Understanding Requires Tracking: Noise and Knowledge Interact in Bilingual Comprehension

Esti Blanco-Elorrieta[1,2], Nai Ding[3], Liina Pylkkänen[1,2], and David Poeppel[1,4]

Abstract

■ Understanding speech in noise is a fundamental challenge for speech comprehension. This perceptual demand is amplified in a second language: It is a common experience in bars, train stations, and other noisy environments that degraded signal quality severely compromises second language comprehension. Through a novel design, paired with a carefully selected participant profile, we independently assessed signal-driven and knowledge-driven contributions to the brain bases of first versus second language processing. We were able to dissociate the neural processes driven by the speech signal from the processes that come from speakers' knowledge of their first versus second languages. The neurophysiological data show that, in combination with impaired access to top–down linguistic information in the second language, the locus of bilinguals' difficulty in understanding second language speech in noisy conditions arises from a failure to successfully perform a basic, low-level process: cortical entrainment to speech signals above the syllabic level. ■

INTRODUCTION

Speaking more than one language is the norm for most of the world's population (U.S. Census Bureau, 2013; Craik & Bialystok, 2006), and multilingualism is increasing notably (Cenoz et al., 2006). Although language proficiency can improve remarkably through exposure over the years, even to the point of reaching native-like proficiency, there is a familiar phenomenon that remains challenging throughout the life of a bilingual individual: In noisy environments, comprehension is hard in a second language (L2) but seems relatively effortless in a first language (L1). Our understanding of the computational and neural foundations of this ubiquitous phenomenon is rather limited. A few hypotheses have attempted to account for this experience (Hervais-Adelman, Pefkou, & Golestani, 2014; Golestani, Hervais-Adelman, Obleser, & Scott, 2013; Ferreira, Engelhardt, & Jones, 2009; Hahne & Friederici, 2001), and although somewhat different in scope, they have all proposed a lack of successful use of top–down linguistic information as the source of this effect. The rationale is as follows: In any given situation where humans listen to a degraded signal, they use top–down linguistic knowledge such as the sentential context or predicted semantic meaning of the sentence to calculate and repair the message that has been obscured by the poor quality of the input. Researchers have argued that bilingual individuals not having as easy an access to this top–down semantic information in their second language leads to an inability to repair the speech signal and to consequently not understand the message (Hervais-Adelman et al., 2014). Here, we tested the possibility that, in addition to a failure to accurately apply high-level linguistic information, the source of this persistent difficulty may also lie in an inability to perform a lower-level process reported to aid comprehension (Zoefel, Archer-Boyd, & Davis, 2018): the neurophysiologically well-established concept of neural entrainment to speech (Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Buzsáki & Draguhn, 2004).

To characterize quantitatively the effect of noise across different L2 proficiency levels, we recruited bilingual (Mandarin Chinese, American English) participants who were (i) Mandarin Chinese native speakers with low English proficiency, (ii) Mandarin native speakers with high English proficiency (these participants lived in China until adulthood and had learned English since they were young, but only in an academic setting), and (iii) native speakers of Mandarin who were English dominant (born to at least one Mandarin-speaking caregiver in an English-speaking country). Thus, our carefully selected participant sample covered the full spectrum of possible language proficiency combinations in both languages. We recorded magnetoencephalographic (MEG) responses while participants listened to four-word sentences at different signal-to-noise ratios (SNRs), varying from completely clear to fully unintelligible speech. We discovered that the neural responses that track the physical speech rhythm are affected by noise—but not by language proficiency. In contrast, responses tracking linguistic structure reflect the interaction between noise and

[1] New York University, [2] New York University Abu Dhabi, [3] Zhejiang University, [4] Max Planck Institute for Empirical Aesthetics

https://doi.org/10.1162/jocn_a_01610

knowledge. Hence, complementing previous research suggesting that greater availability of top–down linguistic information may account for the difference in L1 versus L2 comprehension (Hervais-Adelman et al., 2014), the data show that an automatic lower-level mechanism tracking speech also contributes to the prevalent effect of impoverished comprehension of L2 speech in noise.

METHODS

Participants

Fifty-one right-handed Mandarin–English bilingual individuals participated in the experiment (16 men, 35 women; age: M = 20 years, SD = 2.45 years). To meaningfully characterize the effect of noise across varied second-language proficiency levels, we selected Mandarin–English bilinguals with diverse language backgrounds. Sixteen of the participants were native speakers of Mandarin who had acquired English later in life and had always lived in a Mandarin-speaking environment (age of acquisition: Mandarin = 1.31 years [SD = 1.53 years], English = 7.62 years [SD = 3.5 years]). Their self-reported oral (speaking and understanding) proficiency was 96.8% in Mandarin (SD = 4.7%) and 68.5% in English (SD = 7.1%), and their score in the Woodcock–Muñoz English-language survey was 51.8%. English had never been their language for socializing, and they had rarely used it outside a classroom context. Seventeen participants were native speakers of Mandarin with high proficiency in English. They had acquired English earlier in life at international schools but had grown up in a Mandarin-speaking environment (age of acquisition: Mandarin = 1.33 years [SD = 0.88 years], English = 5.91 years [SD = 4.42 years]). They had moved to the United States for undergraduate education at some point in the past 3 years and had since been in an English-speaking environment. Their self-reported oral proficiency was 94.1% in Mandarin (SD = 3.2%) and 82.9% in English (SD = 6.4%), and their score in the Woodcock–Muñoz English-language survey was 68.8%. Finally, we tested a group of English-dominant speakers (n = 18), who were born to at least one Mandarin-speaking parent in the United States (age of acquisition: Mandarin = 1.83 years [SD = 0.70 years], English = 2.25 years [SD = 4.24 years]). Hence, they had learned Mandarin from birth, but the dominant language in their environment and everyday use had always been English. Their self-reported oral proficiency was 76.4% in Mandarin (SD = 6.1%) and 95.6% in English (SD = 2.41%), and their score in the English-language survey was 91.4%. In this last group, participants reported that their life unfolded fully in English except for at home, where they spoke Mandarin, and they reported their English to be significantly better than their Chinese.

