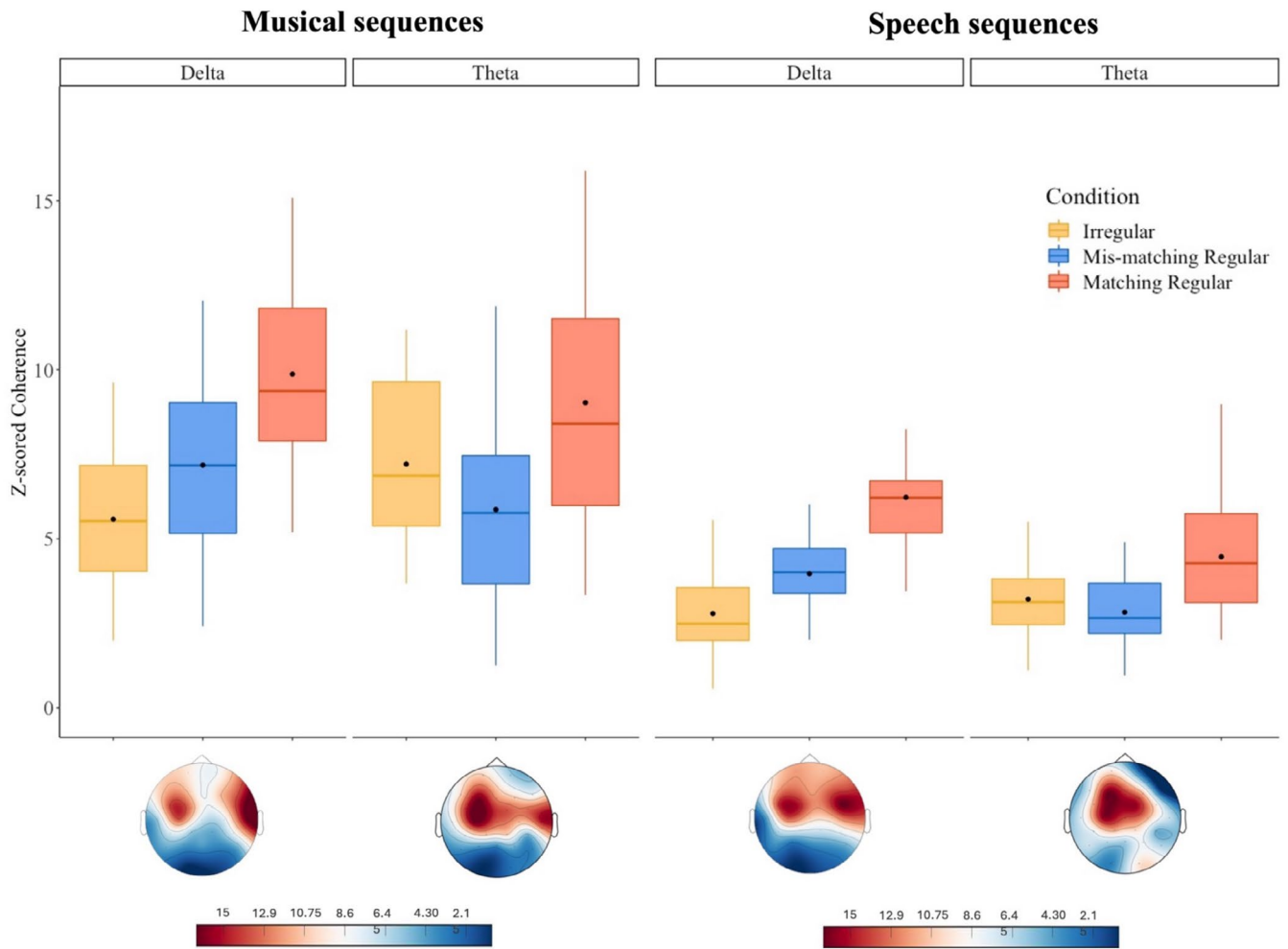
FIGURE 3 | Cortical tracking (z-scores indicating stimulus-brain coherence) of musical and speech sequences in the Spanish blocks. The top left and right panels show the delta and theta frequency bands for music (left) and speech (right). The bottom left and right panels show the topographies of the maximum coherence values extracted for the analysis per frequency and stimulus type.

was transformed into a percentage. Due to a technical problem with the data collection, we could not collect a complete set of responses from one participant. Behavioral scores were compared across conditions using ANOVA.

Analyses of the cortical tracking data were conducted using Linear mixed-effects models constructed with the z-scores transformed maximum coherence values as the dependent variable, including a two-way interaction: Condition (Matching regular, Mis-matching regular, and irregular) × Frequency band (Delta and Theta) per stimuli type (Speech and Music). As noted above, the Delta and Theta frequency bands were analyzed because they reflect the prosodic and syllabic rhythms of Spanish and Basque. Language was not included as a variable in the model since the goal of the study was to measure cortical tracking outcomes based on the different rhythmic structures of each language. In addition, we did not expect our balanced bilinguals to perform better in one language compared to the other. Random intercepts were specified per participant. The continuous variables entered into the analyses were scaled and centered around zero to assist with model convergence. Main effects and interactions were further explored by post hoc comparisons using Bonferroni corrections. Analyses were conducted using the

lme4 (Bates 2018) and lmertest (Kuznetsova et al. 2017) packages in R.

We also conducted exploratory analyses using the language proficiency and music experience data that are reported in Experiment 2. Pearson correlations were used for this purpose, given that normality, linearity, and homoscedasticity were assessed and met. Correlations were run between z-scored coherence values for each language, stimulus type, frequency bands, condition, participants' BEST scores, and participants' individual musical background scores.

2.2 | Results

2.2.1 | Participants' Language Proficiency and Musical Background

All participants were highly proficient in Basque and Spanish (Maximum score: 65; Spanish $M = 64$, SD = 1.01; Basque $M = 58.09$, SD = 5.95). Although participants showed higher scores in Spanish compared to Basque, as demonstrated by a paired samples t-test, $t(32) = 5.038$, $p < 0.001$, it is worth noting
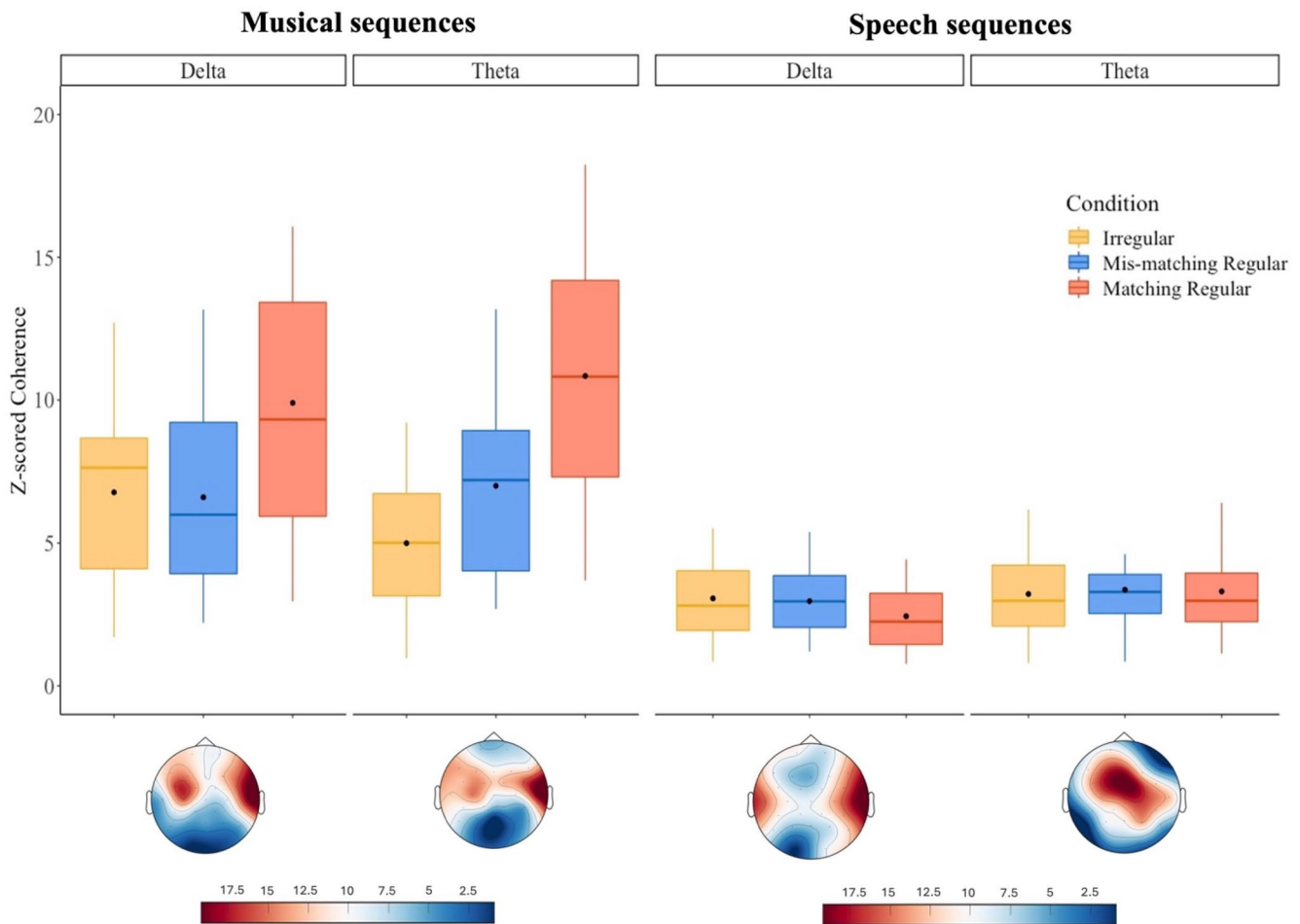
FIGURE 4 | Cortical tracking (*z*-scores indicating stimulus-brain coherence) of musical and speech sequences in the Basque blocks. The top left and right panels show the delta and theta frequency bands for music (left) and speech (right). The bottom left and right panels show the topographies of the maximum coherence values extracted for the analysis per frequency and stimulus type.

that picture naming scores in Basque are typically lower than in Spanish because of lower standardization of the test and greater lexical variability in Basque (see de Bruin et al. 2017). Regarding participants' musical background, all participants self-reported to be actively engaged with music: 7 participants had more than 5 years of formal musical education, but no participant reported being a professional musician (general musical sophistication factor, maximum score = 100; $M = 66.45$; SD = 11.37).

