Language proficiency entails tuning cortical activity to second language speech

Mikel Lizarazu[1,2], Manuel Carreiras[1,3], Mathieu Bourguignon[1,4,5], Asier Zarraga[1], and Nicola Molinaro[1,3]

[1]BCBL, Basque center on Cognition, Brain and Language, Donostia/San Sebastian, Spain.

[2]Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL University, 75005 Paris, France.

[3]Ikerbasque, Basque Foundation for Science, Bilbao, Spain.

[4]Laboratoire de Cartographie fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium.

[5]Laboratory of neurophysiology and movement biomechanics, UNI – ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium

KEYWORDS: second language learning, brain oscillations, cortical speech tracking, top-down modulations, magnetoencephalography

Corresponding author: Mikel Lizarazu
Basque Center on Cognition, Brain and Language
Mikeletegi Pasealekua, 69
20009 Donostia, Gipuzkoa, Spain
Email: m.lizarazu@bcbl.eu phone: +34 943309300

Abstract

Cortical tracking of linguistic structures in speech, such as phrases (<3 Hz, delta band) and syllables (3-8 Hz, theta band), is known to be crucial for speech comprehension. However, it has not been established whether this effect is related to language proficiency. Here, we investigate how auditory cortical activity in second language (L2) learners tracked L2 speech. Using magnetoencephalography, we recorded brain activity from participants listening to Spanish and Basque. Participants were Spanish native (L1) speakers studying Basque (L2) at the same language center at three different levels: beginner (Grade 1), intermediate (Grade 2), and advanced (Grade 3). We found that i) both delta and theta tracking to L2 speech in the auditory cortex were related to L2 learning proficiency; ii) top-down modulations of activity in left auditory regions during L2 speech listening—by the left inferior frontal and motor regions in delta band, and by the left middle temporal regions in theta band—were also related to L2 proficiency. Altogether, these results indicate that the ability to learn a second language is related to successful cortical tracking of L2 speech and its modulation by neuronal oscillations in higher order cortical regions.

Introduction

Anyone who has ever learned a new language will have noticed that understanding a native speaker is anything but easy, especially in the initial stages of learning. This problem may, in part, relate to poor lexical or syntactic knowledge or to difficulty recognizing individual words in the speech stream, in other words, segmenting second language (L2) speech. Functional neuroimaging studies have shown that speech segmentation and comprehension rely on phase coupling between the temporal speech envelope and cortical activity, here referred to as the cortical tracking of speech (CTS) (Rose, 1992; Greenberg et al., 2003; Poeppel, 2003; Poeppel, Idsardi and Van Wassenhove, 2008; Molinaro and Lizarazu, 2018; Bourguignon et al., 2020). It is widely debated whether CTS reflects a purely rhythmic (i.e., oscillatory) neural process or, alternatively, a recurrent series of evoked responses elicited by the acoustics of the signal (for a discussion, see Haegens and Zion-Golumbic, 2018; Obleser and Kayser, 2019). However, the latter explanation cannot fully account for a series of findings (Kosem et al., 2018; Zoefel, Archer-Boyd and Davis, 2018; Zoefel, Ten Oever and Sack, 2018; Doelling et al., 2019; Forseth et al., 2020) showing that neural oscillations play a prominent role in supporting CTS.

Further, it has been found that CTS in the auditory cortex is top-down modulated by oscillatory activity in higher-order regions of the frontal lobe (Park et al., 2015; Keitel, Gross and Kayser, 2018) that form part of a perisylvian network involved in speech perception (Poeppel and Assaneo, 2020). Evidence that CTS is endogenously modulated is robust (Ding and Simon, 2012; Mesgarani and Chang, 2012; O'Sullivan et al., 2015;).

However, whether and how CTS is related to L2 acquisition remains largely unknown. Here, we investigate whether CTS in a L2 is associated with proficiency in that language. Previous research has established a link between spontaneous speech synchronization and artificial language learning (Assaneo et al., 2019; Poeppel and Assaneo, 2020). Indeed, CTS seems to play a key role in literacy acquisition and the development of normal phonological skills (Carr et al., 2014; Rios-López et al., 2017; Lallier et al., 2017, 2018; Lizarazu, Lallier and Molinaro, 2018; Lizarazu et al., 2021a, b). Even so, until now, no similar link has been established with L2 learning. Previous research has indicated that efficient auditory processing is at the root of language learning in children and adults (Mueller, Friederici and Männel, 2012). Thus, neural sensitivity to the speech envelope of a L2 is expected to increase across the learning experience and could be an important factor supporting L2 learning. Importantly, the present study has been designed to investigate if it is only basic auditory mechanisms (in the present study, CTS in auditory regions) or whether top-down modulations of CTS in the auditory cortex may also be related to L2 proficiency.

CTS is typically seen in the delta (<3 Hz) and theta (3–8 Hz) frequency bands, which correspond to the phrasal and syllabic rates of speech, respectively (Ahissar et al., 2001; Boemio et al., 2005; Luo and Poeppel, 2007; Bourguignon et al., 2013; Ding and Simon, 2013; Gross et al., 2013; Doelling et al., 2014; Molinaro and Lizarazu, 2018). According to neurocognitive models of speech processing (Giraud and Poeppel,2012; Peelle and Davis, 2012), CTS in the auditory cortex largely reflects a bottom-up analysis of the speech envelope for syllabic and phrasal information (Keitel et al., 2018; Kaufeld et al., 2020). Phonemic level processing could be more related to a phase-amplitude coupling mechanism involving gamma band activity (Lizarazu et al., 2019, see also Gross et al., 2013), however this hypothesis is still very controversial, since phoneme-level information could contribute to CTS as evidenced by more recent temporal modeling approaches (Di Liberto et al., 2015). Yet, there is increasing evidence that CTS partly builds on top-down modulations of sound processing based on temporal expectations, linguistic priors, and contextual information in a broader sense (Ahissar et al., 2001; Luo and Poeppel, 2007; Gross et al., 2013; Doelling et al., 2014; Park et al., 2015; Pérez et al., 2015; Rimmele et al., 2015; Molinaro and Lizarazu, 2018).

Here, we hypothesized that, during language learning, brain mechanisms including CTS and its top-down modulation will increase their fitting. We used magnetoencephalography (MEG) to record brain activity while participants listened to intelligible (natural) and unintelligible (spectrally rotated) speech in their L1 (Spanish) and L2 (Basque). Participants were studying the L2 at a local language center at different levels: beginner (Grade 1), intermediate (Grade 2), and advanced (Grade 3). For each language and Grade, we characterized the prominent spectral components of the CTS in the left and right auditory cortices and used effective connectivity analysis (transfer entropy) (Massey, 1990; Schreiber, 2000) to identify the top-down components driving this coupling. Based on previous studies, we hypothesized that both CTS in the auditory cortex and its top-down effects would be stronger for intelligible than unintelligible speech in both languages across all Grades. We also expected the strength of both CTS in the auditory cortex and its top-down effects during L2 speech perception to relate to L2 proficiency.

Methods

Subjects

Thirty-eight right-handed native Spanish-speaking (L1, native language) healthy subjects (Mean (*M*) age: 45.03; standard deviation (*SD*): 10.45) were included in this study. All participants had no history of neurological illness and presented normal peripheral hearing. All participants were studying Basque (L2, second language) at the same language center at different levels (Bai&By, https://www.baiby.com/en/): Grade 1 (beginners; n = 13), Grade 2 (intermediate; n=13) and Grade 3 (advanced; n=12). Participants in Grades 1, 2 and 3 had already achieved levels A1, B1, and C1 respectively in the Common European Framework of Reference for Languages (CEFR). The three Grade groups did not differ significantly in age (Grade 1: M age = 42.46, SD: 12.14; Grade 2: M age = 45.92, SD: 8.69; Grade 3: M age = 47, SD = 10.58) ($p > 0.34$) or sex (Grade 1: 5/13 Females; Grade 2: 8/13 Females; Grade 3: 6/12 Females ($p > 0.26$). They attended an online language course, performing regular exercises online, which were revised during a meeting with the tutor every week. Participants were specifically recruited in order to ensure a similar number of participants per Grade. The Basque Center on Cognition Brain and Language (BCBL) ethical committee approved the experiment (following the principles of the Declaration of Helsinki) and all participants signed an informed consent form.

Behavioral screening

L1 and L2 skills were evaluated using the Basque, English, and Spanish Test (BEST) (De Bruin, Carreiras and Duñabeitia, 2017). The BEST consists of two parts. Firstly, expressive vocabulary was assessed with a picture-naming task. The test consisted of 65 pictures corresponding to non-cognate words that had to be named in each language. Secondly, participants completed a short semi-structured oral proficiency interview in each of their three languages (Gollan et al., 2012). This 5-min interview consisted of a set of questions ranging in difficulty and requiring the participant to use different types of grammatical constructions (e.g., questions requiring different tenses). The interview was conducted and assessed by a group of linguists who were native speakers of Basque and Spanish with high proficiency in English. One linguist evaluated each participant, but a total of four linguists with previous professional experience in assessing linguistic competence took part in the process. The scoring was based on a Likert-like scale from 1 ('lowest level') to 5 ('native or native-like level'). Picture-naming and interview scores were transformed into percentage scores. Additionally, participants also reported their percentage of daily use and listening exposure to each language. For the present study, BEST scores for English were not relevant, so we only report the scores for Basque and Spanish. Given the phonological similarity between Spanish and Basque we did not include phonological awareness measures. Such a measure could, however, be crucial when studying interactions between more phonologically diverse languages.

Stimuli and procedure

Four types of speech stimuli were prepared: L1, L2, spectrally-rotated L1, and spectrally-rotated L2. The stimuli were a series of disconnected sentences. The L1 speech stimuli consisted of forty meaningful Spanish sentences ranging in duration from 7.42 to 12.65 seconds (*M*=9.9; S*D*=1.13). Similarly, the L2 speech stimuli consisted of forty meaningful Basque sentences ranging in duration from 7.24 to 14.73 seconds (*M*=9.6; *SD*=1.35). We ensured that, in both languages, the total duration of speech was 6.4 minutes. Basque sentences were direct translations of Spanish sentences that maintained similar levels of conceptual and grammatical complexity. All sentences were simple, featuring high-frequency words and relatively low grammatical complexity, so that even low-proficiency Basque learners could perform the tasks. Sentences were uttered by a Spanish-Basque bilingual female who was instructed to read each sentence as clearly and naturally as possible. We evaluated the prosodic and syllabic rates in the stimuli. Prosodic boundaries within sentences were automatically detected using a thresholding algorithm that identifies temporal edges in a continuous stimulus (Gross et al., 2013). The number of syllables per second was manually counted. L1 prosodic and syllabic rates were evident at 0.32 Hz and 5.5 Hz, respectively. L2 prosodic and syllabic rates were present at 0.29 Hz and 5.89 Hz, respectively.

The original Spanish and Basque speech was digitized at 44.1 kHz using a digital recorder (Marantz PMD670) and audio files (*.wav) were segmented using Praat software. The rotated speech was a spectral inversion of the original speech, produced by applying a customized digital implementation of the original algorithm (Blesser, 1972) to flip the spectrum of the original sentences around a center frequency of 1.5 kHz. Rotated speech has very similar temporal and spectral complexity to ordinary speech, but is not intelligible. Supplementary Figure 1 (Figure S1) shows the power spectrum of the speech envelope for each condition. The speech envelope was transformed into frequency power spectra by applying a Fast Fourier transform with a sliding window of 2 seconds. Then, power values were averaged across sentences. The maximum power difference between L1 (Spanish) and (L2) Basque in the delta (<3 Hz) and theta (3–7 Hz) bands was 0.13 dB and 0.71 dB, respectively. For L1, the maximum power difference between the speech and rotated speech in delta (<3 Hz) and theta (3–7 Hz) bands was 1.72 dB and 0.33 dB, respectively. For L2, the maximum power difference between the speech and rotated speech in delta and theta bands was 1.08 dB and 0.26, respectively. Power changes in auditory signals below 2 dB are imperceptible to the human ear under normal conditions (Gray, 2000). Indeed, power values of the speech envelope in delta and theta were statistically compared between different conditions by averaging power values for delta (<3 Hz) and theta (3 – 7 Hz) frequency bands for each sentence. Then, for each frequency band, we computed an ANOVA on the mean power values, with Language (L1 and L2) and Intelligibility (Speech and Rotated speech) as within-items factors. We didn't find a significant effect of Language in delta ($F(1,39)=1.74$, $p=0.19$, $\eta_p^2=0.04$) or theta ($F(1,39)=0.87$, $p=0.36$, $\eta_p^2=0.02$) bands, or any interaction between Language and Intelligibility factors in delta ($F(1,39)=1.83$, $p=0.18$, $\eta_p^2=0.04$) or theta ($F(1,39)=1.43$, $p=0.24$, $\eta_p^2=0.03$) bands.