Our grouping criterion was validated post hoc by submitting participants to K-means clustering based on all the collected language profile variables (i.e., age of acquisition, exposure, self-reported proficiency, and quantitative measures of proficiency) and showing that our criterion matched the output of this unsupervised clustering algorithm, $t(49) = 7.22, p < .001$ (see additional materials for detailed language background information). Information about their language use and proficiency level was gathered with a modified version of the language background questionnaire of Marian, Blumenfeld, and Kaushanskaya (2007; see additional materials[1] for full language background information). All participants were neurologically intact with normal or corrected-to-normal vision, and all provided informed written consent following New York University institutional review board protocols.

Stimuli

Participants listened to four-syllable sentences, concatenated and isochronously presented, in English and Mandarin. Mandarin stimuli were 50 four-syllable sentences taken from Ding, Melloni, Zhang, Tian, and Poeppel (2016; MandarinMaterials, Supplementary Table 1,[2] four-syllable sentences), in which the first two syllables formed a noun phrase and the last two formed a verb phrase (Figure 1B, left). The combination of the noun and the verb phrase formed a complete sentence. All syllables were presented isochronously, lasted between 75 and 354 msec (mean duration = 224 msec; Figure 1B, top right), and were adjusted to 250 msec by truncation or padding silence at the end. Each trial consisted of the sequential presentation of 10 of these sentences, and crucially, no acoustic gaps were inserted between sentences, as such a gap would constitute an unwanted acoustic cue for segmentation. For this reason, the intensity of the stimulus, as shown by the sound envelope, only fluctuated at the syllabic rate (see Figure 1AB, bottom right, Supplementary Figure 1 for stimulus spectrograms[3]). That being said, the sequences of four syllables constituted clearly segmentable sentences. English materials consisted of 60 four-syllable sentences. Each sentence consisted of four monosyllabic words combined to form two-word noun phrases (adjective + noun) and two-word verb phrases (verb + noun). The combination of these two phrases resulted in four-word sentences (e.g., "big rocks block roads"; Figure 1A, left). All syllables were between 250 and 347 msec in duration and were adjusted to 320 msec by padding or truncation (Figure 1A, top right). For both English and Mandarin, a 25-msec cosine window smoothed the offset of each syllable. All sentences in English and Mandarin are displayed in the supplementary materials. Although the syllable duration in the two languages was different, the manipulation was designed to ensure isochronous presentation, which is the essential feature of the experimental design.

We embedded all sentences in four different levels of white noise. We first measured the power of the sentences in isolation and then added white noise to reach

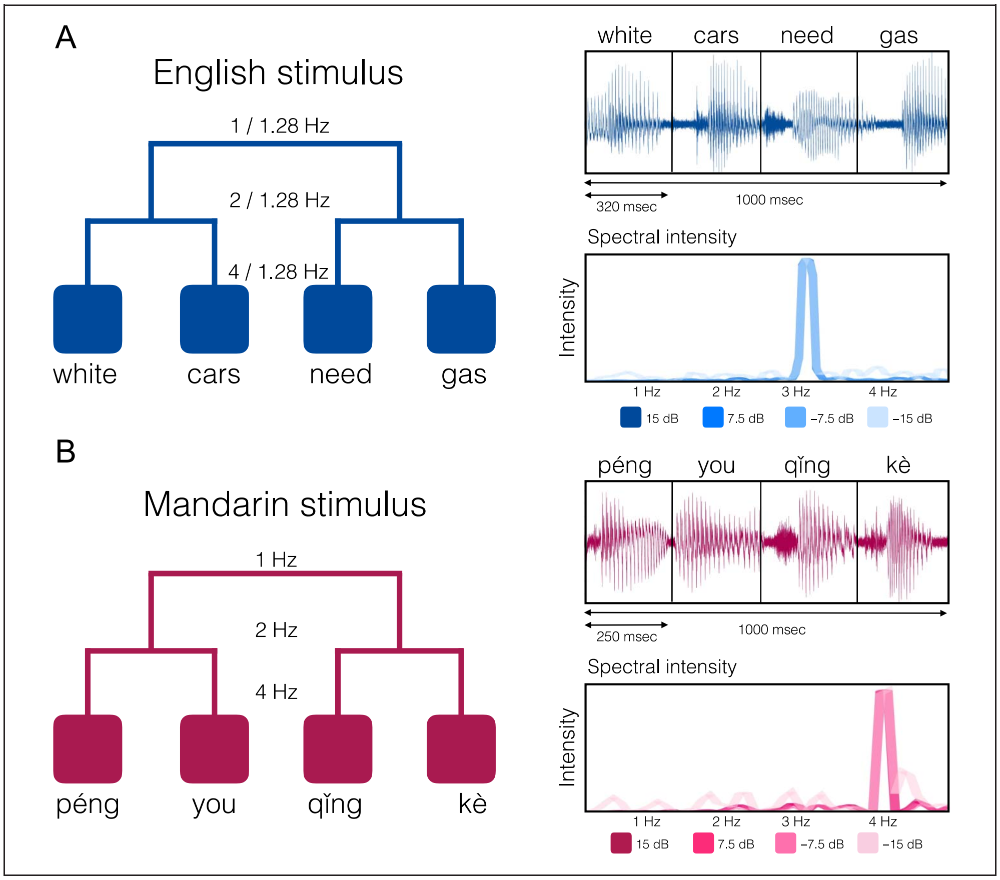


Figure 1. Sample English (A) and Mandarin (B) stimuli. Monosyllabic words were presented isochronously, forming phrases and sentences. (Left) Presentation rate and syntactic structure of the stimuli. In each of A and B: top right, waveform of a sample stimulus; bottom right, spectral intensity of the stimulus at each tested level of noise, revealing a syllabic-scale rhythm at all levels of noise but no phrasal or sentential rate modulation.