2.2.2 | Behavioral Data Analysis

Overall, participants' accuracy scores were high (Spanish: $M = 87.3\%$, SD = 8.9%; Basque: $M = 89.2\%$, SD = 8.15%), indicating that participants were attentive during the task. An ANOVA was run for each language to test if there were any differences in accuracy between the three conditions. No differences between conditions were found in Spanish ($F_{(2, 62)} = 0.394$, $p = 0.67$) or Basque ($F_{(2, 62)} = 0.480$, $p = 0.621$).

2.2.3 | EEG Analysis

We conducted separate analyses for each language portion of the experiment (Basque and Spanish) and stimulus type (speech

sequences and musical sequences). Prior to assessing how exposure to musical sequences with different rhythmic properties impacted cortical tracking of speech, we assessed whether the rhythm manipulation impacted participants' tracking of music. Therefore, this section is divided into Spanish and Basque, and subdivided into cortical tracking of musical sequences and cortical tracking of speech sequences.

2.2.4 | Spanish

2.2.4.1 | Cortical Tracking of Musical Sequences. First, we assessed participants' cortical tracking of musical sequences with different rhythmical properties. Here, we report cortical tracking of musical sequences in the Spanish blocks. The model revealed a main effect of condition ($F_{(2, 144)} = 25.53$, $p < 0.001$), no main effect of frequency ($F_{(1, 144)} = 0.186$, $p = 0.66$), and a significant frequency by condition interaction ($F_{(2, 144)} = 5.486$, $p < 0.005$). The model output can be found in Supporting Information, Appendix, and these results are shown in Figure 3. Post Hoc comparisons revealed that there was higher coherence in the Matching Regular condition than in the Irregular condition ($\beta = -0.938$, SE = 0.147, $t = -6.392$, $p_{bonf} < 0.001$) and in the Mis-matching Regular condition ($\beta = 0.888$, SE = 0.150, $t = 5.943$, $p_{bonf} < 0.001$). There was no significant

difference between the Irregular and Mis-matching regular conditions ($\beta = -0.049$, SE $= 0.148$, $t = -0.333$, $p_{\text{bonf}} = 1$).

Post hoc comparisons conducted on the frequency by condition interaction revealed that in the case of delta (1.6–2.4 Hz), coherence in the Matching Regular condition was higher than in the Mis-matching ($\beta = -0.800$, SE $= 0.216$, $t = 3.699$, $p_{\text{bonf}} = 0.004$) and Irregular conditions ($\beta = -1.309$, SE $= 0.212$, $t = -6.182$, $p_{\text{bonf}} < 0.001$). There was no difference between the Mis-matching and the Irregular conditions ($\beta = -0.508$, SE $= 0.213$, $t = -2.384$, $p_{\text{bonf}} = 0.276$). In the case of theta (3.6–4.4 Hz), coherence in the Matching Regular condition was higher than in the Mis-matching condition ($\beta = 0.976$, SE $= 0.206$, $t = 4.732$, $p_{\text{bonf}} < 0.001$). There was no difference between the Matching Regular condition and the Irregular condition ($\beta = -0.566$, SE $= 0.203$, $t = -2.796$, $p_{\text{bonf}} = 0.088$), nor between the Mis-matching Regular condition and the Irregular condition ($\beta = 0.409$, SE $= 0.206$, $t = 1.986$, $p_{\text{bonf}} = 0.734$).

2.2.4.2 | Cortical Tracking of Speech Sequences. Next, we assessed cortical tracking of Spanish speech sequences. The model revealed a main effect of condition ($F(2, 148) = 54.769$, $p < 0.001$), a main effect of frequency ($F(1, 150) = 16.885$, $p < 0.001$), and a significant frequency by condition interaction ($F(2, 148) = 11.103$ $p < 0.001$). The model output can be found in Supporting Information, Appendix, and these results are shown in Figure 4. Post Hoc comparisons revealed that there was higher coherence in the Matching Regular condition than in the Irregular condition ($\beta = -1.288$, SE $= 0.133$, $t = -9.711$, $p_{\text{bonf}} < 0.001$) and in the Mis-matching Regular condition ($\beta = 1.079$, SE $= 0.132$, $t = 8.175$, $p_{\text{bonf}} < 0.001$). There was no significant difference between the Irregular and Mis-matching Regular condition ($\beta = -0.209$, SE $= 0.133$, $t = -0.1573$ $p_{\text{bonf}} = 0.353$). Coherence in delta (1.6–2.4 Hz) was also significantly higher than coherence in theta (3.6–4.4 Hz) ($\beta = 0.445$, SE $= 0.109$, $t = 4.105$, $p_{\text{bonf}} < 0.001$).

Post hoc comparisons conducted on the frequency by condition interaction revealed that in the case of delta (1.6–2.4 Hz), coherence in the Matching Regular condition was higher than in the Mis-matching Regular ($\beta = 1.253$, SE $= 0.190$, $t = 6.582$, $p_{\text{bonf}} < 0.001$) and in the Irregular conditions ($\beta = -1.896$, SE $= 0.189$, $t = -10.021$, $p_{\text{bonf}} < 0.001$). Coherence in the Mis-matching Regular condition was significantly higher than in the Irregular condition ($\beta = -0.642$, SE $= 0.191$, $t = -3.369$, $p_{\text{bonf}} = 0.014$). In the case of theta (3.6–4.4 Hz), coherence in the Matching Regular condition was higher than in the Mis-matching Regular ($\beta = 0.903$, SE $= 0.183$, $t = 4.947$, $p_{\text{bonf}} < 0.001$) and in the Irregular conditions ($\beta = -0.680$, SE $= 0.186$, $t = -3.658$, $p_{\text{bonf}} = 0.005$). There was no difference between the Mis-matching Regular condition and the Irregular condition ($\beta = 0.223$, SE $= 0.186$, $t = 1.204$, $p_{\text{bonf}} = 1$).

2.2.5 | Basque

2.2.5.1 | Cortical Tracking of Musical Sequences. Next, we assessed cortical tracking of musical sequences in the Basque blocks. The model revealed a main effect of condition ($F(2, 149) = 37.329$, $p < 0.001$), no main effect of frequency ($F(1, 150) = 0.158$, $p = 0.691$), and a significant frequency by condition interaction ($F(2, 149) = 3.180$, $p = 0.044$). The model output can be found in Supporting Information, Appendix, and these results are shown in Figure 4. Post Hoc comparisons revealed that there was higher coherence in the Matching Regular condition than in the Mis-matching Regular condition ($\beta = 0.881$, SE $= 0.135$, $t = 6.539$, $p_{\text{bonf}} < 0.001$) and in the Irregular condition ($\beta = -1.104$, SE $= 0.136$, $t = -8.119$, $p_{\text{bonf}} < 0.001$). There was no significant difference between the Irregular and the Mis-matching Regular condition ($\beta = -0.223$, SE $= 0.137$, $t = -1.63$, $p_{\text{bonf}} = 0.315$).

Post hoc comparisons conducted on the frequency by condition interaction revealed that in the case of delta (1.2–2 Hz), coherence in the Matching Regular condition was higher than in the Mis-matching Regular condition ($\beta = 0.794$, SE $= 0.191$, $t = 4.170$, $p_{\text{bonf}} < 0.001$) and in the Irregular condition ($\beta = -0.772$, SE $= 0.189$, $t = -4.09$, $p_{\text{bonf}} = 0.001$). There was no difference between the Mis-matching Regular condition and the Irregular condition ($\beta = 0.022$, SE $= 0.193$, $t = 0.118$, $p_{\text{bonf}} = 1$). In the case of theta (3.6–4.4 Hz), coherence in the Matching Regular condition was higher than in the Mis-matching Regular condition ($\beta = 0.966$, SE $= 0.190$, $t = 5.078$, $p_{\text{bonf}} < 0.001$) and in the Irregular condition ($\beta = -1.435$, SE $= 0.196$, $t = -7.332$, $p_{\text{bonf}} < 0.001$). There was no difference between the Mis-matching Regular condition and the Irregular condition ($\beta = -0.469$, SE $= 0.194$, $t = 2.417$, $p_{\text{bonf}} = 0.252$).