During MEG recording, sentences were presented auditorily to the participants at 75-80 decibel sound pressure level. Supplementary Figure 2 (Figure S2) shows the experimental design that we followed. Each trial began with a 1 s auditory tone (at 500 Hz tone) followed by a 2 s silence before the sentence presentation. A probe word was presented auditorily 2 s after the end of each sentence and participants had to decide if that word had been present in the previous sentence or not. While listening to the sentence, participants were asked to fixate a white sticker attached to a front-facing screen that was switched off. Participants answered the question by pressing the corresponding button (Yes/No). After each response, the next trial was presented. The response hand was counterbalanced across participants and the presentation order of the sentences in each of the four different lists was pseudo-randomized. Across the experiment, the Basque version of a sentence preceded the Spanish version half of the time (and vice versa), but we ensured that translations of the same sentence were never presented in immediate succession. Participants were asked to avoid head movements and to try to blink only in between sentences. Stimuli were delivered using Presentation software (http://www.neurobs.com/).

Data acquisition

MEG data were acquired in a magnetically shielded room using the whole-scalp MEG system (Elekta-Neuromag, Helsinki, Finland; http://www.bcbl.eu/bcbl-facilitiesresources/meg/) installed at the BCBL. The system is equipped with 102 sensor triplets (each comprising a magnetometer and two orthogonal planar gradiometers) uniformly distributed around the head of the participant. The head position inside the helmet was continuously monitored using four Head Position Indicator coils. The location of each coil relative to the anatomical fiducials (nasion, left and right preauricular points) was defined with a 3D digitizer (Fastrak Polhemus, Colchester, VA, USA). This procedure is critical for the compensation of head movements in MEG data. Digitalization of the fiducials together with ~100 additional points evenly distributed over the scalp of the participant were used during subsequent data analysis to spatially align the MEG sensor coordinates with the T1 magnetic resonance brain images acquired on a 3T MRI scan (Siemens Medical System, Erlangen, Germany). MEG recordings were acquired continuously with a bandpass filter at 0.01–330 Hz and a sampling rate of 1 kHz. Eye movements were monitored with two pairs of electrodes in a bipolar montage placed on the external chanti of each eye (horizontal electrooculography (EOG)) and above and below the right eye (vertical EOG).

Data preprocessing

Signal space separation (Taulu and Kajola, 2005) was applied offline to MEG data to subtract environmental magnetic noise and correct for head movements (Maxfilter™ v2.1, Elekta Oy, Helsinki, Finland). Bad EEG channels were detected and reconstructed with an automated pipeline adapted from Bigdely-Shamlo et al. (2015). Subsequent analyses were performed using Matlab R2010 (Mathworks, Natick, MA, USA). Heartbeat and EOG artifacts were detected using Independent Component Analysis (ICA) and linearly subtracted from recordings. The ICA decomposition was performed using the Infomax algorithm implemented in Fieldtrip toolbox (Oostenveld et al., 2011). Across participants, the number of heartbeat and ocular components that were removed varied from 1 – 4 and 1 – 3 components, respectively.

Source activity estimation

Using the Neuromag tool MRIlab, the digitized points from the Polhemus were co-registered to the skin surface. Individual T1-weighted MRI images were segmented into scalp, skull, and brain components using the segmentation algorithms implemented in Freesurfer (Reuter et al., 2012). Leadfield computation was based on a three-shell volume conductor model using a 5-mm grid of sources defined on the MNI template brain (Gramfort et al., 2014). The template grid was transformed into individual headspace by a non-linear space transformation algorithm (Ashburner et al., 1997; Ashburner and Friston, 1999) implemented in Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK). The noise covariance matrix was estimated from the empty room data acquired right before bringing the subject in the MEG room. We used the noise covariance matrix to whiten the forward matrix and the data (Lütkenhöner, 1998; Lin et al., 2006). The cortical sources of the MEG signals were estimated using minimum-norm estimates (MNE) (Hämäläinen and Ilmoniemi, 1994).

For further analysis, brain signals from predefined regions of interest were selected. The regions of interest were the left- and right-hemisphere auditory cortices (Brodmann areas (BA) 41 and BA 42) (Figure S3). These regions were selected from the 3D Brodmann atlas provided with MRIcron software (available at http://www.mccauslandcenter.sc.edu/mricro/mricron).

Coherence analysis

Coherence measures the degree of phase synchronization between two signals in the frequency domain. For each participant and condition, we used coherence to quantify CTS (i.e., the coupling between the speech temporal envelope and cortical oscillations) in each auditory cortex (BA41, BA42). The envelope of the speech signals was estimated using a filter bank that models the passage of the signal through the cochlea (Glasberg and Moore, 1990; Ghitza, 2011; Kösem et al., 2016). CTS was assessed in each source in the left and right auditory cortices in the 0.5–10 Hz frequency range with 0.5 Hz frequency resolution (following Bourguignon et al., 2013 and Molinaro and Lizarazu, 2018). Then, CTS values were averaged across all sources separately for the left and right auditory cortices. For each language (L1 and L2), we calculated ssCTS as the difference between CTS for the natural and spectrally rotated speech. Resulting ssCTS values are positive when auditory cortical activity tracked natural speech more than spectrally rotated speech, and negative otherwise.

The statistical analysis of ssCTS values was performed in two steps. First, we identified the frequency bands that showed significantly positive ssCTS values for each language across all participants, and separately for each Grade. This was done with a cluster-based permutation test that statistically compared ssCTS values to zero (Maris and Oostenveld, 2007). All frequency bins were included in the test. Clusters of frequency bins showing significantly positive ssCTS values (p>0.05, dependent sample t-test) defined the frequency bands of interest. Based on previous studies (Bourguignon et al., 2013; Gross et al., 2013; Molinaro et al., 2016; Meyer and Gumbert, 2018; Molinaro and Lizarazu, 2018; Destoky et al., 2019; Vander Ghinst et al., 2016, 2019) we expected to find significant ssCTS values in delta (<3 Hz) and theta (~[3–8] Hz) frequency bands. Second, we investigated whether the ssCTS values changed depending on language, Grade, and hemisphere. The ssCTS values were averaged in the frequency bands of interest and submitted to a four-way ANOVA with Grade (Grades 1, 2, and 3) as between-subject factor and language (L1 and L2), hemisphere (left, right) and frequency band (delta, theta) as within-subject factors.

Transfer entropy analysis

We used transfer entropy analysis to identify the brain areas that exert top-down modulation (TDM) on oscillations in bilateral auditory cortices (BA41, BA42) (Park et al., 2015). Based on ssCTS results, MEG signals from each source in the brain were bandpass filtered in the delta (0.5–2 Hz) and theta (4–7 Hz) frequency bands (fourth-order Butterworth filter, forward and reverse). Then, we applied the Hilbert transform to extract the phase dynamics for each frequency band. Phase values during speech presentation were quantized in 4 bins with boundaries chosen so that bins were equally populated (i.e., bins represent quartiles of the distribution of phase values between –π and π, following Park et al., 2015 and Park, Thut, and Gross, 2020). TDM from each source in the brain (X) to reference sources (Y) was calculated as follows:

$$TE_d(X \rightarrow Y) = H(Y, Y_d) - H(Y_d) - (H(Y, Y_d, X_d) - H(Y_d, X_d)),$$

where *TE* is the transfer entropy and the subscript *d* indicates signals delayed in time. The calculation was repeated for 20 different delays, from 10 ms to 200 ms in 10 ms steps (following Park et al., 2015 and Park, Thut, and Gross, 2020). The joint entropy H of two processes *A* and *B* is calculated as follows:

$$H(A,B) =- \sum_{a,b=1}^{4} p_{A,B}(a,b) log_2 p_{A,B}(a,b),$$

where the joint distribution $p_{A,B}(a,b)$ is defined as the following sum over all N_t observations:

$$p_{A,B}(a,b) = \frac{1}{N_t} \sum \delta_a(A)\delta_b(B),$$

with $\delta_a\big(A(t)\big)$ a Kronecker delta function taking the value 1 if the binned phase value at $A(t)$is *a* and 0 otherwise.

For each source, individual TDM values were averaged across delays and reference sources in the left and the right auditory cortices separately. This analysis produced volumetric, whole-brain maps where each source value represents the strength of top-down connectivity from this source to the left and right auditory cortices, respectively. These computations were performed for each participant, condition (L1, L2, spectrally-rotated L1, spectrally-rotated L2), and frequency band (delta, theta). For each language, we calculated speech-specific TDM (ssTDM) maps as the difference between the TDM maps obtained for the natural and spectrally rotated speech. Following this procedure, we obtained eight ssTDM maps per participant, one for each possible combination of the two languages (L1, L2), the two frequency bands (delta, theta), and the two reference regions (left and right auditory cortices).

The statistical analysis of ssTDM values was performed in two steps. First, we identified the brain regions showing significantly positive ssTDM values for each language and frequency band across all participants. This was done with a cluster-based permutation test that used a t-test to compare ssTDM values to zero (Maris and Oostenveld, 2007). All sources in the brain were included in the test. Second, brain regions showing significant ($p<0.05$) positive ssTDM values in delta and theta frequency bands for L1 were defined as Regions Of Interest (ROIs). Then, for each reference region and frequency band of interest, ssTDM values were averaged in the ROIs and submitted to a two-way ANOVA with factors Grade (1, 2, and 3) and language (L1 and L2).

Results

Behavioral screening

Language proficiency in L1 (Spanish) and L2 (Basque) was assessed with BEST (Basque English and Spanish Test), a test created specifically to evaluate participants' multilingualism (De Bruin, Carreiras and Duñabeitia, 2017). BEST consists of a picture-naming task and interviews in L1 and L2 to estimate oral proficiency. Participants also reported their daily use and daily listening exposure to each language. All measures were transformed into percentages.

Figure 1 shows means and standard errors across participants for the picture-naming, interview, daily use, and daily listening scores for each language and learning level. For each behavioral score, we ran a two-way ANOVA with Language (L1 and L2) and Grade (Grades 1, 2, and 3) as factors. We observed a main effect of Language on all behavioral scores (see Table S1). L1 scores were higher than L2 scores across all measures. We also found a Language by Grade interaction on all scores (see Table S1). Overall, the data support the contention that, as L2 learning advances, proficiency and use approach L1 levels. Post-hoc tests of the behavioral scores are also summarized in Table S1.

The percentage of correct responses to the word detection task that participants performed in the MEG recording, was 88% ± 5% (mean ± SD). We found a clear interaction between Language (L1 and L2) and Grade (Grades 1, 2, and 3) ($F(1,34) = 24.5$, $p<0.001$). Indeed, participants answered similarly in Spanish across groups and showed increasing accuracy in Basque across Grades. Importantly, all groups answered well above chance, with accuracy higher than 80% on average. This indicates that in the present study participants paid attention to L2 speech stimuli and were able to recognize individual words (hence segmenting L2 speech) with more-than-acceptable accuracy.

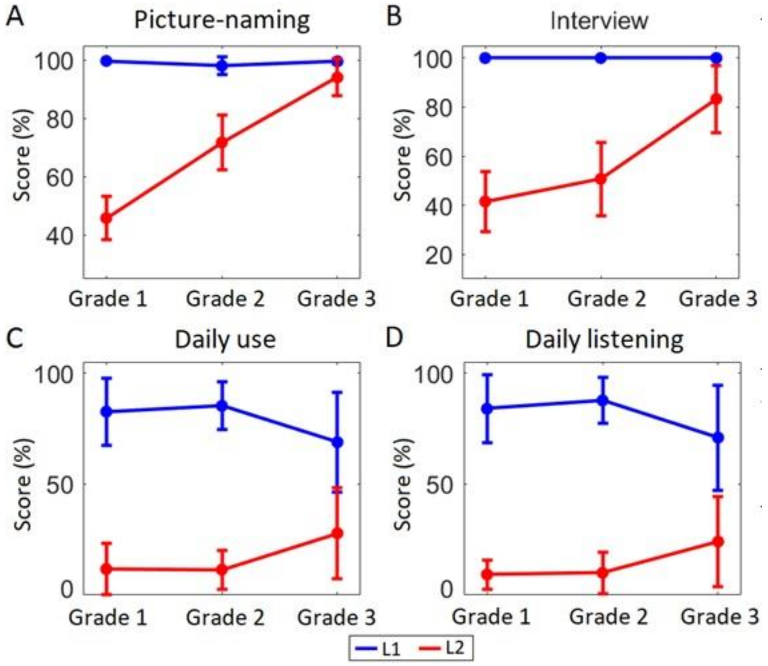


Figure 1: Language dominance. Mean and standard error of the scores (transformed into percentages) for L1 (Spanish) and L2 (Basque) measures across participants. (A) Picture-naming: The test consisted of 65 pictures corresponding to non-cognate words that had to be named in each language. (B) Interview: A personal interview with the participants to estimate their general language skill from 1 (lowest) to 5 (highest). (C) Daily use: approximate percentage of daily use of each language as reported by

participants. (D) Daily listening: approximate percentage of daily listening in each language as reported by participants.