the desired output SNR in decibels. The SNR levels ranged from +15 dB (clear speech) to −15 dB (unintelligible speech in noise) in 7.5-dB intervals. Although ideally (and eventually) babbling noise may be a better background noise to mirror the type of noise experienced in real life, white noise was selected for a first characterization to avoid confounding semantic, phonological, or language interference effects that went beyond the effects of noise qua noise. Participants heard 160 sentences at each noise level for each language, and the experiment took approximately 1 hr to complete. Across participants, each four-syllable sentence was presented an equal number of times at each level of noise, and the presentation of these sentences and noise levels was fully randomized.

Procedure

Before the MEG recording, each participant's head shape was digitized using a Polhemus dual-source handheld FastSCAN laser scanner. MEG data were collected in the Neuroscience of Language Laboratory at New York University using a whole-head 157-channel axial gradiometer system (Kanazawa Institute of Technology) as participants lay in a dimly lit, magnetically shielded room. Trials began with the binaural auditory presentation of the stimuli. Participants listened to sets of 10 randomly presented four-syllable sentences in either language and were then presented with a 1–4 scale on-screen. Listeners had to indicate via button press how much they understood (1 = *nothing at all* and 4 = *everything*). The validity of this comprehension measure has been established by showing qualitatively comparable results between this subjective measure of comprehension and participants' performance on recall tests (Doelling, Arnal, Ghitza, & Poeppel, 2014; Ghitza, 2012). After the button press, the next trial began. After the MEG recording, all participants completed the Woodcock–Muñoz Language Survey to evaluate their proficiency in English. Participants completed the first four parts of this survey, aimed to assess their oral, listening, reading, and writing skills. The completion of this test took around 45 min.

Data Acquisition and Preprocessing

MEG data were recorded at 1000 Hz (200-Hz low-pass filter), noise reduced via the continuously adjusted least-squares method (Adachi, Shimogawara, Higuchi, Haruta, & Ochiai, 2001) in MEG Laboratory software (Yokogawa Electric and Eagle Technology), and epoched from beginning to end of the auditory stimulus. The MEG responses were decomposed into components using a denoising source separation technique (de Cheveigné & Simon, 2008; for a detailed explanation, see Ding et al., 2016), and the first five components were kept for analysis and projected back into sensor space. This technique decomposes MEG recordings to extract the neural response components that are consistent over trials, and it was applied to accurately estimate the strength of neural activity phase-locked to the stimulus. To avoid the transient response at the beginning of each trial, data were only analyzed from the beginning of the second sentence of each 10-sentence trial. Single-trial responses per noise level and participant were Fourier transformed into the frequency domain and subsequently averaged within condition to obtain an evoked response per condition per participant.

Data were source localized with MNE-Python (Gramfort et al., 2013, 2014). To estimate the distributed electrical current image in the brain at each time sample, we used the minimum norm approach (Hämäläinen & Ilmoniemi, 1994) via MNE (MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging). The cortical surfaces were constructed using an icosahedron subdivision of five and mapping an average brain from FreeSurfer (CorTech and MGH/HMS/MIT Athinoula A. Martinos Center for Biomedical Imaging) to the head-shape data gathered from the head-scanning process. This generated a source space of 5124 points for each reconstructed surface, leaving ~6.2 mm of spacing within sources (cortical area per source = ~39 mm^2). Then, the boundary-element model method was used to calculate the forward solution. The 100-msec prestimulus period was used to construct the noise covariance matrix and to apply as a baseline correction. The inverse solution for each participant was then computed from the noise-covariance matrix, the forward solution, and the source covariance matrix and was applied to the evoked response for each condition. The application of the inverse solution determined the most likely distribution of neural activity in source space. Minimum norm current estimates were computed for three orthogonal dipoles, of which the root mean square was retained as a measure of activation at that source (thus, the orientation of the dipole was free unsigned). The resulting minimum norm estimates of neural activity were transformed into normalized estimates of noise at each spatial location using the default regularization factor (SNR = 3). Hence, we obtained noise-normalized SPMs, which provide information about the statistical reliability of the estimated signal at each location in the map with millisecond accuracy. Then, those SPMs were converted to dynamic maps (dSPMs). To quantify the spatial resolution of these maps, the point-spread function for different locations on the cortical surface was computed. The point spread is defined as the minimum norm estimate resulting from the signals coming from a current dipole located at a certain point on the cortex. The calculation of the point-spread function following the approach of Dale et al. (2000) reduces the location bias of the estimates, in particular, the tendency of the minimum norm estimates to prefer superficial currents (i.e., their tendency to misattribute focal, deep activations to extended, superficial patterns). Hence, by transforming our minimum norm estimates to dSPM, we obtained an accurate spatial blurring of the true activity patterns in the spatiotemporal maps (Dale et al., 2000).