2.2.5.2 | Cortical Tracking of Speech Sequences. As the next step, we assessed cortical tracking of Basque speech sequences. The model revealed no main effect of condition ($F(2, 149) = 0.931$, $p = 0.396$), a main effect of frequency ($F(1, 152) = 5.952$, $p = 0.015$), and no frequency by condition interaction ($F(2, 150) = 1.160$, $p = 0.316$). The model output is shown in the Supporting Information, Appendix, and these results are shown in Figure 4. Post hoc comparisons revealed that there was higher coherence in theta (3.6–4.4 Hz) than in the delta (1.2–2 Hz) across conditions ($\beta = -0.353$, SE $= 0.145$, $t = -2.437$, $p_{\text{bonf}} = 0.016$).

2.3 | Interim Discussion

In this first experiment, we were interested in clarifying the relevance of the musical-matching rhythms structure on the efficacy of musical stimulation on cortical tracking to speech. In the Spanish blocks, we showed higher cortical tracking of Matching regular sequences compared to Mis-matching and Irregular sequences. In the case of speech, participants showed higher tracking of Spanish speech sequences in the delta band when the sentences were preceded by rhythmical musical cues matched to the temporal structure of the speech (Matching regular sequences). In line with our expectations, participants also showed higher tracking of Mis-matching regular sequences compared to Irregular sequences in the delta band. These results suggest that rhythm regularity was sufficient to enhance cortical tracking of speech in the Spanish blocks.

In the Basque blocks, we found higher cortical tracking of Matching Regular musical sequences in comparison to Mis-matching and Irregular sequences. Participants efficiently tracked the rhythmically different musical sequences that were presented. However, we did not find any influence of

the preceding music on the Basque speech sequences, as there was no significant difference between the three conditions. Several explanations were considered for this unexpected finding. First, a lack of the music-to-speech cortical tracking effect could be due to participants' language knowledge or dominance. However, this is unlikely because participants' language proficiency scores were high for both languages, and participants were not dominant in one language compared to the other. The possibility of participants not engaging with the Basque task compared to the Spanish task was also considered. However, participants' behavioral measures showed that participants were attentive to both tasks. Thus, we propose that the inconsistent findings for Spanish and Basque were due to the quality of the language-specific stimuli. The speech stimuli used in this experiment were carefully created, controlling for the number of syllables and stressed syllables across languages. Nevertheless, the Basque language allows for some flexibility in positioning the stress in some words due to the different dialect rules and varieties (Aurrekoetxea et al. 2013). In Basque, the unit of accentuation is not the word, as it is in Spanish, but rather the sequence or group of words in each case. Thus, pronunciation patterns in Basque vary. Moreover, the consonant-vowel chains that make up Basque words are incorporated by Basque speakers into different accent patterns, lengths, and intonations depending on their origin, age, linguistic ability, literacy level, among other factors (Amarauna 2025; Oñederra Olaizola 1998). This flexibility allowed us to create stimuli that did not differ significantly in the number of syllables and in the rhythmical patterns of the two languages. However, a later careful inspection of the Basque sentences indicated that in controlling for the position of stressed syllables in the Basque speech sequences, the naturalness of the stimuli was reduced (see Table 1 for a comparison between Experiment 1 and Experiment 2 sentences). To test the possibility that this was responsible for our failure to detect a music-to-speech cortical tracking effect in the Basque portion, we created new Basque stimuli allowing for its natural syllable stress position and conducted Experiment 2 including the same participants from Experiment 1.

3 | Experiment 2

3.1 | Method

3.1.1 | Participants

All 33 participants from Experiment 1 were invited to participate in Experiment 2, and 20 participants were available (6 male, M age: 26 years, SD = 2.36). In this subsample, all participants were highly proficient in Basque and Spanish (Maximum score: 65; Spanish $M = 64.12$, SD: 1.35; Basque $M = 58.12$, SD = 6.08) and were actively engaged with music (general musical sophistication factor, maximum score = 100; $M = 67.89$; SD = 12.61; 5 participants had more than 5 years of musical formal education). Thus, this subsample was representative of the initial sample from Experiment 1.

3.1.2 | Stimuli

The new linguistic stimuli consisted of 64 simple Basque sentences. All were 13 syllables long and were recorded by the same female Basque–Spanish bilingual speaker in natural infant-directed speech. This adjustment from 12 to 13 syllables was based on feedback from a linguist with expertise in Basque grammar and phonology. A 13-syllable structure allowed for greater naturalness in Basque, given that Basque is an agglutinative language that often requires the use of affixes and case markers, which can increase the length of words. This structure provided more flexibility for creating grammatically correct and semantically natural sentences that retained the rhythmic properties essential for the study. Each sentence was cued by a metronome to follow a regular rate of 100 bpm, reflecting the rate of stressed syllables in Basque, obtaining a regular meter with an inter-onset interval of 625 msec between stressed syllables. This 100 bpm pacing was chosen as it better matched the speaker's natural rhythm in Basque, compared to a faster rate of 120 bpm, which was perceived as more natural for Spanish. The bilingual speaker who produced the stimuli confirmed that 100 bpm allowed for a more comfortable and natural pace for Basque sentences, aligning with the rate of stressed syllables typically found in the language. Each sentence had a duration of 3750 ms. Identical to Experiment 1, three types of musical stimuli were created: Matching Regular, Mis-matching Regular, and Irregular conditions. The Matching Regular musical sequences reflected the rhythmic structure and pitch contour of the linguistic stimuli. The Mis-matching Regular musical sequences preserved the pitch contour of the linguistic stimuli but followed a rhythmical meter with notes that did not match the syllabic structure of the speech stimuli. The Irregular musical sequences were created by altering the pitch and note duration of the Matching Regular sequences and altering their regular meter.

The structure of the task was the same as in Experiment 1. In Experiment 2, the stressed syllable rate in Basque was 1.75 Hz (SD = 0.23) and 3.65 Hz (SD = 0.35) for the syllabic rate, which also influenced the frequency selection for coherence analysis. Specifically, the spectral peaks for Basque in Experiment 2 shifted to 1.6 Hz and 3.6 Hz, which are consistent with the syllable rate analysis. Spectral analyses of the new musical and speech sequences are shown in Figure 5.

3.1.3 | Procedure

The procedure was identical to Experiment 1, with the exception that participants only completed the Basque portion of the experiment.

3.1.4 | EEG Data Acquisition and Preprocessing

Data acquisition was the same as in Experiment 1. All processing and analysis steps were identical to the ones from

TABLE 1 | Stimuli comparison from Experiment 1 and Experiment 2 Basque sentences.

Experiment	Sentence (Basque)	Translation	Stress placement	Notes/comparison
1	Gaurko illargia hori horia da	Today's moon is yellow	GaurKO iLLARgiA hoRI hoRIa DA	In this example, we focus on the word **"Gaur,"** where stress falls on the final syllable of **"GaurKO,"** and the suffix **"-KO"** carries the emphasis in the compound word. Although this stress placement is typical in this context, **Experiment 1** presents an interesting case. Here, the stress falls on the final syllable of **"GaurKO,"** but in an ideal stress pattern, the stress should have been placed on **"GAURko,"**—the first syllable of the compound. This would align more closely with the typical stress placement for compound words in Basque, where stress often falls on the first element of the compound.
2	Gaurko illargitxoa hori horia da	Today's little moon is yellow	GAURko iLARgitxoA hoRI hoRIa DA	In **Experiment 2**, however, we corrected this, and the stress falls on the first syllable **"GAUR"** in **"GAURko,"** which follows the expected pattern for compound words in Basque. Besides, the introduction of the diminutive form "iLARgitxoA" in Experiment 2 likely influences the overall rhythm and stress structure, making the stress shift to the first syllable of "GAURko" rather than the final "-KO" of Experiment 1. This change demonstrates how morphological variations, like the use of diminutives, can alter stress placement and the prosodic structure within a sentence.

Note: Stressed syllables are capitalized.

Experiment 1. Given that each sentence was cued by a metronome to follow a regular rate of 100 bpm, the length of the sentences was 3.75 s. Therefore, data were segmented from −0 to 3.75 s after the onset of each trial, which gave a total of 384 trials. Since the trial length (3750 ms) led to a frequency resolution of ~0.26 Hz, we created new trials of 2500 ms with an overlapping window of 1250 ms to achieve a frequency resolution of ~0.4 Hz for the analyses. The creation of new windows gave a total of 768 trials. A minimum of 75% artifact-free trials per participant was required for inclusion in subsequent analyses (see Supporting Information, Appendix for trial exclusion summary).

3.1.5 | Coherence Analysis

Coherence was calculated following the steps from Experiment 1. Figure 6 bottom left and right show the sensors with the highest coherence values per frequency band and stimuli type. Maximum values were extracted from 1.2 to 2 Hz in the case of delta and 3.2 to 4 Hz in the case of theta. Trials that deviated by 2 SD or more from the mean were removed from subsequent analyses ($M = 15.3$, SD $= 3.2$).

3.1.6 | Statistical Analysis

All statistical analyses were identical to Experiment 1.

3.2 | Results

3.2.1 | Behavioral Data Analysis

Participants' accuracy scores were high, $M = 87.2\%$, SD $= 6.67\%$, indicating that participants were attentive to the task. An ANOVA was run to test differences between conditions. We found a main effect of condition ($F(2, 36) = 3.583$, $p = 0.038$). Post hoc comparisons revealed that participants' accuracy was higher in the Matching Regular condition compared to the Mis-matching Regular condition, ($p = 0.041$). There was no difference in accuracy between the Matching Regular condition and the Irregular condition ($p = 0.2$), nor between the Irregular and the Mismatching Regular condition ($p = 1$).