MEG functional data

L2 learning improves cortical tracking of L2 speech in delta and theta frequency bands

We quantified CTS as coherence between the speech temporal envelope and reconstructed activity in primary auditory cortices. For each language, we calculated speech-specific CTS (ssCTS) as the difference between CTS scores obtained for natural and spectrally rotated speech. We identified the frequency bands of significant ssCTS in the left and right auditory cortices across all Grades (Figure 2A), and separately for each Grade (Figure S4). L1 ssCTS values were significantly positive ($p<0.05$, two-tailed cluster-based permutation test) in delta (0.5–1.5 Hz) and theta (3–7 Hz) frequency bands for both left and right auditory cortices. Similarly, L2 ssCTS values were significantly positive ($p<0.05$) in delta (0.5–1.5 Hz) and theta (3.5–6.5 Hz) bands in bilateral auditory cortices. Negative ssCTS were not statistically significant. Next, we used a four-way ANOVA to examine how mean ssCTS across delta (0.5–1.5 Hz) and theta (3.5–6.5 Hz) bands depends on language (L1 and L2), Grade (1, 2 and 3), hemisphere (left, right), and frequency band (delta, theta). Figure 2B & C presents the mean ssCTS values across subjects. It clearly highlights the evolution of ssCTS with Grade level for L2 but not L1, as substantiated by a Language by Grade interaction ($F(2,35)=5.32$, $p=0.01$, $\eta_p^2=0.16$) and a main effect of Language ($F(1,35)=22.26$, $p<0.01$, $\eta_p^2=0.33$). Indeed, (i) ssCTS was higher in the L1 than L2 in Grade 1 ($t(51)=6.13$, $p<0.01$) and Grade 2 ($t(51)=3.73$, $p<0.01$), but not in Grade 3 ($t(47)=0.45$, $p=0.66$), (ii) ssCTS for L2 increased with Grade (Grade 2 vs. 1, $t(51)=4.27$, $p<0.01$; Grade 3 vs. 1, $t(47)=4.77$, $p<0.01$; Grade 3 vs. 2, $t(47)=2.22$, $p=0.03$), while (iii) ssCTS for L1 did not differ between Grades ($|ts|<0.07$, $ps>0.94$). The ANOVA also revealed that ssCTS was higher in the right than in the left auditory cortex (main effect of Hemisphere, $F(1,35)=16.92$, $p<0.01$, $\eta_p^2=0.31$) and higher in the delta than in the theta band (main effect of Frequency Band, $F(1,35)=91.84$, $p<0.01$, $\eta_p^2=0.72$).

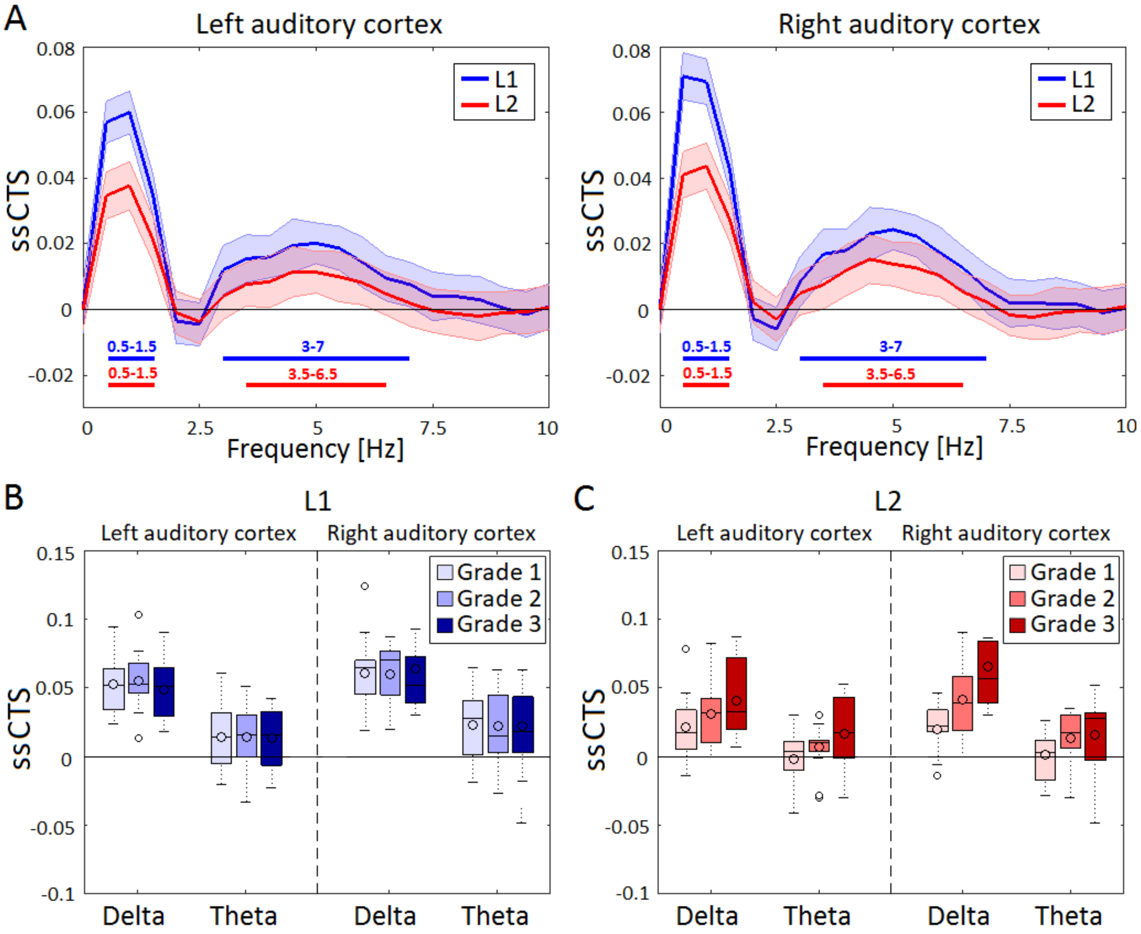


Figure 2: Speech-specific cortical tracking of speech (ssCTS) in delta and theta bands. (A) Spectrum of ssCTS for each language (L1, native language; L2, second language) and auditory cortex (left, right) across all Grades (mean indicated by solid line, standard error by shaded area). Horizontal bars indicate the frequency bands showing significantly positive (above 0) ssCTS. (B & C) Box plot showing the first quartile, median, mean (filled circles), and third quartile of ssCTS for L1 (B) and L2 (C) for each auditory cortex, frequency band (delta, theta), and Grade (1, 2, and 3). The outliers (unfilled circles) were the points that fell more than 1.5 times the interquartile range above the third quartile or below the first quartile. Extreme lines represent the highest and lowest value (excluding outliers).

L2 learning increases delta and theta top-down control of auditory cortices during L2 speech processing

We used transfer entropy to quantify top-down oscillatory modulation (TDM) of auditory cortical activity during speech processing. We focused the analysis of TDM on the phase of delta (0.5–1.5 Hz) and theta (3–7 Hz) brain oscillations since we observed that they were relevant for CTS. The TDM analysis produced volumetric, whole-brain maps in which each source value represents the strength of directional connectivity from that source to the left and right auditory cortices, respectively. For each language, we calculated the volumetric maps of speech-specific TDM (ssTDM) as the difference between the TDM values obtained for natural speech and spectrally rotated speech.

Figures 3 and S3 show brain regions with significantly positive ssTDM of left (Figure 3) and right (Figure S5) auditory cortices across all Grades. Negative ssCTS were not significant. Overall, the same network was identified in both L1 and L2 and was more extended in the left than the in the right auditory cortex. Delta-band ssTDM of the left auditory cortex (Figure 3A) involved left premotor, left motor, as well as bilateral inferior frontal and parietal regions, while the right auditory cortex involved only the right inferior frontal and parietal regions (Figure S5A). Theta-band ssTDM involved the left premotor and bilateral middle temporal regions in the left auditory cortex (Figure 3B), but middle temporal regions in the right auditory cortex (Figure S5B). In a further analysis, we defined brain regions showing significant ssTDM values in the delta and theta bands for L1 as Regions Of Interest (ROIs). Next, we used two-way ANOVAs to examine how delta (Table S2) and theta (Table S3) ssTDM values in each ROI depend on language (L1 and L2) and Grade (Grades 1, 2, and 3).

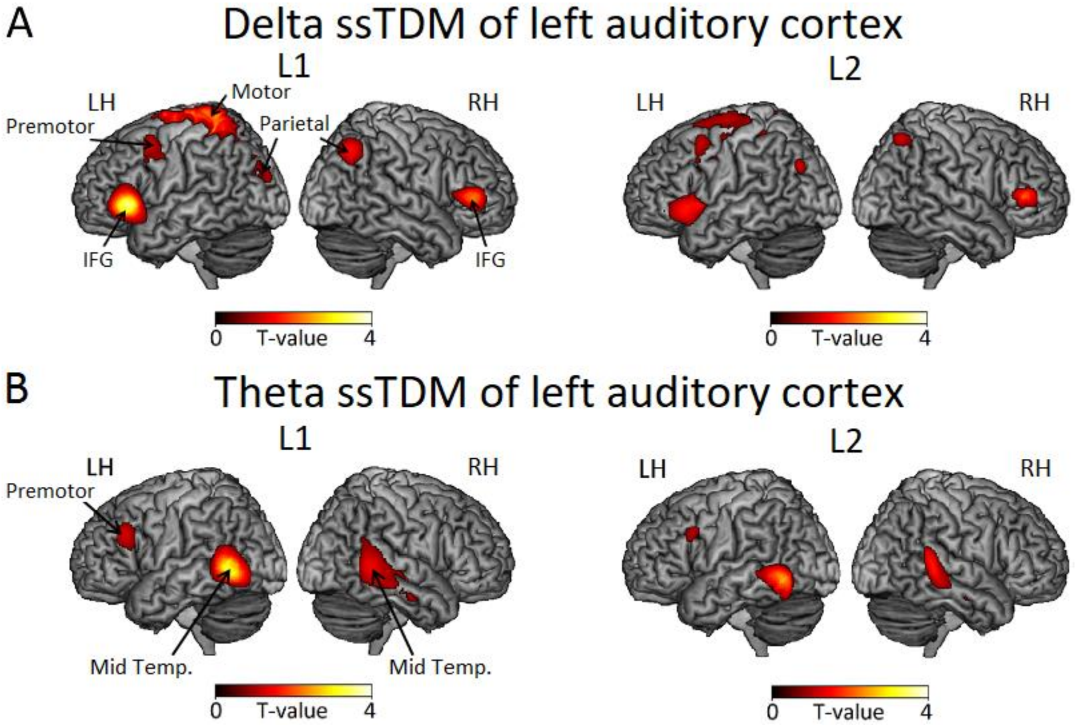


Figure 3: Volumetric maps of speech-specific top-down modulation (ssTDM) of left auditory cortical oscillations. Brain maps show regions with significantly positive ssTDM values in left auditory cortex for delta (A) and theta (B) bands for each language (L1, native language; L2, second language) and hemisphere (LH, left hemisphere; RH, right hemisphere).