Analyses

Behavioral Data

For the main statistical tests, we conducted mixed-effects model analyses using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2012) using noise (−15, −7.5, 7.5, and 15 dB), proficiency, and the interaction between them as fixed effects and participant as a random effect. In addition, we conducted complementary categorical analyses within each group of participants to assess the effect of noise in intelligibility within each language profile specifically. For this analysis, responses for each trial were averaged within participant for each noise level. We subsequently conducted a related samples two-tailed t test across participants to assess whether their comprehension in English and Mandarin at each noise level significantly differed. All reported p values are false discovery rate (FDR) corrected for multiple comparisons.

Tracking Analyses

MEG activity for each trial was averaged within participant at each noise level in the frequency domain. We then subjected the amplitude at the syllabic and phrasal peak to the same linear mixed-effects model used on behavioral data, with noise, proficiency, and their interaction as fixed effects and participant as a random effect (low-frequency environmental noise during the recordings prevented us from analyzing entrainment to the sentential level). Having assessed that the interaction of noise and proficiency significantly accounted for the amplitude of the peaks, we split participants by proficiency to unpack the nature of this effect and assess the effect of noise at both the syllabic and phrasal peak within these groups. For each spectral peak, a one-tailed paired t test was used to test if the neural response across all sensors in a frequency bin was significantly stronger across participants than the average of the four neighboring frequency bins (two bins on each side). We corrected for multiple comparisons across t tests using FDR correction. The application of a 1000-permutation test in lieu of the original t tests revealed the same significant results.

Source Localization Analysis

Having obtained distinct tracking effects across languages for each participant group at −7.5 dB, we turned to the source-localized data to identify from where in the cortex this activity was emerging. For this purpose, we subtracted the response to the English sentences at −7.5 dB from the response to Mandarin sentences at −7.5 dB (averaged across participants). This analysis hence revealed the localization of the oscillatory analysis effect.

RESULTS

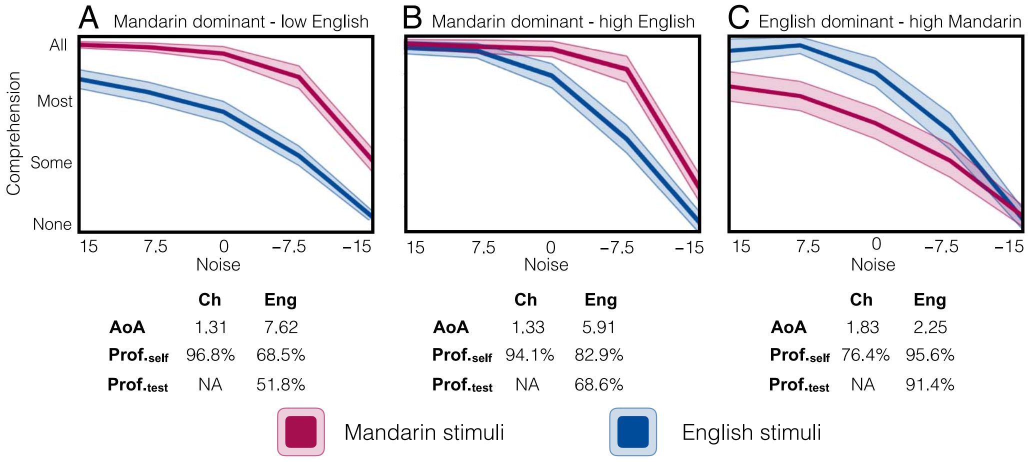
Behavioral Results

Behavioral results showed that the influence of noise on the comprehension of speech varied based on individuals' language proficiency. A linear mixed-effects regression model regressing noise, age of acquisition, and language proficiency on comprehension revealed that, (i) in addition to noise and proficiency significantly influencing comprehension independently, (ii) they also interacted significantly, such that the decrease in comprehension because of noise was greater, the lower the proficiency of the individual in that language. This effect held both for Chinese (noise: $F(1, 4551) = 148, p < .001$; proficiency: $F(1, 46.8) = 31.5, p < .001$; interaction between noise and proficiency: $F(1, 4550) = 4.45, p = .03$) and English (noise: $F(1, 4975) = 44.1, p < .001$; proficiency: $F(1, 47) = 7.72, p = .007$; interaction between noise and proficiency: $F(1, 4975) = 23.39, p < .001$) stimuli.

Next, to unpack these results, we complemented the continuous regression analysis with categorical analyses wherein we assessed the influence of noise in comprehension for each language group. We found that, for Mandarin-dominant participants with low English proficiency (Figure 2A), the lower comprehension of English compared to Mandarin was constant across all levels of noise ($p < .001$). However, Mandarin speakers with high English proficiency understood both languages equally in clear speech (15 dB: $p = .09$; 7.5 dB: $p = .14$), but their comprehension of English was severely impaired in noisy conditions (−7.5 and −15 dB: $p < .001$; Figure 2B). Finally, we found that English-dominant bilinguals showed an overall impaired comprehension of Mandarin, except at the highest level of noise; in that case, participants did not report understanding in either language (15, 7.5, and −7.5 dB: $p < .001$; −15 dB: $p = .12$; Figure 2C). Overall, behavioral results reveal that the influence of noise on the comprehension of speech varied contingent on language knowledge.