3.2.2 | EEG Analysis

Identical to the first experiment, we conducted separate analyses for each stimulus type (speech and musical sequences).

3.2.3 | Cortical Tracking of Musical Sequences

First, we assessed the cortical tracking of musical sequences. The model revealed a main effect of condition ($F(2, 81) = 53.737$, $p < 0.001$), a main effect of frequency ($F(1, 82) = 241.352$, $p < 0.001$), and a significant frequency by condition interaction ($F(2, 81) = 3.3706$, $p = 0.039$). The model output is shown in Supporting Information, Appendix, and these results are shown in Figure 6.
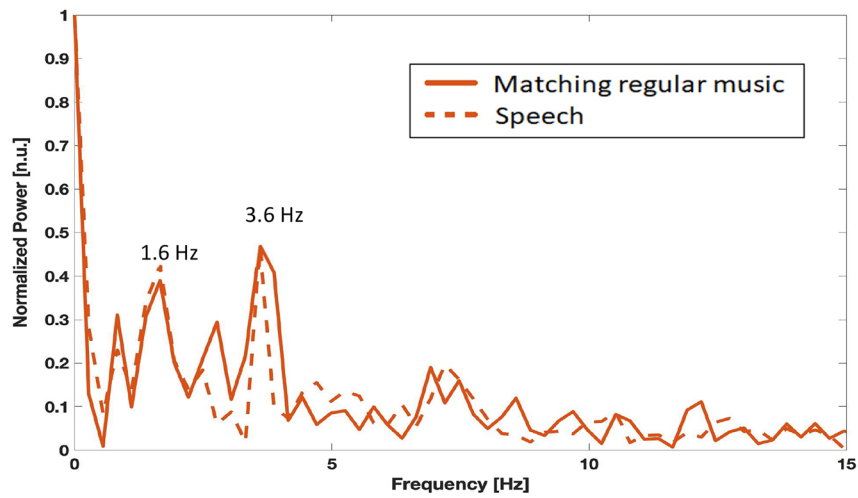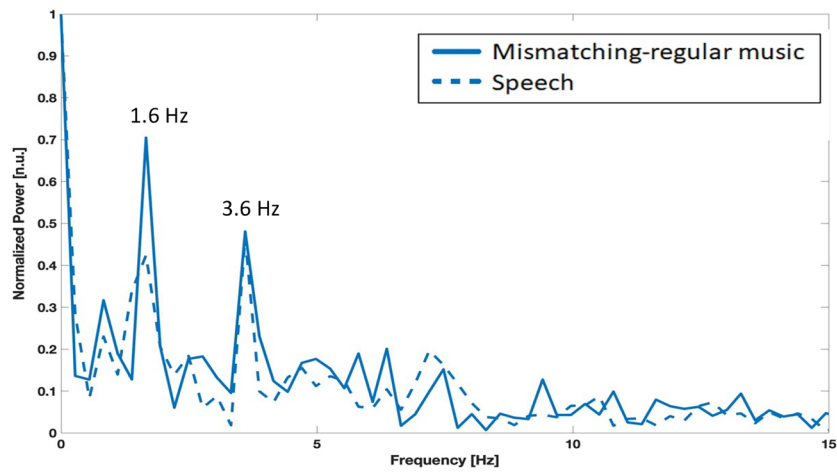
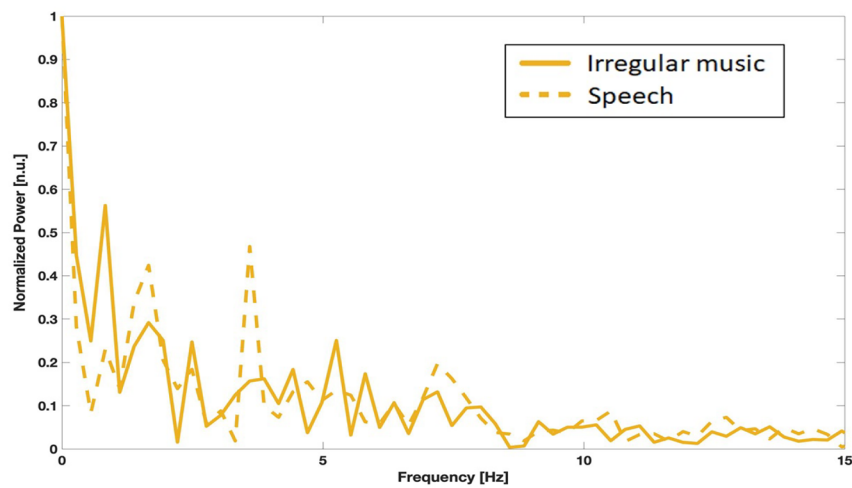**FIGURE 5** | Acoustic properties of Basque musical and speech sequences in Experiment 2. Here, we show the spectral information of speech and musical sequences in normalized power (scaled power of signal). Spectra A and B correspond to the new Basque musical stimuli, showing peaks at 1.6 and 3.6 Hz for the Matching and Mis-matching Regular condition. Spectrum C corresponds to the new Basque stimuli for the Irregular condition.

Post Hoc comparisons revealed that there was higher coherence in the Matching regular condition than in the Mis-matching Regular condition ($\beta=0.609$, SE$=0.102$, $t=5.973$, $p_{\text{bonf}} < 0.001$) and in the Irregular condition ($\beta=-1.037$, SE$=0.100$, $t=-10.322$, $p_{\text{bonf}} < 0.001$). Coherence was also significantly higher in the Mis-matching regular condition than in the Irregular condition ($\beta=-0.428$, SE$=0.101$, $t=-4.245$, $p_{\text{bonf}} < 0.001$). Moreover, there was higher coherence in theta (3.2–4 Hz) than in delta (1.2–2 Hz) ($\beta=-1.28$, SE$=0.0827$, $t=-15.528$, $p_{\text{bonf}} < 0.001$).

Post hoc comparisons conducted on the frequency by condition interaction revealed that in the case of delta (1.2–2 Hz), there was higher coherence in the Matching Regular condition than in the Mis-matching Regular ($\beta=0.717$, SE$=0.144$, $t=4.985$, $p_{\text{bonf}} < 0.001$) and in the Irregular condition ($\beta=-0.885$, SE$=0.141$, $t=-6.261$, $p_{\text{bonf}} < 0.001$). There was no difference in coherence between the Mis-matching Regular and the Irregular condition ($\beta=-0.168$, SE$=0.144$, $t=1.170$, $p_{\text{bonf}}=1$). In the case of theta (3.2 4 Hz), coherence was also significantly higher in the Matching Regular condition than in the Mis-matching Regular ($\beta=0.501$, SE$=0.145$, $t=3.466$, $p_{\text{bonf}}=0.012$) and in the Irregular condition ($\beta=-1.189$, SE$=0.142$, $t=-8.393$, $p_{\text{bonf}} < 0.001$). Coherence in the Mis-matching condition was also higher than in the Irregular condition ($\beta=-0.688$, SE$=0.142$, $t=-4.856$, $p_{\text{bonf}} < 0.001$).

3.2.4 | Cortical Tracking of Speech Sequences

Next, we assessed cortical tracking of Basque speech sequences. The model revealed a main effect of condition ($F(2, 83)=26.268$, $p < 0.001$), a main effect of frequency ($F(1, 82)=9.732$, $p=0.002$), and no significant frequency by condition interaction ($F(2, 82)=1.264$, $p=0.287$). The model output is shown in Supporting Information, Appendix, and these results are shown in Figure 6. Post hoc comparisons revealed that there was higher coherence in the Matching Regular condition than in the Mis-matching Regular condition ($\beta=0.551$, SE$=0.172$, $t=3.194$, $p_{\text{bonf}}=0.005$) and in the Irregular condition ($\beta=-1.218$ SE$=0.169$, $t=-7.228$, $p_{\text{bonf}} < 0.001$). Coherence was also significantly higher in the Mis-matching Regular condition than in the Irregular condition ($\beta=-0.668$, SE$=0.171$, $t=-3.914$, $p_{\text{bonf}} < 0.001$). Moreover, there was higher coherence in theta (3.2–4 Hz) than in delta (1.2–2 Hz) ($\beta=-0.433$, SE$=0.139$, $t=-3.118$, $p_{\text{bonf}}=0.002$).