Figure 4 presents ssTDM of the left auditory cortex in each ROI, averaged across subjects. Clearly, this coupling from almost all left (but not right) hemisphere ROIs increased across Grades for L2 but not L1. Indeed, (i) language by Grade interactions in ssTDM of the left auditory cortex was significant—or close to significant—for left motor regions (delta), left premotor regions (delta, theta), inferior frontal regions (delta), and left middle temporal regions (theta) (*ps*<0.06); most of these interactions were accompanied by significant main effects of language and Grade (see Tables S2 and S3 for delta and theta results). Also, within all these ROIs, (ii) ssTDM values were lower for L2 than L1 for Grade 1 (*ps*<0.01) but not for Grades 2 and 3 (*ps*>0.05), and lower for Grade 1 than Grades 2 and 3 for L2 (*ps*<0.05) but not L1 (*ps*>0.4).

The ssTDM of the right auditory cortex did not increase significantly with Grade (Figure S4 and Tables S2 & S3).

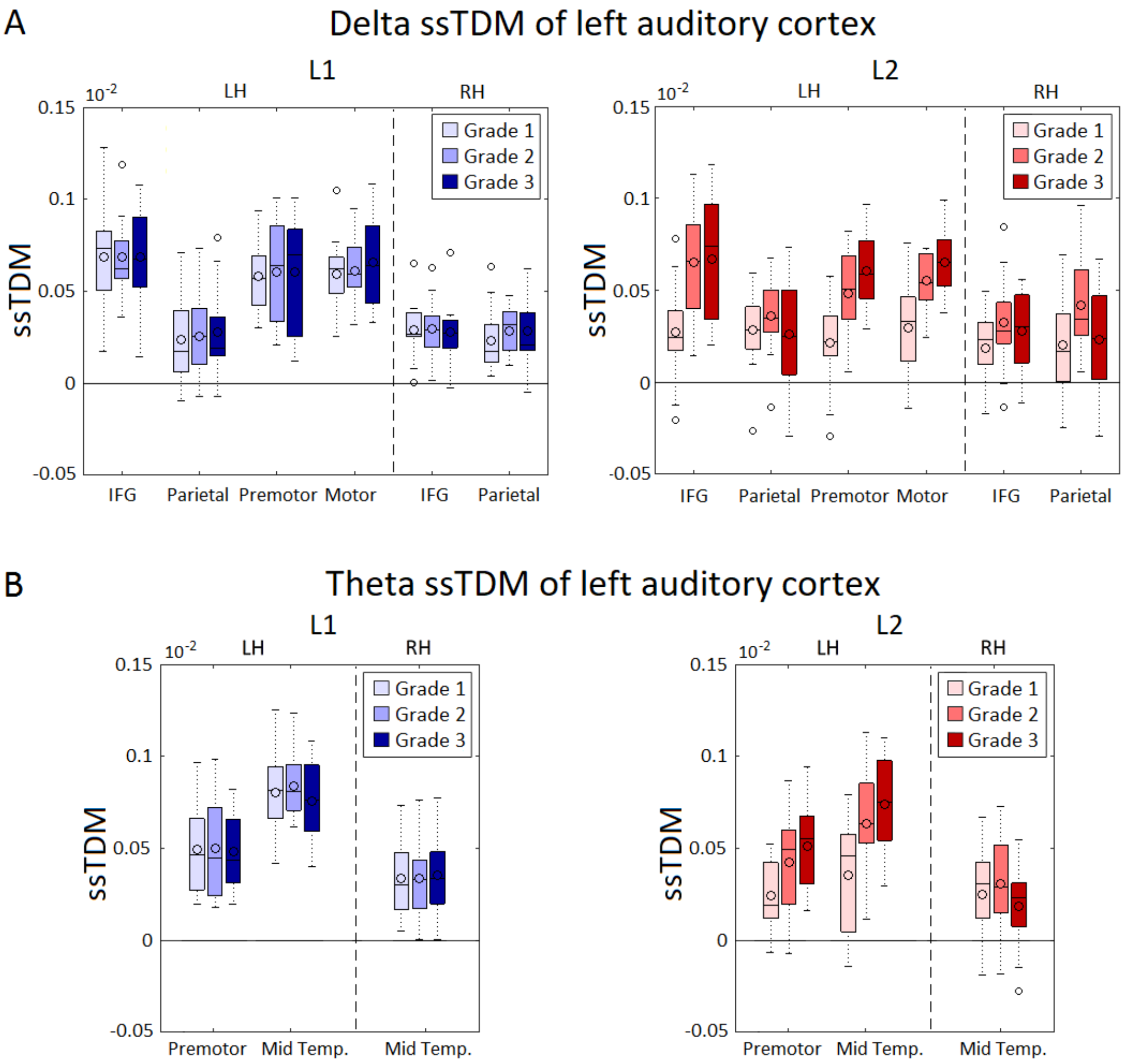


Figure 4: Delta and theta speech-specific top-down modulation (ssTDM) of left auditory cortex from each ROI. (A & B) Box plot showing the first quartile, median, mean (filled circles) and third quartile of ssTDM for delta (A) and theta (B), for each language (L1, native language; L2, second language), hemisphere (LH, left hemisphere; RH, right hemisphere), and Grade (Grades 1, 2 and 3). The outliers (unfilled circles) were the points that fell more than 1.5 times the interquartile range above the third quartile or below the first quartile. Extreme lines represent the highest and lowest value (excluding outliers).

Discussion

CTS has been identified as an ideal mechanism for the brain to segment and, eventually, comprehend speech input. Neurocognitive models assume that cortical tracking can efficiently adapt to the acoustic properties of speech and then map this information onto abstract language representations. In the present study, we provide evidence that language proficiency is functionally related to cortical tracking of speech.

The present findings thus advance our understanding of cortical speech tracking mechanisms. In the speech tracking literature, it is debated to what extent CTS reflects acoustic analysis of the speech signal or top-down modulation based on language-specific proficiency (Rimmele et al., 2018; Meyer, Sun and Martin, 2019; Obleser and Kayser, 2019). To examine this issue, numerous studies have focused on intelligibility manipulations comparing neural entrainment to speech and speech-like but unintelligible acoustic sounds. However, to date, researchers have reported contradictory results (Kösem and Van Wassenhove, 2017). Those studies that support a more direct relationship between neural entrainment and language processing have mainly relied on modifications to speech acoustics for control conditions. For instance, speech has been rendered unintelligible by changes in speed (Ahissar et al., 2001), loss of spectral complexity using noise-vocoded speech (Peelle, Gross and Davis, 2013; Rimmele et al., 2015), time-reversal (Gross et al., 2013), and the addition of background noise (Ding and Simon, 2013). Unfortunately, the alterations in speech spectrotemporal features induced by these manipulations could have affected CTS and could explain the differences reported across such studies. A qualitatively different approach that overcomes this problem is contrasting brain responses while a participant listens to distinct languages. Pérez and colleagues (2015) investigated neural oscillatory patterns associated with listening to native (Spanish), foreign (English), and unknown (French) speech by recording EEG. They found that the oscillatory power in the theta band (as well as the phase synchrony across electrodes in this band) was larger for native than both foreign and unknown speech. Their approach reduced the acoustic differences between conditions. However, (i) it is not clear how such oscillatory effects relate to speech tracking and (ii) if variations in phonetic and stress patterns across languages and speakers might explain the reported effects. Here, we made a step forward by (i) focusing on cortical speech tracking (i.e., the phase alignment between neural activity and speech) and (ii) employing speech passages in two languages that are phonologically similar (but typologically very different: Spanish is defined as a head-initial language; Basque as a head-final language), recorded by the same Spanish-Basque bilingual speaker (thus further reducing potential acoustic variations across languages). Based on these improvements, we convincingly show for the first time that cortical tracking of the speech envelope is proportionate to the language proficiency of the listener.

The second relevant finding is the identification of the network of brain regions whose oscillatory activity influences speech tracking mechanisms in the auditory cortex. Consistent with previous findings (Park et al., 2015; Keitel, Gross and Kayser, 2018), this network comprised frontal (bilateral inferior frontal and also left-lateralized premotor and motor areas) and bilateral parietal regions in the delta band. In the theta band, bilateral middle temporal regions and left premotor regions showed directional coupling with the ipsilateral auditory cortex. Especially relevant for our discussion, modulation of CTS in the auditory cortex depended on language proficiency only for a partial set of these regions. Within these, the relation between auditory speech tracking and its respective top-down modulation was significantly more relevant for the inferior frontal and the motor regions in the delta band and for the middle temporal regions in the theta band (Figure S7). Thus, despite the right-hemisphere dominance of the cortical tracking of speech, "top-down brain regions" whose activity changed at different proficiency levels emerged mainly in the left hemisphere.

The present data speak for a functional distinction between delta and theta band oscillatory cortical activity during speech listening and their relative effects on CTS. It has been suggested that delta-band speech tracking is sensitive to the phrasal time-scale of speech, and relates to word groupings making up hierarchically structured syntactic representations (Ding et al., 2016). Left inferior frontal cortical regions play a relevant role in such mechanisms and have long been associated with syntactic processing (Hagoort, 2005; Friederici, 2011). The delta band modulatory effect displayed by these regions is paralleled by the involvement of the motor cortex. This more dorsal region has been identified as part of the articulatory loop playing a top-down role in speech perception (Morillon and Baillet, 2017; Keitel, Gross and Kayser, 2018). It modulates auditory processing when external input imposes high attentional demands on the perceptual system: the more challenging stimulus comprehension becomes, the stronger the recruitment of the premotor cortex (Morillon, Schroeder and Wyart, 2014; Morillon and Baillet, 2017; Assaneo and Poeppel, 2018). Motor and premotor regions were found to play a pivotal role in predicting the timing of incoming auditory inputs, a crucial attentional component in speech processing (Morillon and Baillet, 2017; Park, Thut and Gross, 2020). Here, we show that the efficiency of this time-sensitive effect is higher if the listener is proficient in the language of the speech stream (Donhauser and Baillet, 2020; Poeppel and Assaneo, 2020), leading to higher modulation of auditory cortex speech tracking.

We also observed that middle temporal regions influence auditory activity in the theta band, i.e., auditory sensitivity to syllabic speech rate. Temporal regions support speech comprehension at the interface between phonological processing and more abstract combinatorial language processing (Hagoort, 2005; Hickok and Poeppel, 2007; Friederici, 2011). The effect reported here may reflect top-down influences based on abstract lexical representations that, in turn, affect auditory synchronization to syllabic speech structure. In other words, word-level representations may constrain cortical tracking of individual syllables, possibly in a predictive fashion. In fact, Donhauser and Baillet (2019) showed that uncertainty about a contextually predicted phoneme predicted gain increases in ongoing theta-band oscillations in the auditory cortex during sentence processing. Such modulation may be driven by more general word context, thus enhancing the estimation of the incoming syllabic unit: the more the listener is proficient in the target language, the greater the modulatory role of temporal regions on auditory tracking of the speech input.

Why then does attending L2 classes improve CTS in the L2? A tentative explanation is that the learning process enhances acquisition of L2 lexical and syntactic structures. This, in turn, could promote acquisition of information that supports segmenting continuous speech input in the second language. Top-down modulation by left IFG and left temporal regions could respectively provide the syntactic and lexical information needed to identify discrete linguistic units such as phrases and words. This would fine-tune phase alignment between the temporal structures of the speech input and the neural activity of the auditory cortex in the delta and theta bands respectively. Efficient phase alignment would gradually improve for L2 speech segmentation, boosting the overall efficiency of the neurocognitive speech comprehension system in a second language.

A possible direction for future research would be investigating what type of L2 learning best supports cortical tracking of L2 speech. In our study, we tested participants that were learning Basque in a local language academy. It is possible that L2 speech tracking could be boosted beyond the proficiency levels considered in our study if learning took place in a language immersion environment. Such environments are known to promote L2 learning. Our results do not address this question, because participants did not benefit from extensive immersion. As is evident in Figure 1, the daily Basque interactions in of our participants did not change drastically across Grades, remaining robustly higher for Spanish (the L1), across all Grades. On the other hand, L2 proficiency measures—such as the Interview and Picture Naming task—demonstrated clear incremental progress, approaching L1 proficiency levels by Grade 3. This last trend closely resembles the increases found in neural correlates for L2 speech processing (both CTS, Figure 2, and TDM, Figure 4) and is further supported by the correlations among these latter measures (reported in Figure S7). Nonetheless, immersion effects should be further evaluated.

It is important to point out that the CTS effects we report here look increasingly similar for the native and second language moving from lower to higher Grades. But this does not in itself indicate that at Grade 3 our learners have already reached native-like processing fluency in Basque. CTS reflects just one component of the neurocognitive infrastructure involved in speech comprehension. More specifically, it is a mechanism involved in segmenting continuous speech input into discrete linguistic units. Our evidence shows that this specific mechanism is gradually tuned to the temporal structure of the second language as proficiency increases and can possibly serve as the basis for better recognizing the distributional regularities of the L2, thereby supporting higher processing fluency in the new language.