MEG: Cortical Tracking Results

While participants listened to the sentences, we recorded neural activity with MEG. Cortical oscillations have been proposed to be likely candidates for segmentation of continuous speech (Zoefel et al., 2018; Gross et al., 2013; Luo & Poeppel, 2007). We performed analyses aimed to quantify how entrainment to different linguistic levels (words vs. phrasal structures) was disrupted by noise, on the one hand, and language proficiency, on the other (low-frequency environmental noise during the recordings prevented us from analyzing entrainment to the sentential level). The MEG responses were transformed into the frequency domain, and we retained for subsequent analysis the five neural response components that were the most consistent over trials, as identified by spatial filters (see Methods). Results revealed the distinct influence of noise on the cortical tracking of different linguistic levels. A linear mixed-effects regression on the amplitude of syllabic and phrasal peaks with age of acquisition, language proficiency, noise, and their interactions as continuous regressors revealed that, whereas the level of tracking at the syllabic level was only affected by noise, $F(3, 144) = 2.64, p = .05$ (age of acquisition: $F(3, 144) = 1.09, p = .35$; language proficiency: $F(3, 144) = 1.98, p = .11$; interaction between noise and language: $F(3, 144) = 1.81, p = .14$), at the phrasal level, there was a



	Ch	Eng
AoA	1.31	7.62
Prof.$_{self}$	96.8%	68.5%
Prof.$_{test}$	NA	51.8%

	Ch	Eng
AoA	1.33	5.91
Prof.$_{self}$	94.1%	82.9%
Prof.$_{test}$	NA	68.6%

	Ch	Eng
AoA	1.83	2.25
Prof.$_{self}$	76.4%	95.6%
Prof.$_{test}$	NA	91.4%

Figure 2. Comprehension performance for participants as a function of noise, averaged across participants (shaded area: 95% confidence interval). Below each graph, we report the average age of acquisition (AoA), self-reported proficiency (Prof.$_{self}$), and English proficiency score in the Woodcock language questionnaire (Prof.$_{test}$) for the participants in each group.

reliable interaction between noise and language proficiency, $F(3, 144) = 2.76$, $p = .04$, whereas the main effects of age of acquisition, $F(3, 144) = 0.98$, $p = .4$, language proficiency, $F(3, 144) = 2.1$, $p = .1$, and noise, $F(3, 144) = 2.03$, $p = .11$, were not significant. This shows that the tracking capabilities of noisy signals vary across individuals depending on their proficiency. Furthermore, there was a significant correlation between the amplitude of the syllabic peak and that of the phrasal peak, $t(202) = 40$, $p < .001$, suggesting a relation between participants' capacity to segment incoming speech and parsing the syntactically relevant structure.

Analogously to the behavioral data analysis, we complemented the continuous analysis with a categorical analysis where we assessed how individuals of different language profiles tracked both syllabic and phrasal structures. In consonance with the regression analysis, we found that participants of all proficiency combinations tracked the syllabic rhythm at all levels of noise, and the amplitude of this tracking response decreased as noise increased (Figure 3: 3.2-Hz response at the top, 4-Hz response at the bottom). However, the tracking of phrasal structure was heavily dependent on language proficiency and comprehension (Figure 3: 1.6-Hz response at the top, 2-Hz response at the bottom).

Specifically, with regard to speech rhythm tracking, Mandarin speakers with low English proficiency tracked syllabic rhythm in Mandarin (15 dB: $p < .001$; 7.5 dB: $p = .002$; −7.5 dB: $p = .001$; −15 dB: $p = .008$) and English (15 dB: $p < .001$; 7.5 dB: $p < .001$; −7.5 dB: $p = .003$; −15 dB: $p = .002$), as did Mandarin native speakers with high English proficiency (Mandarin: 15 dB, $p = .04$; 7.5 dB, $p = .04$; −7.5 dB, $p = .04$; −15 dB, $p = .03$; English: 15 dB, $p < .04$; 7.5 dB, $p = .04$; −7.5 dB, $p = .03$; −15 dB, $p = .04$) and English-dominant speakers (English: all noise levels, $p < .001$; Mandarin: 15 and 7.5 dB, $p < .001$; −7.5 dB, $p = .021$; −15 dB, $p = .01$).

In contrast, there was a clear disparity in the tracking response to phrase-level structure across languages depending on the proficiency combinations of the participants. Mandarin-dominant speakers with low English proficiency did not track English phrases (15 dB: $p = .33$; 7.5 dB: $p = .52$; −7.5 dB: $p < .1$; −15 dB: $p = .21$; Figure 3A), although they did track phrases in Mandarin at all levels of noise (15, 7.5, and −7.5 dB: $p < .001$), except during pure noise (−15 dB: $p = .53$; Figure 3D). As proficiency in English increased, so did the tracking of the English phrases. Mandarin speakers with high English proficiency did track phrases at the clearest levels of speech in English (15 dB: $p < .04$; 7.5 dB: $p = .02$), although not at the two noisier levels (−7.5 dB: $p = .16$; −15 dB: $p = .22$; Figure 3B), while also being able to track Mandarin phrases at all levels of noise (15 dB: $p = .04$; 7.5 dB: $p = .04$; −7.5 dB: $p = .02$) except during pure noise (−15 dB: $p = .15$; Figure 3E). Finally, English-dominant speakers showed phrasal tracking of