3.2.5 | Exploratory Analyses: Experiments 1 and 2

We were interested in exploring the relationship between individuals' musical training, language proficiency, and cortical tracking of speech. For this purpose, we conducted Pearson

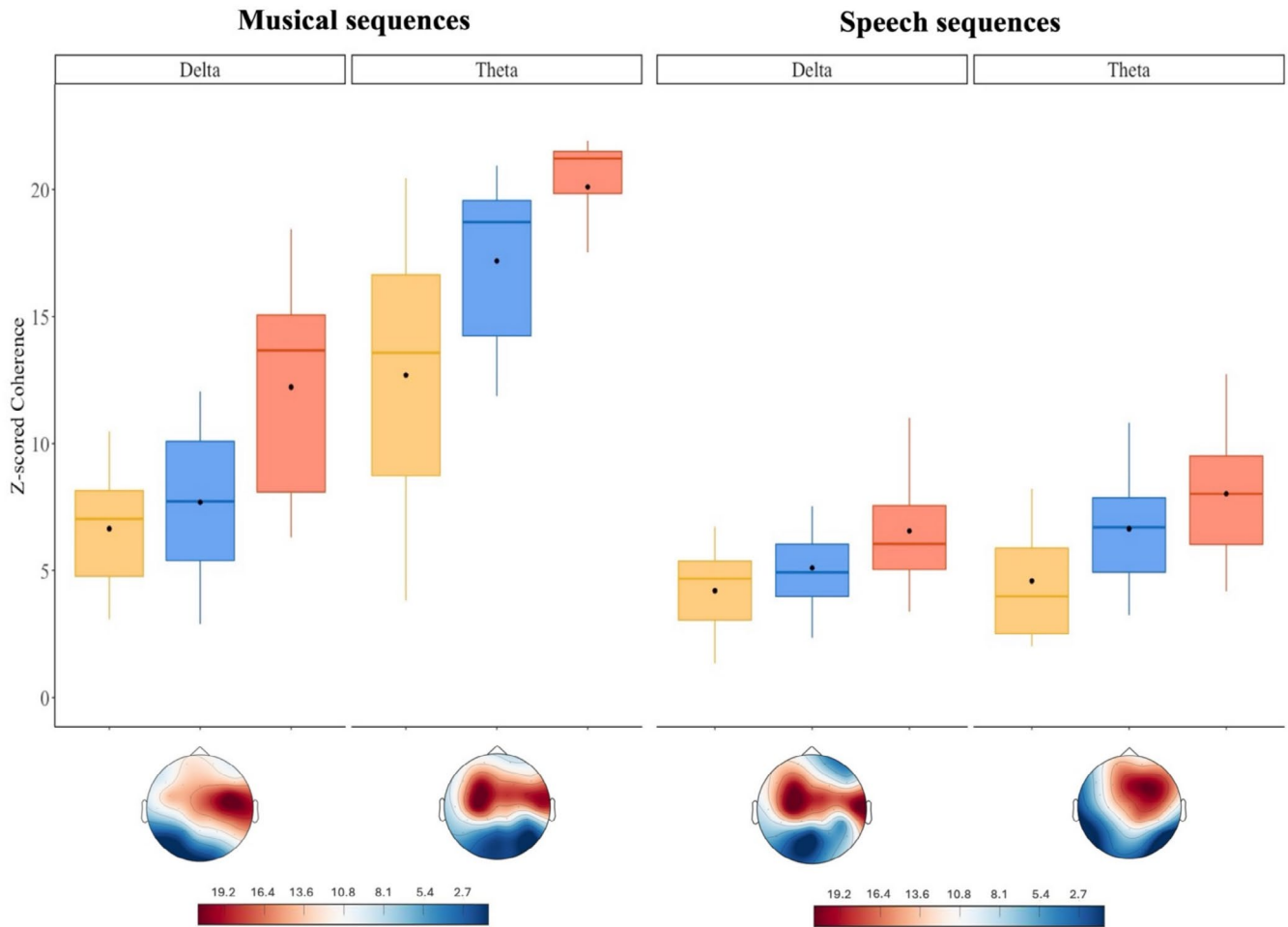**FIGURE 6** | Cortical tracking (z-scores indicating stimulus-brain coherence) of musical and speech sequences in the Basque blocks of Experiment 2. The top left and right panels show the delta and theta frequency bands for music (left) and speech (right). The bottom left and right panels show the topographies of the maximum coherence values extracted for the analysis per frequency and stimulus type.

correlation analyses between the cortical tracking values of musical sequences, speech sequences, participants' language proficiency scores, and participants' musical backgrounds. Given the unexpected findings from the Basque Language in Experiment 1, we decided to include participants' Spanish scores from Experiment 1 ($n = 33$) and participants' Basque scores from Experiment 2 ($n = 19$). These analyses are reported in Table S8 in theSupporting Information, Appendix. Contrary to our expectations, the results of the correlational analyses did not show any consistent relations between musical background and participants' cortical tracking of music and speech across experiments. Furthermore, there was no evidence of significant relations between the language proficiency scores (BEST scores) in either language and cortical tracking of speech across experiments.

3.3 | Discussion

The main aim of this study was to assess whether and how short music exposure might affect subsequent cortical tracking of speech. Specifically, we explored the precise aspects of rhythmic stimulation that lead to an improvement in speech processing through cortical tracking. Second, we assessed the effects of language-specific rhythm patterns on cortical tracking of speech in our balanced bilingual sample. To this end, we conducted two experiments where we measured participants' EEG responses to sentences in Basque and Spanish that were preceded by musical sequences that differed in their rhythmical structure.

In Experiment 1, we found that rhythmical regularities in the Matching and Mis-matching Regular musical sequences were sufficient to enhance cortical tracking of speech sequences in the Spanish language task. Moreover, we showed that when the musical sequences were matched to the meter and beat level, reflecting both the prosodic and syllabic structure of the sentences (Matching Regular), tracking was higher than in the other two conditions. Surprisingly, we did not find this enhancement for the Basque speech sequences. To further explore this discrepancy, we conducted Experiment 2 using modified Basque stimuli that better reflected the natural prosodic structure of the language, addressing limitations in the initial stimulus design. With these adjustments, Experiment 2 found that rhythmical regularities in the Matching and Mis-matching musical sequences were again sufficient to enhance cortical tracking of Basque speech, similar to the Spanish task in Experiment 1. This suggests that the Basque-specific adaptation was crucial for achieving reliable tracking results in Basque, reinforcing the importance of language-specific rhythm in the music-to-speech effect. This adaptation also raises questions about the robustness of the cortical tracking effect observed across the two experiments. While we demonstrated language-specific enhancements following short-term music exposure in both tasks, the variation in the Basque results highlights the influence of prosodic alignment in achieving consistent cortical tracking. Thus, Experiment 2 specifically provides critical insight into the language specificity of music-to-speech cortical tracking, suggesting that language prosody and syllabic timing play key roles in modulating these effects. Overall, these findings suggest that the closer the alignment between the temporal structure of the musical stimuli and the syllabic structure of the speech, the greater the improvement in cortical tracking of the speech sequences.

We will first discuss the findings on cortical tracking of musical sequences and speech sequences. Then, we will discuss the relevance of the musical matching rhythmical structures and, finally, the role of the language-specific rhythms on cortical tracking in relation to our findings.

3.3.1 | Cortical Tracking of Musical Sequences—The Speech-to-Music Effect

To disentangle the origin of the music-to-speech benefit, it was critical to design musical stimuli that varied in their rhythmicity. To do so, we created musical sequences by converting the prosodic and syllabic structure of the recorded speech to the beat and meter structure of the music (Matching Regular conditions). Besides, we then further manipulated the musical sequences to eliminate the note structure (Mis-Matching Regular condition) and then also the meter structure (Irregular condition). To understand whether this difference in rhythmicity could modulate participants' cortical tracking of the speech sequences, we first inspected participants' cortical tracking of the musical sequences. Consistent with a wealth of research showing synchronization to rhythmic stimuli (e.g., Fujioka et al. 2009, 2012; Henry and Obleser 2012; Nozaradan et al. 2011), we showed that participants' cortical tracking of musical sequences was better when they were rhythmically regular (Matching and Mis-matching Regular) than when they were rhythmically irregular. Surprisingly, despite both the Matching and the Mis-matching musical sequences having a regular meter and structure, participants tracked better the musical sequences that matched the two languages compared to the other two conditions. Based on previous evidence showing that participants effectively track natural music (Harding et al. 2019; der Nederlanden et al. 2020), we expected participants to show effective tracking of both rhythmically regular musical sequences. However, given the design of our study where participants listened to blocks of four music-speech stimulus pairs continuously, we also considered the possibility of speech affecting music. To address these bidirectional effects, we conducted an additional analysis modeling cortical coherence to music as a function of the interaction between condition (Condition) and the trial order within the block (Trial: 1–4). Our results did not show a significant effect of trial order within the block ($F(2, 702) = 0.26$, $p = 0.77$), or a significant interaction between condition and trial ($F(6, 702) = 0.98$, $p = 0.43$). This suggests that the observed coherence effects are not influenced by the directionality of the benefit (music-to-speech or speech-to-music) within a block but are instead specific to the experimental manipulation of condition. These findings support our interpretation that the differences observed specifically reflect music-to-speech benefits. However, it is important to consider that, although being both highly rhythmical, the main difference between the Matching and Mis-matching conditions was the alignment to the temporal structure of speech. Therefore, participants may better track the musical sequences that best resembled natural speech.

3.3.2 | Cortical Tracking to Speech Sequences—The Music-to-Speech Effect

One of the explanations for why non-musicians would benefit from short-term musical exposure proposes a shared processing

of the structure in language and music (Fiveash et al. 2021; Tierney and Kraus 2014). The similarities between domains make music an ideal tool to investigate whether this shared structure can be used to enhance speech processing. In this study, we created a shared metrical context between music and speech to investigate the effect of the former on the latter. We provided the first evidence that, by creating musical sequences with a metrical structure based on the prosodic structure of speech, we can use these sequences to modulate how the brain synchronizes to subsequent speech through the mechanism of cortical tracking. Although a previous study showed that rhythmic scaffolding that matched the syllabic structure of the subsequent speech led to a benefit in cortical tracking of speech (Falk et al. 2017), the specific music properties that best enhance the subsequent synchronization of brain oscillations with speech were unknown. Importantly, we also showed that participants better-tracked speech sequences when these were preceded by musical sequences that shared (1) metrical and temporal structure and/or (2) only metrical structure. Our findings suggest that, when presented with a rhythmical stimulus that resembles natural speech, brain oscillations immediately synchronize to the external frequency, supporting the idea that the shared temporal regularity found in music and speech improves synchronization to that regularity.