Finally, it is also important to note that in the present study we used spectrally rotated speech as the control condition to obtain speech specific CTS and TDM values for both languages. There is no “ideal” control condition to use as a baseline in this kind of analysis. Previous studies have used time-reversed speech (Gross et al., 2013; Mai, Minett, and Wang, 2016), noise-vocoded speech (Peelle et al., 2013; Rimmele et al., 2015), background noise (Ding and Simon, 2013), or speech without strong acoustic edges (Doelling et al., 2014) as baselines. Interestingly, in Molinaro and Lizarazu (2018) we compared neural entrainment differences between speech and white noise and between speech and spectrally rotated speech, and showed that results were very similar. In both cases, speech elicited stronger neural entrainment than white noise and the spectrally rotated conditions in fronto-temporal brain regions. The advantage of using spectrally rotated speech as a control condition is that its temporal and spectral properties are highly similar to those of ordinary speech, but unintelligible. Considering this and based on our previous studies, we chose spectrally rotated speech as baseline; but we are confident our results would have been similar if we had chosen a different type of baseline.

Overall, the present findings highlight the complex network of regions influencing the auditory cortex and show how this oscillatory network activity is related to language learning. This is the first study to demonstrate that cortical tracking of speech modulates with L2 proficiency and to identify which brain regions control this modulation by influencing oscillatory activity in the auditory cortices. It is worth mentioning that this work highlights the role of the usual suspects, that is, left hemisphere brain regions that have long been a focus of attention in the neurobiology of language: the inferior frontal and middle temporal regions. Critically, however, we functionally distinguish between these two brain clusters, with (i) the inferior frontal cortex playing a modulatory role in the delta band, corresponding to phase alignment of the auditory cortex with phrase clusters in the speech input, and (ii) the middle temporal regions interacting with the auditory cortex in the theta band, i.e., the rate of syllabic speech structure.

Research transparency

None of the study procedures or analyses were preregistered prior to the research implementation. We have reported how we determined our sample size, all data inclusion/exclusion criteria, all manipulations, and all measures used in the study. The Matlab based scripts for spectral rotation of the speech signal, thresholding algorithm used to detect prosodic boundaries, as well as the experimental code, stimuli and transcriptions of the stimuli are archived at the following link: https://osf.io/svwb8/ (DOI 10.17605/OSF.IO/SVWB8). Legal copyright restrictions prevent public archiving of the BEST database, which can be obtained from the cited references. The conditions of our ethics approval do not permit any public archiving of data obtained in this study. Readers seeking access to the data should contact the lead author Mikel Lizarazu (m.lizarazu@bcbl.eu) or the local ethics committee at the Basque Center on Cognition Brain and Language (BCBL). Access will be granted to named individuals in accordance with ethical procedures governing the reuse of sensitive data. A data sharing agreement must be signed before access to the data will be granted.

References

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences. 98(23): 13367-13372.

Assaneo MF, Poeppel D. 2018. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. Science advances. 4(2): eaao3842.

Ashburner J, Neelin P, Collins L, Evans A, Friston K. 1997. Incorporating prior knowledge into image registration. Neuroimage. 6(4): 344-352.

Ashburner J, Friston KJ. 1999. Nonlinear spatial normalization using basis functions. Human brain mapping. 7(4): 254-266.

Assaneo MF, Ripollés P, Orpella J, Lin WM, de Diego-Balaguer R, Poeppel D. 2019. Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nature neuroscience. 22(4): 627-632.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 57(1): 289-300.

Bigdely-Shamlo N, Mullen T, Kothe C, Su KM, Robbins KA. 2015. The PREP pipeline: standardized preprocessing for large-scale EEG analysis. Frontiers in neuroinformatics. 9: 16.

Blesser B. 1972. Speech perception under conditions of spectral transformation: I. Phonetic characteristics. Journal of Speech and Hearing Research. 15(1): 5-41.

Boemio A, Fromm S, Braun A, Poeppel D. 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nature neuroscience. 8(3): 389-395.

Bourguignon M, De Tiege X, de Beeck MO, Ligot N, Paquier P, Van Bogaert P, ..., Jousmäki V. 2013. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Human brain mapping. 34(2): 314-326.

Bourguignon M, Molinaro N, Lizarazu M, Taulu S, Jousmäki V, Lallier M, ..., De Tiège X. 2020. Neocortical activity tracks the hierarchical linguistic structures of self-produced speech during reading aloud. NeuroImage. 216: 116788.

Carr KW, White-Schwoch T, Tierney AT, Strait DL, Kraus N. 2014. Beat synchronization predicts neural speech encoding and reading readiness in preschoolers. Proceedings of the National Academy of Sciences. 111(40): 14559-14564.

De Bruin A, Carreiras M, Duñabeitia JA. 2017. The BEST dataset of language proficiency. Frontiers in Psychology. 8: 522.

Destoky F, Philippe M, Bertels J, Verhasselt M, Coquelet N, Vander Ghinst M, ..., Bourguignon M. 2019. Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope. Neuroimage. 184: 201-213.

Di Liberto GM, O'Sullivan JA, Lalor EC. 2015. Low-frequency cortical entrainment to speech reflects phoneme-level processing. Current Biology. 25(19): 2457-2465.

Ding N, Simon JZ. 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences. 109(29): 11854-11859.

Ding N, Simo JZ. 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. Journal of Neuroscience. 33(13): 5728-5735.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nature neuroscience. 19(1): 158.

Doelling KB, Arnal LH, Ghitza O, Poeppel D. 2014. Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage. 85: 761-768.

Doelling KB, Assaneo MF, Bevilacqua D, Pesaran B, Poeppel D. 2019. An oscillator model better predicts cortical entrainment to music. Proceedings of the National Academy of Sciences. 116(20): 10113-10121.

Donhauser PW, Baillet S. 2020. Two distinct neural timescales for predictive speech processing. Neuron. 105(2): 385-393.

Forseth KJ, Hickok G, Rollo PS, Tandon N. 2020. Language prediction mechanisms in human auditory cortex. Nature communications. 11(1): 1-14.

Friederici AD. 2011. The brain basis of language processing: from structure to function. Physiological reviews. 91(4): 1357-1392.

Ghitza O. 2011. Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. Frontiers in psychology. 2: 130.

Giraud AL, Poeppel D. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nature neuroscience. 15(4): 511.

Glasberg BR, Moore BC. 1990. Derivation of auditory filter shapes from notched-noise data. Hearing research. 47(1-2): 103-138.

Gollan TH, Weissberger GH, Runnqvist E, Montoya RI, Cera CM. 2012. Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. Bilingualism: language and cognition. 15(3): 594-615.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, ..., Hämäläinen MS. 2014. MNE software for processing MEG and EEG data. Neuroimage. 86: 446-460.

Gray L. 2000. Properties of Sound. Journal of Perinatology. 20(8).

Greenberg S, Carvey H, Hitchcock L, Chang S. 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics. 31(3-4): 465-485.

Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S. 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS biology. 11(12).

Haegens S, Golumbic EZ. 2018. Rhythmic facilitation of sensory processing: a critical review. Neuroscience & Biobehavioral Reviews. 86: 150-165.

Hagoort P. 2005. On Broca, brain, and binding: a new framework. Trends in cognitive sciences. 9(9): 416-423.

Hämäläinen MS, Ilmoniemi RJ. 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Medical & biological engineering & computing. 32(1): 35-42.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. Nature reviews neuroscience. 8(5): 393-402.

Kaufeld G, Bosker HR, Ten Oever S, Alday PM, Meyer AS, Martin AE. 2020. Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. Journal of Neuroscience. 40(49): 9467-9475.

Keitel A, Gross J, Kayser C. 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS biology. 16(3): e2004473.

Kösem A, Basirat A, Azizi L, van Wassenhove V. 2016. High-frequency neural activity predicts word parsing in ambiguous speech streams. Journal of neurophysiology. 116(6): 2497-2512.

Kösem A, Van Wassenhove V. 2017. Distinct contributions of low-and high-frequency neural oscillations to speech comprehension. Language, Cognition and Neuroscience. 32(5): 536-544.

Kösem A, Bosker HR, Takashima A, Meyer A, Jensen O, Hagoort P. 2018. Neural entrainment determines the words we hear. Current Biology. 28(18): 2867-2875.

Lallier M, Molinaro N, Lizarazu M, Bourguignon M, Carreiras M. 2017. Amodal atypical neural oscillatory activity in dyslexia: A cross-linguistic perspective. Clinical Psychological Science: 5(2): 379-401.

Lallier M, Lizarazu M, Molinaro N, Bourguignon M, Ríos-López P, Carreiras M. 2018. From auditory rhythm processing to grapheme-to-phoneme conversion: How neural oscillations can shed light on developmental dyslexia. In Reading and Dyslexia. (pp. 147-163). Springer, Cham.

Lehongre K, Ramus F, Villiermet N, Schwartz D, Giraud AL. 2011. Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. Neuron. 72(6): 1080-1090.

Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hämäläinen MS. 2006. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage. 31(1): 160-171.

Lizarazu M, Lallier M, Molinaro N, Bourguignon M, Paz-Alonso PM, Lerma-Usabiaga G, Carreiras M. 2015. Developmental evaluation of atypical auditory sampling in dyslexia: Functional and structural evidence. Human brain mapping. 36(12): 4986-5002.

Lizarazu M, Lallier M, Molinaro N. 2018. Desincronización entre los ritmos del habla y las oscilaciones neuronales: una posible causa de los problemas fonológicos en la dislexia. Anales de la Real Academia Nacional de Medicina. 135(2): 47-51.

Lizarazu M, Lallier M, Molinaro N. 2019. Phase− amplitude coupling between theta and gamma oscillations adapts to speech rate. Annals of the New York Academy of Sciences. 1453(1): 140.

Lizarazu M, Lallier M, Bourguignon M, Carreiras M, Molinaro N. 2021. Impaired neural response to speech edges in dyslexia. Cortex. 135: 207-218.

Lizarazu M, di Covella LS, van Wassenhove V, Rivière D, Mizzi R, Lehongre K, ..., Ramus F. 2021b. Neural entrainment to speech and nonspeech in dyslexia: conceptual replication and extension of previous investigations. Cortex.

Luo H, Poeppel D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 54(6): 1001-1010.

Luo H, Poeppel D. 2012. Cortical oscillations in auditory perception and speech: evidence for two temporal windows in human auditory cortex. Frontiers in psychology. 3: 170.

Lütkenhöner B. 1998. Dipole source localization by means of maximum likelihood estimation. I. Theory and simulations. Electroencephalography and clinical neurophysiology. 106(4): 314-321.

Mai G, Minett JW, Wang WSY. 2016. Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. Neuroimage. 133: 516-528.

Marchesotti S, Nicolle J, Merlet I, Arnal LH, Donoghue JP, Giraud AL. 2020. Selective enhancement of low-gamma activity by tACS improves phonemic processing and reading accuracy in dyslexia. PLoS biology. 18(9): e3000833.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods. 164(1): 177-190.

Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. Nature. 485(7397): 233-236.

Massey J. 1990. Causality, feedback and directed mutual information. IEEE ISITA, Honolulu, HI.

Meyer L, Gumbert M. 2018. Synchronization of electrophysiological responses with speech benefits syntactic information processing. Journal of cognitive neuroscience. 30(8): 1066-1074.

Meyer L, Sun, Y, Martin AE. 2019. Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing. Language, Cognition and Neuroscience. 1-11.

Molinaro N, Lizarazu M, Lallier M, Bourguignon M, Carreiras M. 2016. Out-of-synchrony speech entrainment in developmental dyslexia. Human Brain Mapping. 37(8): 2767-2783.

Molinaro N, Lizarazu M. 2018. Delta (but not theta)-band cortical entrainment involves speech-specific processing. European Journal of Neuroscience. 48(7): 2642-2650.

Morillon B, Schroeder CE, Wyart V. 2014. Motor contributions to the temporal precision of auditory attention. Nature communications. 5(1): 1-9.

Morillon B, Baillet S. 2017. Motor origin of temporal predictions in auditory attention. Proceedings of the National Academy of Sciences. 114(42): E8913-E8921.