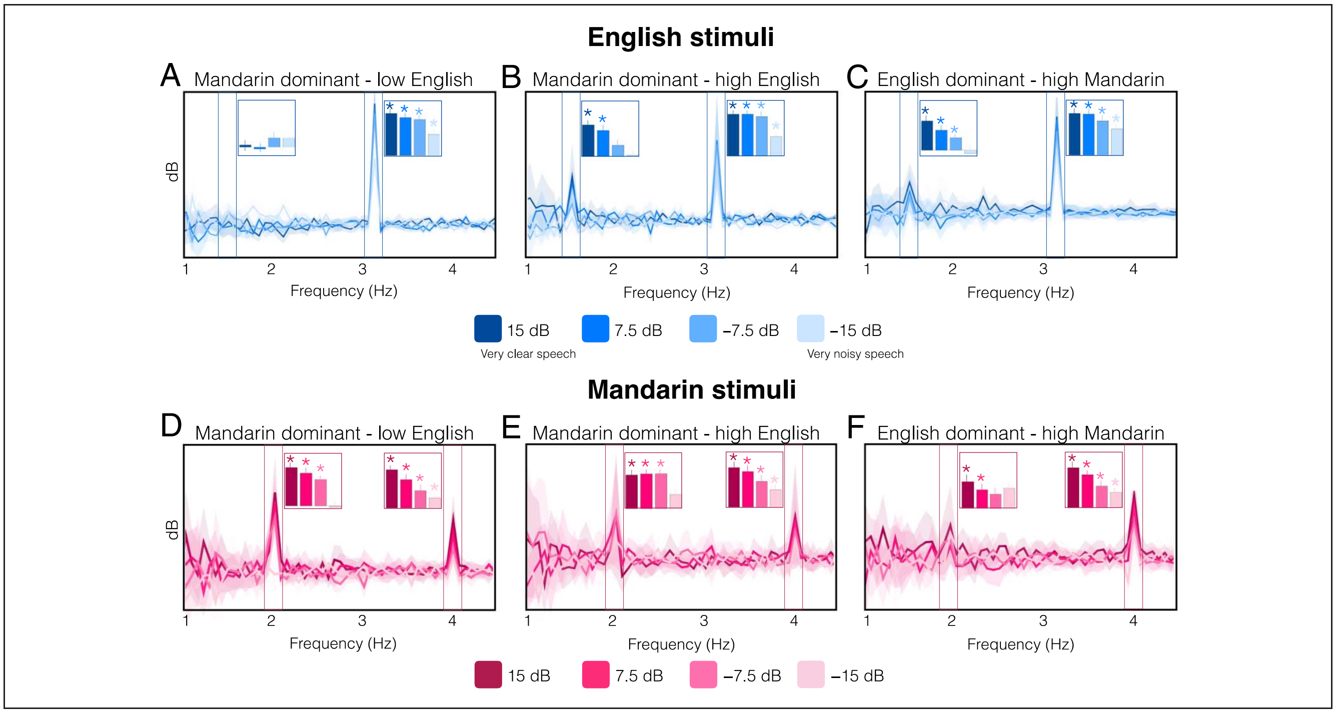


Figure 3. MEG-derived neural response spectra for each language group (Mandarin dominant − low English [$n = 16$], Mandarin dominant − high English [$n = 12$], and English dominant − high Mandarin [$n = 12$]). Solid lines indicate average response; shading indicates 95% CI. Spectral peaks at corresponding frequencies reflect whether there was neural tracking of syllabic or phrasal rhythms at a given level of noise. Frequency bins with significantly stronger power than two neighbors on each side are marked with an asterisk (*) of the corresponding color ($p < .05$, paired one-sided t test, FDR corrected).

English phrases at 15 and 7.5 dB, as did the Mandarin-dominant speakers with high English proficiency (p = .04 and p = .002, respectively). However, crucially, the speakers with higher English proficiency were additionally able to track phrases at −7.5 dB (p = .04; Figure 3C). Hence, the increase in English proficiency was accompanied by tracking of phrasal structures at higher levels of noise. In contrast, these same English-dominant speakers whose Mandarin was mildly worse were only able to track phrases in Mandarin at the two clearest levels of speech (15 dB [p = .03] and 7.5 dB [p = .01]), but not at −7.5 dB (p = .09) or −15 dB (p = .1) like Mandarin-dominant speakers had done (Figure 3F). Hence, the categorical analyses confirmed what the regression analysis revealed: Noise and proficiency interact critically in the tracking of higher-level linguistic structures.

These results build on previous findings reporting an effect of intelligibility on cortical tracking of speech (Park, Ince, Schyns, Thut & Gross, 2015; Doelling et al., 2014; Peelle, Gross, & Davis, 2013) but reveal a more complex and informative pattern than previously known. Specifically, we show (i) that not all levels of entrainment are affected equally by noise and (ii) that not only the physical properties of the stimuli but also the language proficiency of the listener affects the degree of entrainment to different linguistic structures. These results complement research showing that overall neural oscillatory activity underlying speech processing also varies with second language proficiency (Pérez & Duñabeitia, 2019; Pérez, Carreiras, Dowens, & Duñabeitia, 2015).

MEG: Source Localization

Finally, we performed analyses to identify from where in the cortex the reported activation patterns were emerging. We source-localized the same five MEG components submitted to the frequency-domain analysis and compared neural responses to English and Mandarin at −7.5 dB, as this was the SNR at which language proficiency clearly determined the presence or absence of phrasal entrainment. The analysis revealed that activity in areas surrounding the auditory cortex elicited increased activity in response to the better-understood language. Furthermore, this increase widened commensurate with the imbalance between languages (Figure 4). This is consistent with