The present results are significant in three major respects. First, we showed that the regularity found in the rhythmic structure of music acts as a temporal guide for brain oscillations. Previous studies showed that brain oscillations phase-align with the beats of the rhythm (Poeppel and Assaneo 2020) and that temporal regularity, as found in musical rhythm, could help the brain predict when upcoming beats are coming. Our findings are consistent with the idea that, when presented with rhythmic input, neurons realign the phase of their oscillations so that when an event occurs, they are in their high excitability phase (Schroeder and Lakatos 2009). This process of alignment has been shown to enhance the neural processing of such events and has sometimes been related to predictive processing. In our case, we provided evidence that a shared temporal framework between speech and music helped maintain the temporal dynamics of the synchronization to the speech signal, even after only very short musical exposure. Our results integrate well with findings supporting the use of short rhythmic exposure and especially the use of priming paradigms to test improvements in speech perception (Cason and Schön 2012; Chern et al. 2018; Falk et al. 2017).

Second, our findings suggest that not only is the regularity in music crucial but so is also adjusting this regularity to optimally reflect the rhythmic characteristics of the language. Although our musical sequences were natural, they were created by extracting the syllabic and melodic patterns of the speech sequences. Therefore, the rhythmical patterns to which we exposed the participants were highly influenced by the speech, that is, they reflected the syllabic and prosodic patterns of each language. In this study, we showed that adjustments to the music play a crucial role in modulating cortical tracking of speech. The importance of closely matching the prosodic characteristics of the music and speech stimuli to detect an effect of short music exposure on cortical tracking of speech was corroborated by the need to modify the Basque speech sequences used in Experiment 1 to more accurately reflect the natural prosodic patterns of the language. Importantly, following this adaptation, the results for Basque in Experiment 2 were closely aligned with those for Spanish in Experiment 1.

Third, by testing Basque–Spanish bilinguals in this study, we aimed to identify the role of language-specific rhythms on the music-to-speech cortical tracking effect. Extensive research has shown efficient tracking of listeners' native languages (see Meyer 2018 for review), but little is known about the differences in cortical tracking of speech that may arise from cross-linguistic differences at the prosodic and syllabic levels. Recent evidence indicates that language-specific features that influence cortical tracking of speech are driven by the specific syllabic and prosodic rates that govern the language rhythms (Peter et al. 2022). While we found a strong effect in both languages, some cross-frequency differences arose for each language.

In Spanish, given its rhythmical characteristics (Molnar et al. 2016; Molnar et al. 2014), we expected to observe a music-to-speech effect mainly in the theta band, based on Falk et al. (2017), who found this effect only in the theta frequency band in speakers of French, a syllabic language. However, while we found similar results when comparing the Matching Regular speech sequences to the two other conditions, we also found an effect in the delta frequency band, where the Matching and the Mismatching regular speech sequences were better tracked than the irregular speech sequences.

It is important to note that our stimuli were recorded in infant-directed speech, and the speaker was cued by a metronome to follow the rate of stressed syllables. The temporal modulation structure of infant-directed speech shows greater modulation energy in the delta frequency band compared to the theta band (Leong et al. 2017). While previous evidence has shown higher cortical tracking to speech in the delta compared to the theta frequency band in Spanish monolingual speakers (Molinaro and Lizarazu 2018), it is possible that our recordings were mainly influenced by, first, the modulations found in infant-directed speech and, second, an emphasis on the prosodic rate.

In the case of Basque, our findings from Experiment 2 showed higher tracking for the Matching and the Mismatching regular speech sequences than the irregular speech sequences in both frequency bands. While few studies have focused on language-specific differences in bilinguals' cortical tracking of speech, Pérez-Navarro et al. (2024) found cortical tracking in both languages only in the delta frequency band but not the theta band on Spanish-Basque bilingual children. In their study, participants' cortical tracking of speech was modulated by their accumulated experience with the language, being cortical tracking higher for the language they had less exposure to, which was Spanish. Similarly, in a longitudinal study on Spanish-Basque bilingual children, participants showed cortical tracking to Basque speech only in the delta frequency band (Ríos-López et al. 2020). Contrary to Pérez-Navarro et al. (2024), participants' cortical tracking was not related to participants' language dominance. In our study, although we did not find any correlation between participants' language dominance and their cortical tracking measures, it is worth noting that we tested an adult sample that was highly proficient in both Spanish and Basque. It is likely, however, that participants have a greater exposure to

one language or the other, something that the BEST language proficiency test does not measure. In another study, Lizarazu et al. (2023) tested Spanish monolinguals who were attending Basque classes. They found that proficiency in Basque modulated participants' theta-gamma phase-amplitude coupling. This suggests that exposure to the rhythms of one language can influence the cortical tracking of rhythms in another language. In a tapping study, English-French bilinguals' tapping rates were measured as they tapped to English and French utterances (Lidji et al. 2011). Participants tapped more regularly to English, a stress-timed language, compared to French, a syllable-timed language, suggesting that linguistic experience with a stress-timed language can influence (behavioral) synchronization to speech rhythms. This is an important issue for future research that could inform the developmental trajectory of cortical tracking patterns in multilingual communities.

3.3.3 | Effects of Prior Musical Experience

Finally, we investigated whether musical experience was a modulating factor on the music-to-speech cortical tracking effects we reported here. In a recent study comparing cortical tracking to rhythmically similar spoken utterances and piano melodies, Harding et al. (2019) found greater cortical tracking was significantly related to musical expertise, and no intermediate effect was observed for non-musicians. This suggests that musical training, rather than the musical structure itself, predicted cortical tracking of speech. Contrary to expectations, we did not find a significant correlation between cortical tracking of musical sequences and participants' musical expertise (although, in our study, only seven participants reported more than 5 years of formal musical education). One potential explanation for this null effect is that the Gold MSI questionnaire we used targets several factors related to musical sophistication, but only one of these factors assesses active musical engagement, including participants' musical exposure. This factor includes questions about the time spent listening to music and doing other music-related activities, such as going to live concerts. This study was conducted during the COVID-19 pandemic, so scores for the active musical engagement factor might have been lowered as participants were not able to attend music-related activities. Therefore, as a follow-up test, we extracted participants' musical exposure and correlated just this factor with their cortical tracking of musical sequences. However, there was still no significant correlation. It is maybe even more remarkable that we found a robust effect of short music exposure on cortical tracking of speech, given that this effect was not influenced by participants' musical background.

3.4 | Limitations and Future Directions

Our findings raise intriguing questions regarding the effects of acoustic features of music on cortical tracking from a general perspective. In this study, we demonstrated that participants' cortical tracking of speech was enhanced by exposure to rhythmical piano music. However, one issue that was not controlled in this study was differences in the spectral density of the musical sequences (Figures 1 and 5). As can be seen in the spectral density plots, there is a higher peak in the Matching Regular

musical sequences for both Spanish and Basque in Experiment 1 in the theta band. However, this limitation does not affect our main finding of more robust music-to-speech effects in the delta frequency band. In addition, although these differences would likely have affected our analyses if we only considered the power of the oscillations, our EEG analyses comprised phase synchronization by the relative amplitude, which is less likely to have been affected.

We still do not know whether participants would also benefit from rhythmical exposure to different acoustic shapes. When considering how oscillations might synchronize to a rhythmic input, it is generally assumed that the behavior should remain consistent across different acoustic forms, whether it is a sharp pure tone, a percussion beat, or speech. However, little consideration has been paid to how such features might affect the neural ability to synchronize with them (Doelling and Assaneo 2021). Optimal cortical tracking of different speech rhythms depends on rapid neural responses to large-amplitude acoustic edges or rise times (Huss et al. 2011; Lizarazu et al. 2021), and these edges work as temporal landmarks for cortical tracking of speech. Different instruments have different attack times or rise times, which would lead to differences in their auditory processing. Although how stimulus shape affects neural synchronization and temporal perception is understudied, our results provide further support for the hypothesis that the brain synchronizes to external rhythmic input from piano music. Importantly, while our results show that the rhythmical manipulation is crucial, additional research is needed to better understand how stimulus shape affects the ease with which the brain is able to synchronize to external stimuli and whether other types of musical stimulation would be better phase-resetters for brain oscillations.

Throughout this article, we have highlighted the importance of the similarities between the speech syllabic and prosodic rhythms and the musical meter and rhythmical structure. However, it is worth noting that together with preserving a regular rhythm, the melodic contour in our study also mimicked the pitch contour of the speech sequences. A limitation of this study is that, given it is one of the main features of prosody, we did not manipulate the pitch contour for the Matching and Mismatching regular musical sequences. However, we did manipulate it for the irregular musical sequences by scrambling the pitches and the silences of the musical sequences, and thus distorting the pitch contour based on Falk et al. (2017) irregular musical sequences. Future research could address, first, whether preserving the pitch contour but distorting the rhythm of the irregular sequences could influence the music-to-speech cortical tracking effects we found. Second, whether not preserving the pitch contour of speech by adding melody to the musical sequences could result in a null music-to-speech effect.