Mueller JL, Friederici AD, Männel C. 2012. Auditory perception at the root of language learning. Proceedings of the National Academy of Sciences. 109(39): 15953-15958.

Obleser J, Kayser C. 2019. Neural entrainment and attentional selection in the listening brain. Trends in cognitive sciences.

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational intelligence and neuroscience. 2011.

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, Shinn-Cunningham BG, ..., Lalor EC. 2015. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. Cerebral cortex. 25(7): 1697-1706.

Park H, Ince RA, Schyns PG, Thut G, Gross J. 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Current Biology. 25(12): 1649-1653.

Park H, Thut G, Gross J. 2020. Predictive entrainment of natural speech through two fronto-motor top-down channels. Language, Cognition and Neuroscience. 1-13.

Peelle JE, Davis MH. 2012. Neural oscillations carry speech rhythm through to comprehension. Frontiers in psychology. 3: 320.

Peelle JE, Gross J, Davis MH. 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral cortex. 23(6): 1378-1387.

Pérez A, Carreiras M, Dowens MG, Duñabeitia JA. 2015. Differential oscillatory encoding of foreign speech. Brain and language. 147: 51-57.

Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. Speech communication. 41(1): 245-255.

Poeppel D, Idsardi WJ, Van Wassenhove V. 2008. Speech perception at the interface of neurobiology and linguistics. Philosophical Transactions of the Royal Society B: Biological Sciences. 363(1493): 1071-1086.

Poeppel D, Assaneo MF. 2020. Speech rhythms and their neural foundations. Nature Reviews Neuroscience. 1-13.

Power AJ, Mead N, Barnes L, Goswami U. 2013. Neural entrainment to rhythmic speech in children with developmental dyslexia. Frontiers in human neuroscience. 7: 777.

Reuter M, Schmansky NJ, Rosas HD, Fischl B. 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage. 61(4): 1402-1418.

Rimmele JM, Golumbic EZ, Schröger E, Poeppel D. 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex. 68: 144-154.

Rimmele JM, Gross J, Molholm S, Keitel A. 2018. Brain oscillations in human communication. Frontiers in human neuroscience. 12: 39.

Ríos-López P, Molnar MT, Lizarazu M, Lallier M. 2017. The role of slow speech amplitude envelope for speech processing and reading development. Frontiers in psychology. 8: 1497.

Rosen S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences. 336(1278): 367-373.

Schreiber T. 2000. Measuring information transfer. Physical review letters. 85(2): 461.

Severini TA. 2000. Likelihood methods in statistics. Oxford University Press.

Taulu S, Kajola M. 2005. Presentation of electromagnetic multichannel data: the signal space separation method. Journal of Applied Physics. 97(12): 124905.

Vander Ghinst M, Bourguignon M, de Beeck MO, Wens V, Marty B, Hassid S, ..., Goldman S. 2016. Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. Journal of Neuroscience. 36(5): 1596-1606.

Vander Ghinst M, Bourguignon M, Niesen M, Wens V, Hassid S, Choufani G, ..., De Tiège X. 2019. Cortical tracking of speech-in-noise develops from childhood to adulthood. Journal of Neuroscience. 39(15): 2938-2950.

Zoefel B, Archer-Boyd A, Davis MH. 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Current Biology. 28(3): 401-408.

Zoefel B, Ten Oever S, Sack AT. 2018. The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. Frontiers in neuroscience. 12: 95.

Supplementary information

Table S1: Statistical analysis of the behavioral scores. For each test, we used a two-way ANOVA to investigate main and interaction effects of Language (L1, native language; L2, second language) and Grade (Grades 1, 2 and 3). We performed Bonferroni post-hoc comparisons to better characterize main effects and interactions observed in the ANOVA. Abbreviations: N_1, N_2 and N_3 are the numbers of participants in Grades 1, 2 and 3, respectively. F, F-value; P, p-value; η_p^2, effect size; T, t-value.

Test	N_1, N_2, N_3	ANOVA				Post-Hoc comparison			
		Effect	**F**	**p**	η_p^2	**Comparison**	**T**	**p**	η_p^2
Picture-naming	**13, 13, 12**	LANGUAGE	567	<0.01	0.65	L1 *vs.* L2:	8.47	<0.01	1.37
		GRADE	81.66	<0.01	0.82	Grade 2 *vs.* Grade1:	6.57	<0.01	1.07
						Grade 3 *vs.* Grade1:	12.78	<0.01	2.07
						Grade 3 *vs.* Grade2:	9.2	<0.01	1.03
		LANGUAGE by GRADE	135	<0.01	0.31	L1 *vs.* L2 in:			
						Grade 1:	12.75	<0.01	2.5
						Grade 2:	5.99	<0.01	1.2
						Grade 3:	6.38	<0.01	1.28
						For the L1:			
						Grade 2 *vs.* Grade1:	-1.91	0.07	-0.75
						Grade 3 *vs.* Grade1:	-0.47	0.64	-0.19
						Grade 3 *vs.* Grade2:	1.67	0.11	0.67
						For the L2:			
						Grade 2 *vs.* Grade1:	7.45	<0.01	2.92

						Grade 3 *vs.* Grade1:	16.72	<0.01	6.7
						Grade 3 *vs.* Grade2:	6.56	<0.01	2.63
Interview	**13, 13, 12**	LANGUAGE	320	<0.01	0.77	L1 *vs.* L2:	11.44	<0.01	1.86
		GRADE	29.22	<0.01	0.63	Grade 2 *vs.* Grade1:	1.65	0.32	0.27
						Grade 3 *vs.* Grade1:	7.32	<0.01	1.19
						Grade 3 *vs.* Grade2:	5.7	<0.01	0.93
		LANGUAGE by GRADE	29.22	<0.01	0.14	L1 *vs.* L2 in:			
						Grade 1:	---	---	---
						Grade 2:	---	---	---
						Grade 3:	---	---	---
						For the L1:			
						Grade 2 *vs.* Grade1:	---	---	---
						Grade 3 *vs.* Grade1:	---	---	---
						Grade 3 *vs.* Grade2:	---	---	---
						For the L2:			
						Grade 2 *vs.* Grade1:	1.65	0.11	0.65
						Grade 3 *vs.* Grade1:	7.69	<0.01	3.1
						Grade 3 *vs.* Grade2:	5.43	<0.01	2.17
Daily use	**13, 10, 12**	LANGUAGE	149.5	<0.01	0.08	L1 *vs.* L2:	11.44	<0.01	1.96
		GRADE	0.420	0.7	0.03				

		LANGUAGE by GRADE	3.38	0.05	0.04	L1 *vs.* L2 in:			
						Grade 1:	17.91	<0.01	3.82
						Grade 2:	8.44	<0.01	1.69
						Grade 3:	7.06	<0.01	1.54
						For the L1:			
						Grade 2 *vs.* Grade1:	0.61	0.55	0.26
						Grade 3 *vs.* Grade1:	-1.58	0.13	-0.63
						Grade 3 *vs.* Grade2:	-1.89	0.08	-0.83
						For the L2:			
						Grade 2 *vs.* Grade1:	0.21	0.83	0.09
						Grade 3 *vs.* Grade1:	2.38	0.03	0.95
						Grade 3 *vs.* Grade2:	1.83	0.08	0.8
Daily Listening	**13, 10, 12**	LANGUAGE	117	<0.01	0.76	L1 *vs.* L2:	10.1	<0.01	1.73
		GRADE	0.42	0.66	0.03				
		LANGUAGE by GRADE	3.48	0.04	0.05	L1 *vs.* L2 in:			
						Grade 1:	14.27	<0.01	3.04
						Grade 2:	7.3	<0.01	1.46
						Grade 3:	6.45	<0.01	1.41
						For the L1:			
						Grade 2 *vs.* Grade1:	0.43	0.68	0.18

						Grade 3 *vs.* Grade1:	-1.73	0.1	-0.69
						Grade 3 *vs.* Grade2:	-0.19	0.07	-0.84
						For the L2:			
						Grade 2 *vs.* Grade1:	0.05	0.96	0.02
						Grade 3 *vs.* Grade1:	2.34	0.03	0.94
						Grade 3 *vs.* Grade2:	2.13	0.05	094

Table S2: Statistical analysis of delta speech-specific top-down modulation (ssTDM) of left and right auditory cortices. For each region of interest, we used two-way ANOVAs to investigate main and interaction effects of Language and Grade factors on delta ssTDM values. We performed Bonferroni post-hoc comparisons to better characterize main effects and interactions observed in the ANOVAs. Abbreviations: F, F-value; P, p-value; η_p^2, effect size; T, t-value; LAC, Left Auditory Cortex; RAC, Right Auditory Cortex.

Band	**ssTDM**	**Region of Interest**	**ANOVA**				**Post-Hoc comparison**			
			Effect	**F**	**p**	η_p^2	**Comparison**	**T**	**p**	η_p^2
Delta	LAC	Left premotor	LANGUAGE	6.91	0.01	0.14	L1 *vs.* L2:	2.56	0.02	0.41
			GRADE	5.22	0.01	0.23	Grade 2 *vs.* Grade1:	2.26	0.09	0.37
							Grade 3 *vs.* Grade1:	3.12	0.01	0.51
							Grade 3 *vs.* Grade2:	0.9	0.99	0.15
			LANGUAGE by GRADE	3.08	0.06	0.13	L1 *vs.* L2 in:			
							Grade 1:	3.92	<0.01	1.09
							Grade 2:	1.13	0.28	0.31
							Grade 3:	-0.01	0.99	-0.01

							For the L1:			
							Grade 2 *vs.* Grade1:	0.33	0.75	0.09
							Grade 3 *vs.* Grade1:	0.7	0.5	0.2
							Grade 3 *vs.* Grade2:	0.02	0.99	0.01
							For the L2:			
							Grade 2 *vs.* Grade1:	2.65	0.02	0.73
							Grade 3 *vs.* Grade1:	4.51	<0.01	1.3
							Grade 3 *vs.* Grade2:	1.87	0.09	0.54
		Left motor	LANGUAGE	6.98	0.01	0.12	L1 *vs.* L2:	2.52	0.02	0.41
			GRADE	5.53	<0.01	0.24	Grade 2 *vs.* Grade1:	2.2	0.1	0.36
							Grade 3 *vs.* Grade1:	3.25	<0.01	0.53
							Grade 3 *vs.* Grade2:	1.1	0.84	0.18
			LANGUAGE by GRADE	3.97	0.03	0.16	L1 *vs.* L2 in:			
							Grade 1:	3.29	<0.01	0.91
							Grade 2:	0.94	0.37	0.26
							Grade 3:	0.06	0.95	0.02
							For the L1:			
							Grade 2 *vs.* Grade1:	0.44	0.67	0.12
							Grade 3 *vs.* Grade1:	0.89	0.69	0.26
							Grade 3 *vs.* Grade2:	0.77	0.46	0.22

							For the L2:			
							Grade 2 *vs.* Grade1:	2.92	0.01	0.81
							Grade 3 *vs.* Grade1:	4.94	<0.01	0.43
							Grade 3 *vs.* Grade2:	0.86	0.41	0.25
		Left Inferior Frontal	LANGUAGE	4.9	0.03	0.11	L1 *vs.* L2:	2.14	0.04	0.35
			GRADE	4.63	0.02	0.21	Grade 2 *vs.* Grade1:	2.59	0.04	0.42
							Grade 3 *vs.* Grade1:	2.66	0.04	0.43
							Grade 3 *vs.* Grade2:	0.12	0.99	0.02
			LANGUAGE by GRADE	3.33	0.05	0.14	L1 *vs.* L2 in:			
							Grade 1:	3.19	<0.01	0.88
							Grade 2:	0.42	0.68	0.12
							Grade 3:	0.12	0.91	0.04
							For the L1:			
							Grade 2 *vs.* Grade1:	0.06	0.96	0.02
							Grade 3 *vs.* Grade1:	-0.79	0.45	-0.23
							Grade 3 *vs.* Grade2:	-0.62	0.55	-0.18
							For the L2:			
							Grade 2 *vs.* Grade1:	3.47	<0.01	0.96
							Grade 3 *vs.* Grade1:	3.09	0.01	0.89
							Grade 3 *vs.* Grade2:	0.18	0.86	0.05

		Left Parietal	LANGUAGE	0.67	0.42	0.02	–
			GRADE	0.25	0.78	0.01	–
			LANGUAGE by GRADE	0.42	0.66	0.02	–
		Right Inferior Fontal	LANGUAGE	0.35	0.56	0.01	–
			GRADE	0.83	0.44	0.05	–
			LANGUAGE by GRADE	0.79	0.46	0.04	–
		Right Parietal	LANGUAGE	0.13	0.72	0.01	–
			GRADE	1.82	0.18	0.09	–
			LANGUAGE by GRADE	1.53	0.23	0.08	–
	RAC	Right Inferior Fontal	LANGUAGE	0.23	0.63	0.01	–
			GRADE	0.16	0.85	0.01	–
			LANGUAGE by GRADE	2.61	0.09	0.13	–

		Right Parietal	LANGUAGE	2.17	0.15	0.05	–
			GRADE	0.21	0.81	0.01	–
			LANGUAGE by GRADE	1.39	0.26	0.07	–

Table S3: Statistical analysis of theta speech-specific top-down modulation (ssTDM) of left and right auditory cortex. For each region of interest, we used two-way ANOVAs to investigate main and interaction effects of Language and Grade factors on theta ssTDM values. We performed Bonferroni post-hoc comparisons to better characterize main effects and interactions observed in the ANOVAs. Abbreviations: F, F-value; P, p-value; η_p^2, effect size; T, t-value; LAC, Left Auditory Cortex; RAC, Right Auditory Cortex.