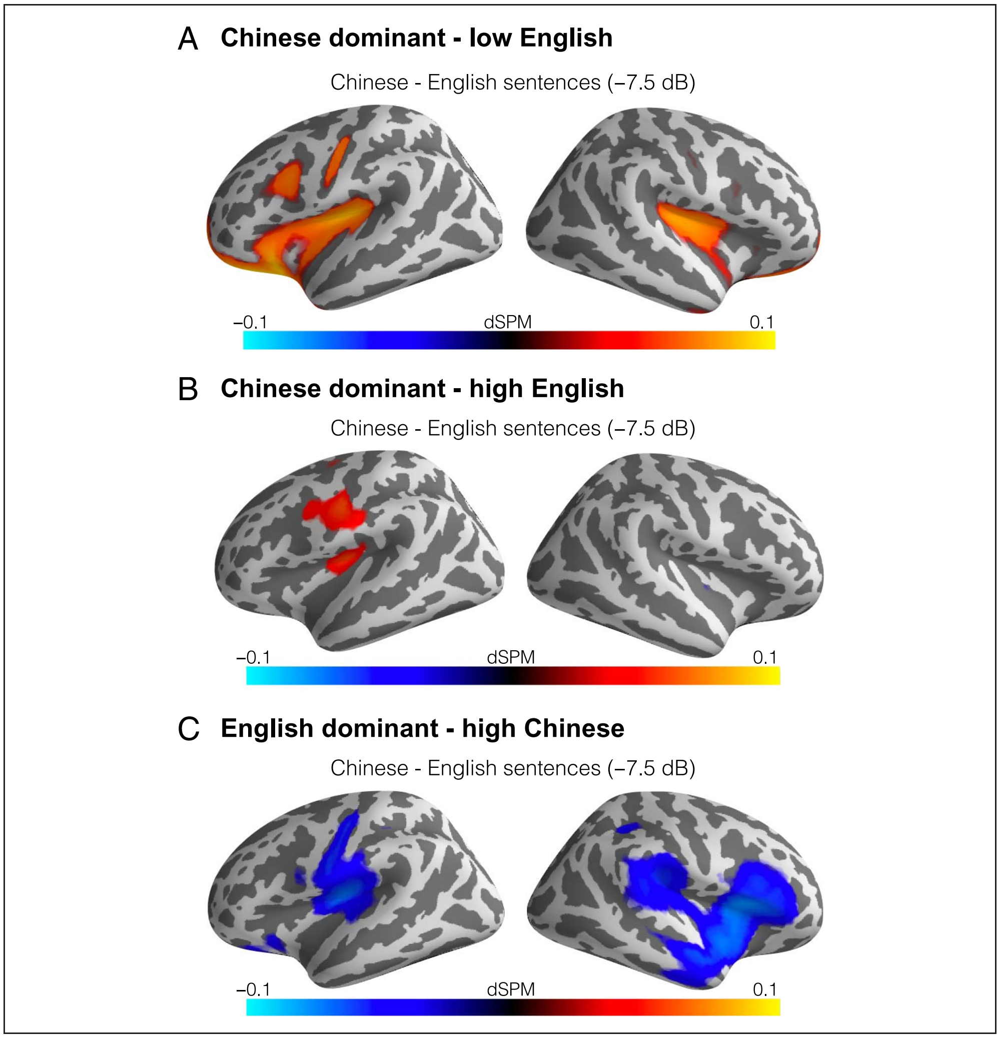

A Chinese dominant - low English
Chinese - English sentences (−7.5 dB)
−0.1
dSPM
0.1
B Chinese dominant - high English
Chinese - English sentences (−7.5 dB)
−0.1
dSPM
0.1
C English dominant - high Chinese
Chinese - English sentences (−7.5 dB)
−0.1
dSPM
0.1

Figure 4. Whole-brain source localization of the five neural response components most consistent over trials, identified with a denoising source separation technique. The whole-brain images show the result of subtracting the average activity elicited by English sentences (at −7.5 dB) from the average activity elicited by Mandarin sentences (at −7.5 dB), averaged across participants. Activity is displayed in noise-normalized SPMs (dSPM; blue indicates higher activity for English stimuli; and red, for Mandarin stimuli).

previous research that localized intelligibility effects (Keitel, Gross, & Kayser, 2018; Peelle et al., 2013).

DISCUSSION

The combined behavioral and neurophysiological data we present capitalize on recent findings and illuminate speech and language processing in new ways. In particular, it has been shown that neural entrainment to speech signals enables both tracking of the acoustic input, principally the amplitude envelope, but also higher-order structure building operations that are not accessible from the physical input alone (Figures 1 and 3; Keitel et al., 2018; Ding et al., 2016; Luo & Poeppel, 2007). In this study, we found that neural responses at the theta band that track the physical speech rhythm are only affected by noise—but not by language proficiency. In contrast, neural tracking of phrasal structure at delta level was affected by the interaction of noise and language knowledge. We advance current understanding in two significant ways, deriving from the parametric nature of the experimental design, in which we concurrently vary the quality of the speech signal (Figure 1) and the language proficiency of the listeners (Figure 2).

First, the paradigm allows us to establish that the language knowledge of the listener determines the spectral information required for speech recognition (cf. Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). In other words, there is no such thing as a categorical limit on how impoverished the signal can be before comprehension is compromised. Instead, this boundary is malleable and shifts in concordance with the linguistic capabilities of the listener.

Relatedly, we show that, neurophysiologically, it is at the phrasal level (cf. Figure 1) that differential knowledge of language is especially influential. The minimum level of SNR necessary to facilitate phrase-level tracking is between −15 and −7.5 dB in the native language and between −7.5 and 7.5 dB in the nonnative language (Figure 3), which in turn forms the basis for structure building. Listeners who track their L1 well at −7.5 dB fail at tracking their L2 at that same SNR, suggesting that the tracking impairment is not because of peripheral causes but implicates higher-order limitations. This finding suggests that it is not only conscious comprehension but also unconscious neural processes that are sensitive to the interaction between noise and language knowledge. This finding is consistent with research suggesting that delta oscillations are not primarily involved in early sound analysis and phonological processing, but rather, they reflect the encoding of abstract syntactic structures (Kösem & van Wassenhove, 2017), with recent findings showing that delta band tracking yields a significant prediction of speech comprehension (Etard & Reichenbach, 2019). Importantly, some level of comprehension was achieved even in the absence of phrasal tracking, suggesting that tracking is not sine qua non to achieve basic understanding. Rather, it would seem that tracking enhances comprehension and it may be necessary to reach full comprehension.