Finally, while the consistent 2 Hz rhythm used across all stimuli reflects the natural syllabic and prosodic rates of speech in syllable-timed languages and was chosen to ensure ecological validity, we acknowledge that prolonged exposure to stimuli with similar temporal structures could theoretically influence cortical tracking measurements. However, the natural variability in acoustic features across stimuli and the alternation between speech and musical sequences likely mitigated the risk of neural adaptation or saturation at this frequency. This

approach aligns with previous studies employing similar rhythmic regularities without evidence of confounding effects (e.g., Falk et al. 2017).

4 | Conclusion

Despite the extensive evidence that musical training influences speech processing, the precise aspects that lead to these benefits have not yet been identified. Here, by creating a short-music exposure cross-linguistic paradigm, we showed that rhythm, inherent in musical signals, guides the adaptation of brain oscillations. This suggests that cortical tracking is the mechanism through which the brain benefits from music training, by adapting the temporal dynamics of the oscillatory activity to the rhythmic scaffolding of the musical signal. Moreover, we also showed that it is crucial to adjust this rhythm to optimally reflect the rhythmic characteristics of the language. These findings are framed in a cross-linguistic context, demonstrating the need to consider the specific rhythms of each language in future studies addressing the music-to-speech cortical tracking effect.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All code used for data processing and analysis is available at https://doi.org/10.6084/m9.figshare.22740353.v3. Data is available upon request to the authors. All stimuli used for Experiment 1 and Experiment 2 are available at https://osf.io/epa4n/?view_only=76177c6fe76b45fbb8eeb33e8446b048.

References

Amarauna. 2025. El Acento en Euskera. https://www.amarauna.euskadi.eus/gl/recurso/el-acento-en-euskera/0e48a234-f938-45dc-9b33-b4c247e5ff08.

Aurrekoetxea, G., I. Gaminde, and A. Iglesias. 2013. "Prosodic Variation in the Basque Language: Stress Areas." In *Current Approaches to Limits and Areas in Dialectology*, edited by E. Carrilho, C. Magro, X. Álvarez, 247–265. Cambridge Scholars Publishing.

Bánki, A., A. Brzozowska, S. Hoehl, and M. Köster. 2022. "Neural Entrainment vs. Stimulus-Tracking: A Conceptual Challenge for Rhythmic Perceptual Stimulation in Developmental Neuroscience." *Frontiers in Psychology* 13: 878984.

Bates, D. M. 2018. lme4: Mixed-Effects Modeling With R.

Besson, M., M. Barbaroux, and E. Dittinger. 2017. "Music in the Brain: Music and Language Processing." In *The Routledge Companion to Music Cognition*, edited by R. Ashley and R. Timmers, 37–48. Routledge.

Bourguignon, M., X. de Tiege, M. O. Beeck, et al. 2013. "The Pace of Prosodic Phrasing Couples the Listener's Cortex to the Reader's Voice." *Human Brain Mapping* 34, no. 2: 314–326. https://doi.org/10.1002/hbm.21442.

Cason, N., and D. Schön. 2012. "Rhythmic Priming Enhances the Phonological Processing of Speech." *Neuropsychologia* 50, no. 11: 2652–2658.

Chern, A., B. Tillmann, C. Vaughan, and R. L. Gordon. 2018. "New Evidence of a Rhythmic Priming Effect That Enhances Grammaticality Judgments in Children." *Journal of Experimental Child Psychology* 173: 371–379.

de Bruin, A., M. Carreiras, and J. A. Duñabeitia. 2017. "The BEST Dataset of Language Proficiency." *Frontiers in Psychology* 8: 522.

Degé, F., and G. Schwarzer. 2011. "The Effect of a Music Program on Phonological Awareness in Preschoolers." *Frontiers in Psychology* 2: 10340.

der Nederlanden, C. M. V. B., M. F. Joanisse, and J. A. Grahn. 2020. "Music as a Scaffold for Listening to Speech: Better Neural Phase-Locking to Song Than Speech." *NeuroImage* 214: 116767.

Ding, N., A. D. Patel, L. Chen, H. Butler, C. Luo, and D. Poeppel. 2017. "Temporal Modulations in Speech and Music." *Neuroscience & Biobehavioral Reviews* 81: 181–187.

Doelling, K. B., L. H. Arnal, O. Ghitza, and D. Poeppel. 2014. "Acoustic Landmarks Drive Delta–Theta Oscillations to Enable Speech Comprehension by Facilitating Perceptual Parsing." *NeuroImage* 85: 761–768.

Doelling, K. B., and M. F. Assaneo. 2021. "Neural Oscillations Are a Start Toward Understanding Brain Activity Rather Than the End." *PLoS Biology* 19, no. 5: e3001234.

Du, Y., and R. J. Zatorre. 2017. "Musical Training Sharpens and Bonds Ears and Tongue to Hear Speech Better." *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 51: 13579–13584.

Falk, S., C. Lanzilotti, and D. Schön. 2017. "Tuning Neural Phase Entrainment to Speech." *Journal of Cognitive Neuroscience* 29, no. 8: 1378–1389.

Fernández-Merino, L., and M. Lizarazu. 2023. Speech, Music, and Brain Coherence Analysis Package—EEG (Version 3). *Figshare*. https://doi.org/10.6084/m9.figshare.22740353.v3.

Fiveash, A., N. Bedoin, R. L. Gordon, and B. Tillmann. 2021. "Processing Rhythm in Speech and Music: Shared Mechanisms and Implications for Developmental Speech and Language Disorders." *Neuropsychology* 35, no. 8: 771–791.

Fujioka, T., L. J. Trainor, E. W. Large, and B. Ross. 2009. "Beta and Gamma Rhythms in Human Auditory Cortex During Musical Beat Processing." *Annals of the New York Academy of Sciences* 1169, no. 1: 89–92.

Fujioka, T., L. J. Trainor, E. W. Large, and B. Ross. 2012. "Internalized Timing of Isochronous Sounds Is Represented in Neuromagnetic Beta Oscillations." *Journal of Neuroscience* 32, no. 5: 1791–1802.

Gerry, D., A. Unrau, and L. J. Trainor. 2012. "Active Music Classes in Infancy Enhance Musical, Communicative and Social Development." *Developmental Science* 15, no. 3: 398–407.

Ghitza, O. 2011. "Linking Speech Perception and Neurophysiology: Speech Decoding Guided by Cascaded Oscillators Locked to the Input Rhythm." *Frontiers in Psychology* 2: 9259.

Giraud, A. L., and D. Poeppel. 2012. "Cortical Oscillations and Speech Processing: Emerging Computational Principles and Operations." *Nature Neuroscience* 15, no. 4: 511–517.

Goswami, U. 2022. "Language Acquisition and Speech Rhythm Patterns: An Auditory Neuroscience Perspective." *Royal Society Open Science* 9, no. 7: 211855.

Gross, J., N. Hoogenboom, G. Thut, et al. 2013. "Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain." *PLoS Biology* 11, no. 12: e1001752.

Habibi, A., B. R. Cahn, A. Damasio, and H. Damasio. 2016. "Neural Correlates of Accelerated Auditory Processing in Children Engaged in Music Training." *Developmental Cognitive Neuroscience* 21: 1–14.

Harding, E. E., D. Sammler, M. J. Henry, E. W. Large, and S. A. Kotz. 2019. "Cortical Tracking of Rhythm in Music and Speech." *NeuroImage* 185: 96–101.

Henry, M. J., and J. Obleser. 2012. "Frequency Modulation Entrains Slow Neural Oscillations and Optimizes Human Listening Behavior." *Proceedings of the National Academy of Sciences of the United States of America* 109, no. 49: 20095–20100.

Herholz, S. C., and R. J. Zatorre. 2012. "Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure." *Neuron* 76, no. 3: 486–502.

Huss, M., J. P. Verney, T. Fosker, N. Mead, and U. Goswami. 2011. "Music, Rhythm, Rise Time Perception and Developmental Dyslexia: Perception of Musical Meter Predicts Reading and Phonology." *Cortex* 47, no. 6: 674–689.

Jasper, H. H. 1958. "Ten-Twenty Electrode System of the International Federation." *Electroencephalography and Clinical Neurophysiology* 10: 371–375.

Kong, Y. Y., A. Somarowthu, and N. Ding. 2015. "Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech." *Journal of the Association for Research in Otolaryngology* 16, no. 6: 783–796.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. "lmerTest Package: Tests in Linear Mixed Effects Models." *Journal of Statistical Software* 82, no. 13: 1–26.

Ladányi, E., V. Persici, A. Fiveash, B. Tillmann, and R. L. Gordon. 2020. "Is Atypical Rhythm a Risk Factor for Developmental Speech and Language Disorders?" *Wiley Interdisciplinary Reviews: Cognitive Science* 11, no. 5: e1528. https://doi.org/10.1002/wcs.1528.

Lakatos, P., A. Barczak, S. A. Neymotin, et al. 2016. "Global Dynamics of Selective Attention and Its Lapses in Primary Auditory Cortex." *Nature Neuroscience* 19: 1707–1717. https://doi.org/10.1038/nn.4386.