Band	**ssTDM**	**Region of Interest**	**ANOVA**				**Post-Hoc comparison**			
			Effect	**F**	**p**	η_p^2	**Comparison**	**t**	**p**	η_p^2
Theta	LAC	Left premotor	LANGUAGE	4.1	0.05	0.09	L1 *vs.* L2:	1.99	0.05	0.32
			GRADE	1.38	0.26	0.07	–			
			LANGUAGE by GRADE	2.98	0.06	0.13	L1 *vs.* L2 in:			
							Grade 1:	4.43	<0.01	1.23
							Grade 2:	0.84	0.42	0.23
							Grade 3:	-0.34	0.74	-0.1
							For the L1:			
							Grade 2 *vs.* Grade1:	0.01	0.99	<0.01

							Grade 3 *vs.* Grade1:	-0.35	0.73	-0.1
							Grade 3 *vs.* Grade2:	-0.45	0.66	-0.13
							For the L2:			
							Grade 2 *vs.* Grade1:	2.32	0.04	0.64
							Grade 3 *vs.* Grade1:	3.04	0.01	0.88
							Grade 3 *vs.* Grade2:	1.06	0.31	0.31
		Left Inf. Temporal	LANGUAGE	11.3	<0.01	0.21	L1 *vs.* L2:	3.24	<0.01	0.53
			GRADE	4.3	0.02	0.2	Grade 2 *vs.* Grade1:	2.46	0.06	0.4
							Grade 3 *vs.* Grade1:	2.59	0.04	0.42
							Grade 3 *vs.* Grade2:	0.18	0.99	0.03
			LANGUAGE by GRADE	3.54	0.04	0.13	L1 *vs.* L2 in:			
							Grade 1:	3.33	<0.01	0.92
							Grade 2:	2.09	0.06	0.58
							Grade 3:	0.15	0.88	0.04
							For the L1:			
							Grade 2 *vs.* Grade1:	0.52	0.62	0.14
							Grade 3 *vs.* Grade1:	-0.54	0.6	-0.16
							Grade 3 *vs.* Grade2:	-1.15	0.28	-0.33
							For the L2:			

							Grade 2 *vs.* Grade1:	2.36	0.04	0.65
							Grade 3 *vs.* Grade1:	4.54	<0.01	1.31
							Grade 3 *vs.* Grade2:	0.93	0.37	0.27
		Right Inf Frontal	LANGUAGE	2.67	0.11	0.07	–			
			GRADE	0.34	0.71	0.02	–			
			LANGUAGE by GRADE	0.45	0.64	0.02	–			
	RAC	Right Inf. Temporal	LANGUAGE	0.21	0.65	0.01	–			
			GRADE	1	0.38	0.05	–			
			LANGUAGE by GRADE	0.36	0.7	0.02	–			

Figure S1: Power spectrum of the speech signals. For each condition, the power spectrum of the speech envelope was calculated using the Fourier Transform. Power values were obtained in the 0.5 to 10 Hz frequency range with 0.5 Hz frequency resolution.

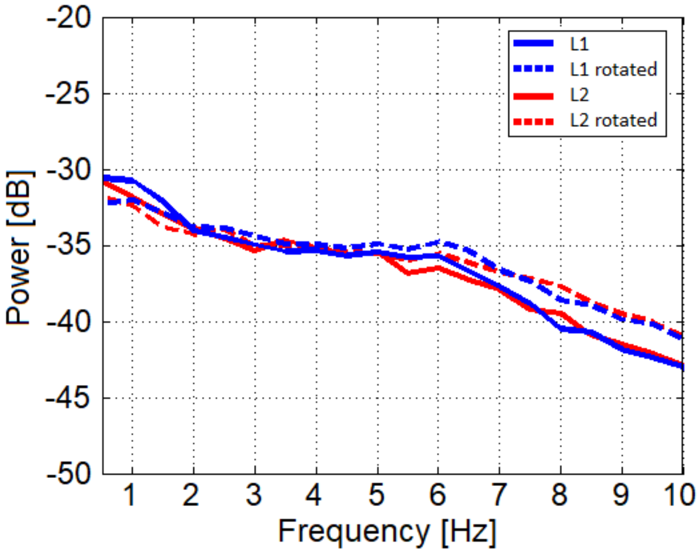


Figure S2: Experimental design. Time course of an individual trial. The identical structure was used for both L1 and L2 trials.

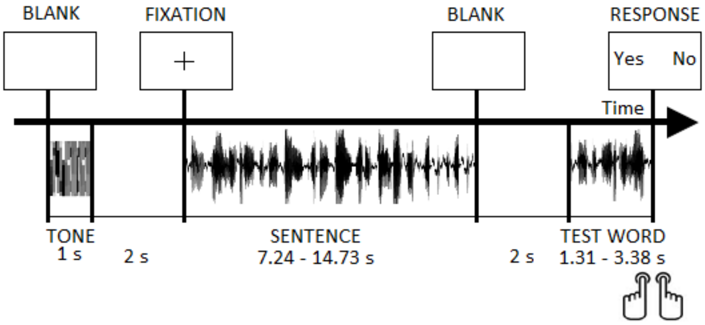


Figure S3: Selection of regions of interest (ROIs). Brodmann areas 41 (red) and 42 (blue) were selected as ROIs. The brain slice in the axial plane (Z = 12, 14, 15, 16 in MNI coordinates) illustrates the depth of the ROIs. BA41 and BA42 in the left hemisphere of the MNI brain were also included.

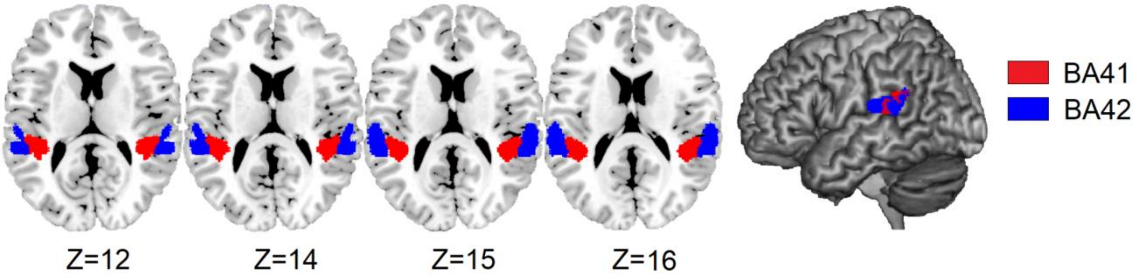


Figure S4: Speech-specific cortical tracking of speech (ssCTS) in delta and theta bands for each Grade. Spectrum of ssCTS for each language (L1, native language; L2, second language) and auditory cortex (left and right) for (A) Grade 1, (B) Grade 2 and (C) Grade 3 (mean indicated by solid line, standard error by shaded area). Horizontal bars indicate the frequency bands showing significantly positive (above 0) ssCTS.

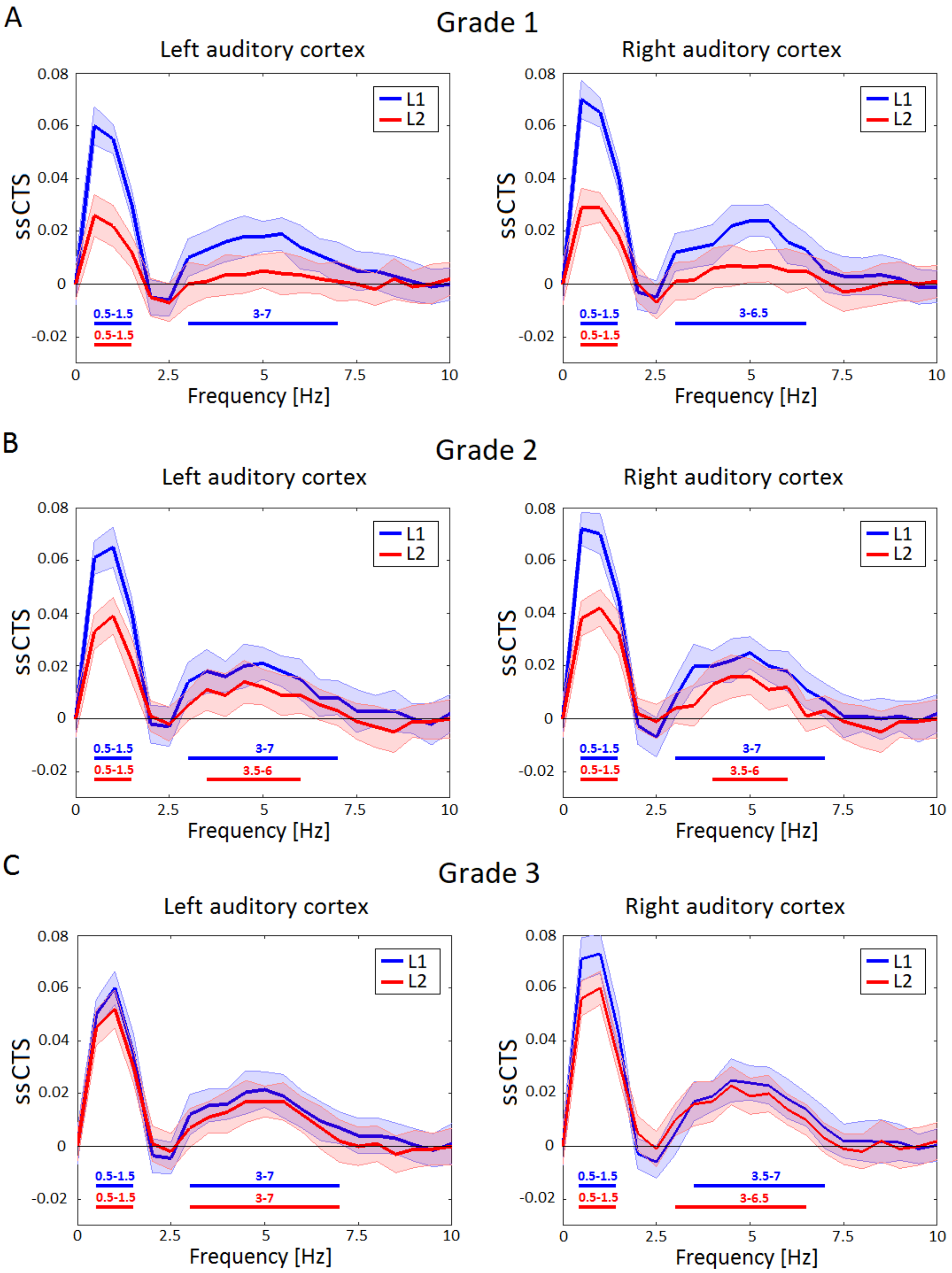


Figure S5: Volumetric maps of speech-specific top-down modulation (ssTDM) of right auditory oscillations. Brain maps show the regions with significantly positive ssTDM in right auditory cortex for delta (A) and theta (B) bands for each language (L1, native language; L2, second language) and hemisphere (LH, left hemisphere; RH, right hemisphere).