These results shed new light on the conceptualization of multilingual language comprehension. Although previous accounts revealed that the source of the prevalent bilingual impairment to comprehend speech in noise emerges from deficiencies in access to lexical or syntactic information in this language (Hervais-Adelman et al., 2014; Ferreira et al., 2009; Hahne & Friederici, 2001), our results suggest that this impairment is additionally reflected, and perhaps instigated, by a failure to successfully complete lower-level processes. Although this experiment cannot prove causality by itself, this proposal is supported by recent research in monolingual individuals showing that the capability to entrain to linguistic structures can in fact causally affect comprehension (Zoefel et al., 2018).

Mechanistically, by hypothesis, comprehension may be enhanced by a feedback loop such that bottom–up rhythmic structure and top–down information mutually aid the prediction and processing of upcoming signals (Peelle et al., 2013). In bilinguals, both of these processes may be compromised: Although previous research has focused on the information availability aspect (Hervais-Adelman et al., 2014; Golestani et al., 2013), we show that the impoverished comprehension of speech in noise by L2 learners is also determined by a disruption in the entrainment to linguistic structures. Importantly, these processes do not seem dissociable: There was a significant correlation between the amplitude of the syllabic peak and the amplitude of the phrasal peak, suggesting a meaningful relation between participants' capacity to segment incoming speech (i.e., a signal-based low-level process) and parsing the syntactically relevant structure (i.e., a high-level process).

In summary, our results characterize the minimum SNR requirements for neural entrainment to different linguistic structures, specify the differential influence of noise and knowledge on syllabic and phrasal tracking, and reveal a neurophysiological pattern that may underlie the widely experienced phenomenon of compromised comprehension of second language speech in noisy environments.

Acknowledgments

This work was supported by National Institutes of Health grant 5R01DC05660 and Max Planck Society funding to D. P., New York University Abu Dhabi Institute grant G1001 to L. P., and Woodcock Institute Research Award to E. B. E.

Reprint requests should be sent to Esti Blanco-Elorrieta, Department of Psychology, New York University, 10 Washington Place, New York, NY 10003, or via e-mail: eb134@nyu.edu.

Notes

1. The questionnaire can be retrieved from https://estiblancoelorrieta.github.io/Modified_lang_quest.pdf.

2. MandarinMaterials, Supplementary Table 1 can be retrieved from https://staticcontent.springer.com/esm/art%3A10.1038%2Fnn.4186/MediaObjects/41593_2016_BFnn4186_MOESM63_ESM.pdf.

3. Supplementary Figure 1 can be retrieved from https://estiblancoelorrieta.github.io/Modified_lang_quest.pdf.

REFERENCES

Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., & Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. *IEEE Transactions on Applied Superconductivity*, *11*, 669–672.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, *304*, 1926–1929.

Cenoz, J., Nunes, P., Riganti, P., Onofri, L., Puzzo, B., & Sachdeva, R. (2006). Benefits of linguistic diversity and multilingualism. Position Paper of Research Task 1.2. Sustainable Development in a Diverse World (SUS. DIV). European Commission.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: Mechanisms of change. *Trends in Cognitive Sciences*, *10*, 131–138.

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron*, *26*, 55–67.

de Cheveigné, A., & Simon, J. Z. (2008). Denoising based on spatial filtering. *Journal of Neuroscience Methods*, *171*, 331–339.

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage*, *85*, 761–768.

Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, *39*, 5750–5759.

Ferreira, F., Engelhardt, P. E., & Jones, M. W. (2009). Good enough language processing: A satisficing approach. In N. Taatgen, H. Rijn, J. Nerbonne, & L. Schomaker (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 413–418). Austin, TX: Cognitive Science Society.

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, *3*, 238.

Golestani, N., Hervais-Adelman, A., Obleser, J., & Scott, S. K. (2013). Semantic versus perceptual interactions in neural processing of speech-in-noise. *Neuroimage*, *79*, 52–61.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, *11*, e1001752.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, *4*, 123–141.

Hämäläinen, M. S., & Ilmoniemi, R. J. (1994). Interpreting magnetic fields of the brain: Minimum norm estimates. *Medical & Biological Engineering & Computing*, *32*, 35–42.

Hervais-Adelman, A., Pefkou, M., & Golestani, N. (2014). Bilingual speech-in-noise: Neural bases of semantic context use in the native language. *Brain and Language*, *132*, 1–6.

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

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

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, *320*, 110–113.

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

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967.

Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top–down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*, 1649–1653.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*, 1378–1387.

Pérez, A., Carreiras, M., Dowens, M. G., & Duñabeitia, J. A. (2015). Differential oscillatory encoding of foreign speech. *Brain and Language*, *147*, 51–57.

Pérez, A., & Duñabeitia, J. A. (2019). Speech perception in bilingual contexts: Neuropsychological impact of mixing languages at the inter-sentential level. *Journal of Neurolinguistics*, *51*, 258–267.

R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

U.S. Census Bureau. (2013). The 2011 statistical abstract. Languages spoken at home by language: 2008, Table 53. https://www2.census.gov/library/publications/2010/compendia/statab/130ed/tables/11s0053.pdf.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*, 401–408.