Lakatos, P., G. Musacchia, M. N. O'Connel, A. Y. Falchier, D. C. Javitt, and C. E. Schroeder. 2013. "The Spectrotemporal Filter Mechanism of Auditory Selective Attention." *Neuron* 77: 750–761. https://doi.org/10.1016/j.neuron.2012.11.034.

Leong, V., M. Kalashnikova, D. Burnham, and U. Goswami. 2017. "The Temporal Modulation Structure of Infant-Directed Speech." *Open Mind* 1, no. 2: 78–90.

Lidji, P., C. Palmer, I. Peretz, and M. Morningstar. 2011. "Listeners Feel the Beat: Entrainment to English and French Speech Rhythms." *Psychonomic Bulletin & Review* 18: 1035–1041.

Lizarazu, M., M. Carreiras, and N. Molinaro. 2023. "Theta-Gamma Phase-Amplitude Coupling in Auditory Cortex Is Modulated by Language Proficiency." *Human Brain Mapping* 44, no. 7: 2862–2872.

Lizarazu, M., M. Lallier, M. Bourguignon, M. Carreiras, and N. Molinaro. 2021. "Impaired Neural Response to Speech Edges in Dyslexia." *Cortex* 135: 207–218.

Luo, H., and D. Poeppel. 2007. "Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex." *Neuron* 54, no. 6: 1001–1010.

Meyer, L. 2018. "The Neural Oscillations of Speech Processing and Language Comprehension: State of the Art and Emerging Mechanisms." *European Journal of Neuroscience* 48, no. 7: 2609–2621.

Molinaro, N., and M. Lizarazu. 2018. "Delta (But Not Theta)-Band Cortical Entrainment Involves Speech-Specific Processing." *European Journal of Neuroscience* 48, no. 7: 2642–2650.

Molinaro, N., M. Lizarazu, M. Lallier, M. Bourguignon, and M. Carreiras. 2016. "Out-Of-Synchrony Speech Entrainment in Developmental Dyslexia." *Human Brain Mapping* 37, no. 8: 2767–2783.

Molnar, M., M. Carreiras, and J. Gervain. 2016. "Language Dominance Shapes Non-Linguistic Rhythmic Grouping in Bilinguals." *Cognition* 152: 150–159.

Molnar, M., J. Gervain, and M. Carreiras. 2014. "Within-Rhythm Class Native Language Discrimination Abilities of Basque-Spanish Monolingual and Bilingual Infants at 3.5 Months of Age." *Infancy* 19, no. 3: 326–337.

Moreno, S., and M. Besson. 2006. "Musical Training and Language-Related Brain Electrical Activity in Children." *Psychophysiology* 43, no. 3: 287–291.

Moreno, S., C. Marques, A. Santos, M. Santos, S. L. Castro, and M. Besson. 2009. "Musical Training Influences Linguistic Abilities in 8-Year-Old Children: More Evidence for Brain Plasticity." *Cerebral Cortex* 19, no. 3: 712–723.

Müllensiefen, D., B. Gingras, J. Musil, and L. Stewart. 2014. "The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population." *PLoS One* 9, no. 2: e89642.

Musacchia, G., M. Sams, E. Skoe, and N. Kraus. 2007. "Musicians Have Enhanced Subcortical Auditory and Audiovisual Processing of Speech and Music." *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 40: 15894–15898.

Neves, L., A. I. Correia, S. L. Castro, D. Martins, and C. F. Lima. 2022. "Does Music Training Enhance Auditory and Linguistic Processing? A Systematic Review and Meta-Analysis of Behavioral and Brain Evidence." *Neuroscience & Biobehavioral Reviews* 140: 104777.

Nozaradan, S., I. Peretz, M. Missal, and A. Mouraux. 2011. "Tagging the Neuronal Entrainment to Beat and Meter." *Journal of Neuroscience* 31, no. 28: 10234–10240.

Obleser, J., and C. Kayser. 2019. "Neural Entrainment and Attentional Selection in the Listening Brain." *Trends in Cognitive Sciences* 23, no. 11: 913–926.

Oñederra Olaizola, M. L. 1998. "Prosodiaren Oraina eta Ahoskera Arauak." *Euskera Ikerketa Aldizkaria: Revista de Investigación de la Real Academia de la Lengua Vasca* 43, no. 2: 439–449.

Oostenveld, R., P. Fries, E. Maris, and J. M. Schoffelen. 2011. "FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data." *Computational Intelligence and Neuroscience* 2011: 1–9.

Patel, A. D. 2011. "Why Would Musical Training Benefit the Neural Encoding of Speech? The OPERA Hypothesis." *Frontiers in Psychology* 2: 142.

Patel, A. D. 2014. "Can Nonlinguistic Musical Training Change the Way the Brain Processes Speech? The Expanded OPERA Hypothesis." *Hearing Research* 308: 98–108.

Peelle, J. E., and M. H. Davis. 2012. "Neural Oscillations Carry Speech Rhythm Through to Comprehension." *Frontiers in Psychology* 3: 320.

Peelle, J. E., J. Gross, and M. H. Davis. 2013. "Phase-Locked Responses to Speech in Human Auditory Cortex Are Enhanced During Comprehension." *Cerebral Cortex* 23, no. 6: 1378–1387.

Peretz, I. 2006. "The Nature of Music From a Biological Perspective." *Cognition* 100, no. 1: 1–32.

Pérez-Navarro, J., A. Klimovich-Gray, M. Lizarazu, G. Piazza, N. Molinaro, and M. Lallier. 2024. "Early Language Experience Modulates the Tradeoff Between Acoustic-Temporal and Lexico-Semantic Cortical Tracking of Speech." *iScience* 27, no. 7: 110247.

Peter, V., S. van Ommen, M. Kalashnikova, R. Mazuka, T. Nazzi, and D. Burnham. 2022. "Language Specificity in Cortical Tracking of Speech Rhythm at the Mora, Syllable, and Foot Levels." *Scientific Reports* 12, no. 1: 13477.

Piro, J. M., and C. Ortiz. 2009. "The Effect of Piano Lessons on the Vocabulary and Verbal Sequencing Skills of Primary Grade Students." *Psychology of Music* 37, no. 3: 325–347.

Poeppel, D., and M. F. Assaneo. 2020. "Speech Rhythms and Their Neural Foundations." *Nature Reviews Neuroscience* 21: 322–334. https://doi.org/10.1038/s41583-020-0304-4.

Przybylski, L., N. Bedoin, S. Krifi-Papoz, et al. 2013. "Rhythmic Auditory Stimulation Influences Syntactic Processing in Children With Developmental Language Disorders." *Neuropsychology* 27, no. 1: 121–131. https://doi.org/10.1037/a0031277.

Ríos-López, P., N. Molinaro, M. Bourguignon, and M. Lallier. 2020. "Development of Neural Oscillatory Activity in Response to Speech in Children From 4 to 6 Years Old." *Developmental Science* 23, no. 6: e12947.

Ríos-López, P., M. T. Molnar, M. Lizarazu, and M. Lallier. 2017. "The Role of Slow Speech Amplitude Envelope for Speech Processing and Reading Development." *Frontiers in Psychology* 8: 276384.

Schroeder, C. E., and P. Lakatos. 2009. "Low-Frequency Neuronal Oscillations as Instruments of Sensory Selection." *Trends in Neurosciences* 32: 9–18. https://doi.org/10.1016/j.tins.2008.09.012.

Tierney, A., and N. Kraus. 2013. "The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound." *Journal of Neuroscience* 33, no. 38: 14981–14988.

Tierney, A., and N. Kraus. 2014. "Auditory-Motor Entrainment and Phonological Skills: Precise Auditory Timing Hypothesis (PATH)." *Frontiers in Human Neuroscience* 8: 949.

Tierney, A., J. Krizman, E. Skoe, K. Johnston, and N. Kraus. 2013. "High School Music Classes Enhance the Neural Processing of Speech." *Frontiers in Psychology* 4: 60636.

Tierney, A. T., J. Krizman, and N. Kraus. 2015. "Music Training Alters the Course of Adolescent Auditory Development." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 32: 10062–10067.

vanden Bosch der Nederlanden, B., M. F. Joanisse, J. A. Grahn, T. M. Snijders, and J. M. Schoffelen. 2022. "Familiarity Modulates Neural Tracking of Sung and Spoken Utterances." *NeuroImage* 252: 119049.

Varnet, L., M. C. Ortiz-Barajas, R. G. Erra, J. Gervain, and C. Lorenzi. 2017. "A Cross-Linguistic Study of Speech Modulation Spectra." *Journal of the Acoustical Society of America* 142, no. 4: 1976–1989.

Wong, P., E. Skoe, N. M. Russo, T. Dees, and N. Kraus. 2007. "Musical Experience Shapes Human Brainstem Encoding of Linguistic Pitch Patterns." *Nature Neuroscience* 10, no. 4: 420–422.

Zendel, B. R., G. L. West, S. Belleville, and I. Peretz. 2019. "Musical Training Improves the Ability to Understand Speech-In-Noise in Older Adults." *Neurobiology of Aging* 81: 102–115.

Zhao, T. C., and P. K. Kuhl. 2016. "Musical Intervention Enhances Infants' Neural Processing of Temporal Structure in Music and Speech." *Proceedings of the National Academy of Sciences of the United States of America* 113, no. 19: 5212–5217.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.