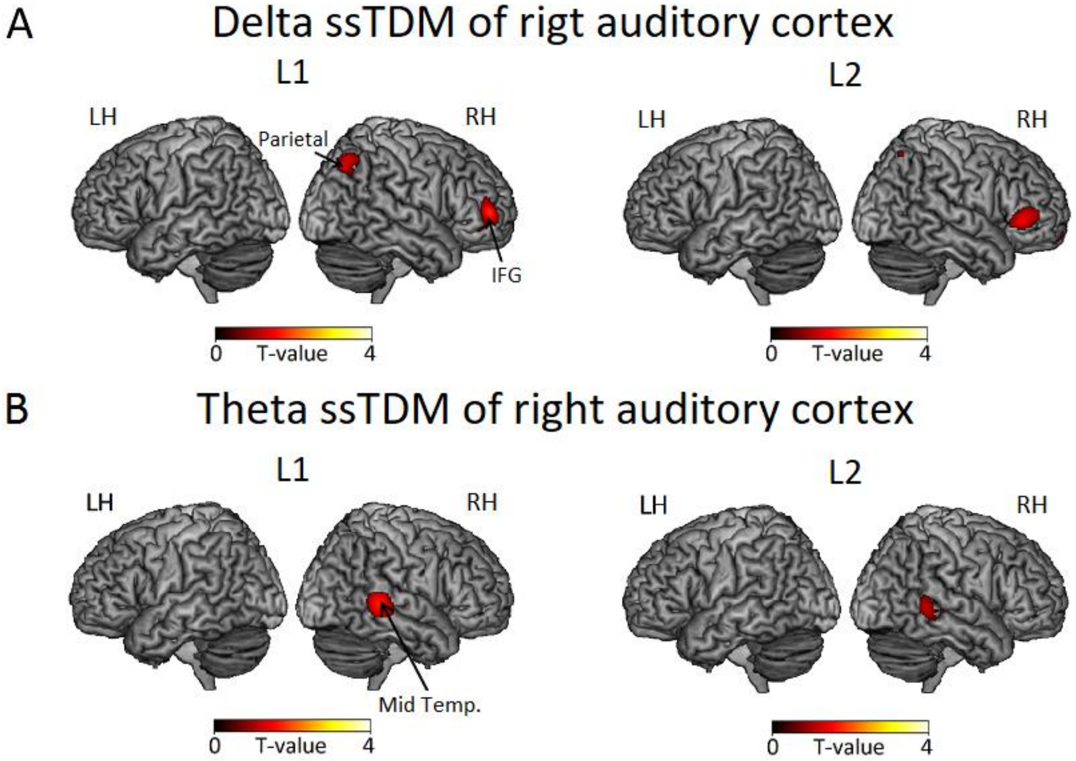
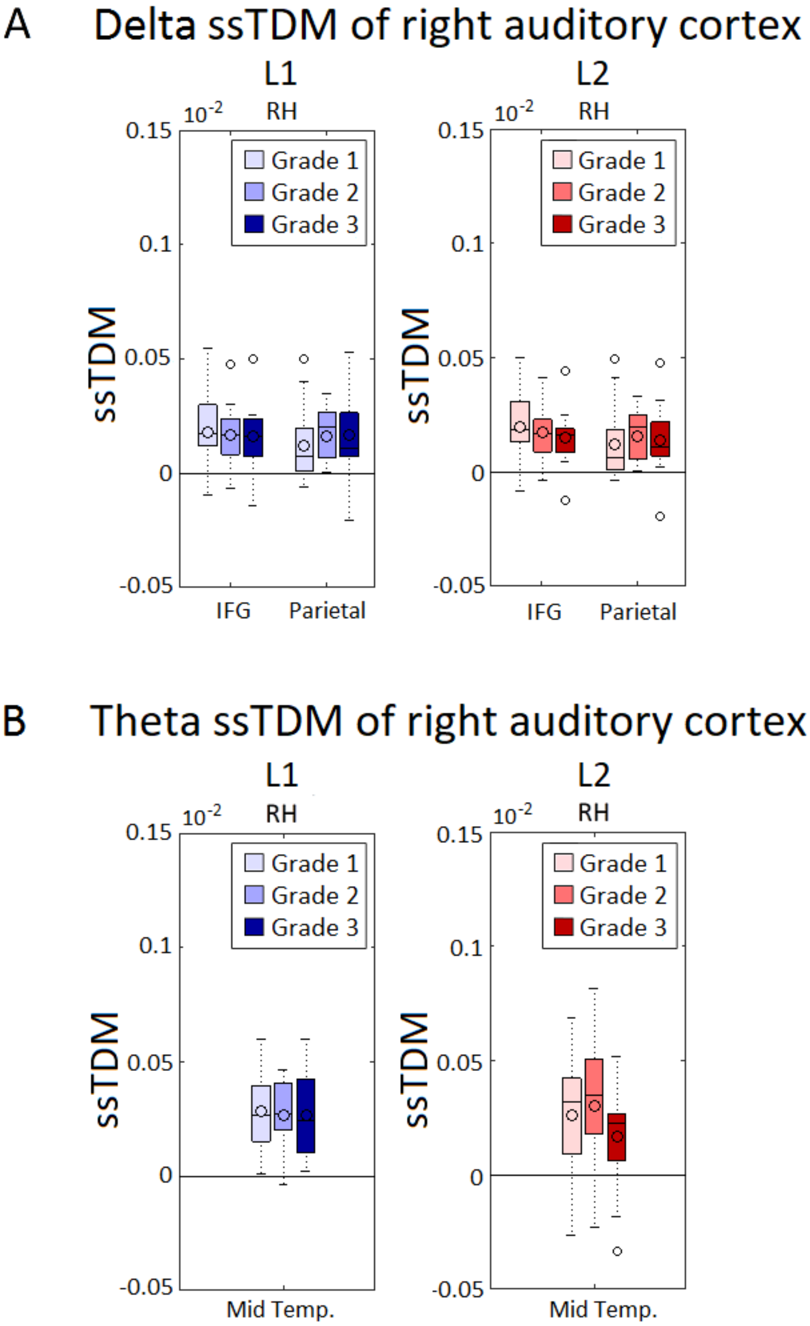


Figure S6: Delta and theta speech-specific top-down modulation (ssTDM) of right auditory cortex from each ROI. (A & B) Box plot showing the first quartile, median, mean (filled circles) and third quartile of ssTDM for delta (A) and theta (B), for each language (L1, native language; L2, second language), for each hemisphere (LH, left hemisphere; RH, right hemisphere) and learning Grade (Grades 1, 2 and 3). The outliers (unfilled circles) were the points that fell more than 1.5 times the interquartile range above the third quartile or below the first quartile. Extreme lines represent the highest and lowest values (excluding outliers).



Correlation analysis

We used linear regression models to evaluate the relationship between behavioral measures (percentage scores of picture-naming, interviews, daily use, daily listening), delta and theta ssCTS, and delta and theta ssTDM in the left auditory cortex for L2 across grades. Furthermore, we used multiple regression models to additionally account for the Grade factor in the relationship between behavioral and brain functional measures. In both models, measures were first centered around their mean. We used likelihood ratio tests (Severini, 2000) to statistically compare linear regression and linear mixed-models. We applied the False Discovery Rate procedure to correct p-values for multiple testing (Benjamini and Hochberg ,1995).

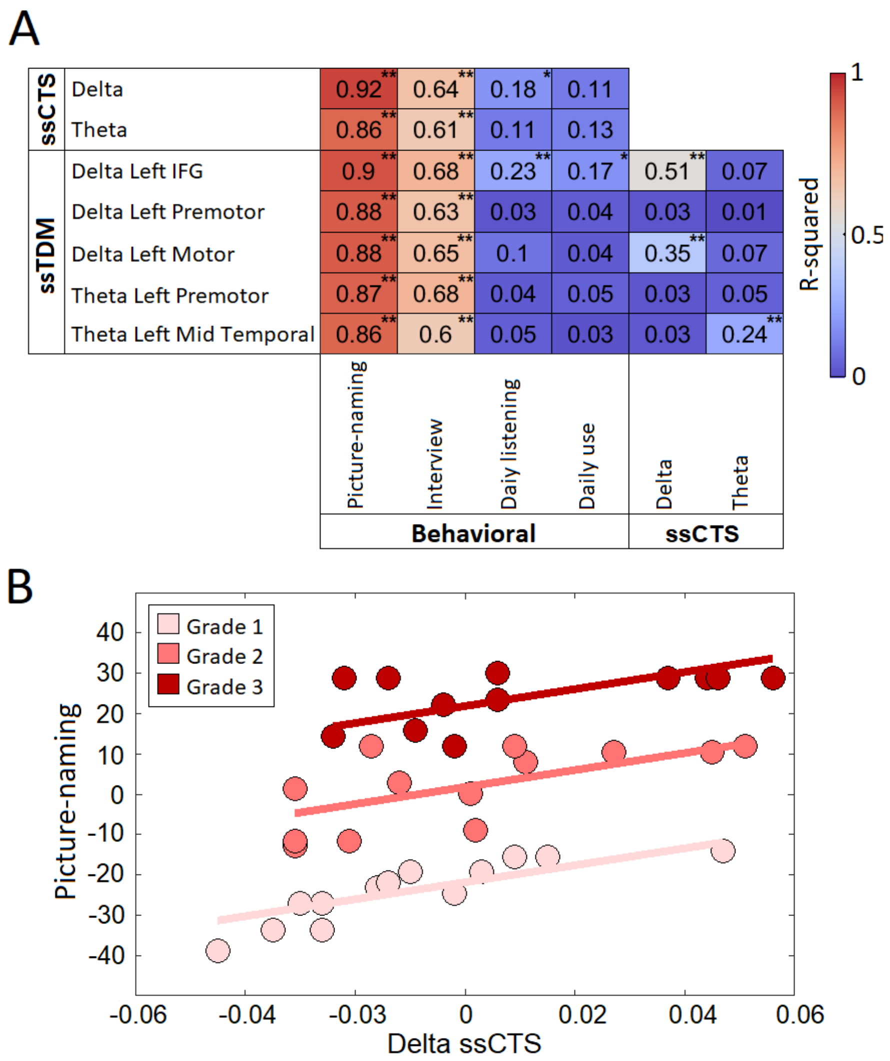
Speech-auditory cortex coupling correlates with L2 proficiency

Our analyses highlighted that L2 ssCTS and ssTDM in several left-hemisphere brain regions increased with learning Grade, that is, with L2 proficiency. We used multiple linear regression models to further explore the hypothesis that participants showing higher neural indices were also more proficient in L2 (Figure S7A[1]). Higher neural indices were associated with higher scores in picture-naming (Figure S7B) and interview scores. This evidence strongly suggests a link between neural indices for neural entrainment with L2 speech and L2 proficiency. This analysis also revealed some significant relations between neural indices (Figure S7A): delta ssCTS subject variability was related to delta ssTDM in both the left IFG and left motor regions; theta ssCTS was significantly related to theta ssTDM in left mid-temporal regions.

---

[1] Despite the relation between Grade and BEST scores, it should be noted that the variance inflation factor was not significantly high (VIF = 7.01).

Figure S7: Relevance of brain coupling measures for L2 proficiency. (A) Multiple linear regression model analyses. Squares of the correlation values (R-square) between behavioral indicators of L2 proficiency, speech-specific cortical tracking of speech (*ssCTS*) for L2, and speech-specific top-down modulation (*ssTDM*) of left auditory cortex for L2. Significant correlations (FDR-corrected) are highlighted with asterisks (*, $p < 0.05$; **, $p < 0.01$). (B) A representative correlation. Circles represent the values obtained for each participant. The black trace is the fit of the linear regression model, colored traces represent the fit of the linear mixed model.



		Behavioral				ssCTS	
		Picture-naming	Interview	Daily listening	Daily use	Delta	Theta
ssCTS	Delta	0.92**	0.64**	0.18*	0.11		
	Theta	0.86**	0.61**	0.11	0.13		
ssTDM	Delta Left IFG	0.9**	0.68**	0.23**	0.17*	0.51**	0.07
	Delta Left Premotor	0.88**	0.63**	0.03	0.04	0.03	0.01
	Delta Left Motor	0.88**	0.65**	0.1	0.04	0.35**	0.07
	Theta Left Premotor	0.87**	0.68**	0.04	0.05	0.03	0.05
	Theta Left Mid Temporal	0.86**	0.6**	0.05	0.03	0.03	0.24**

In the present study, we thus observed a relation between CTS and language proficiency in the form of a cortical network (involving left perisylvian regions) whose oscillatory activity was increasingly well synchronized to L2 speech as learning improved. However, given the relatively low number of participants considered, this correlational evidence should be further explored either with a higher number of participants or in a longitudinal study.