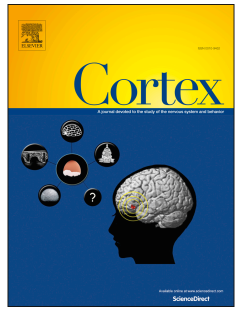
Accepted Manuscript



Successful non-native speech perception is linked to frequency following response phase consistency

Akihiro Omote, Kyle Jasmin, Adam Tierney

PII: S0010-9452(17)30160-0

DOI: 10.1016/j.cortex.2017.05.005

Reference: CORTEX 2015

To appear in: *Cortex*

Received Date: 31 January 2017

Revised Date: 19 April 2017

Accepted Date: 11 May 2017

Title: Successful non-native speech perception is linked to frequency following response phase consistency

Authors: Akihiro Omote[1], Kyle Jasmin[1], Adam Tierney[1†]

Affiliations

[1]Department of Psychological Sciences, Birkbeck, University of London, London, UK

[†]Corresponding author
Adam Tierney, Ph.D.
Birkbeck, University of London
Malet Street, London, WC1E 7HX
United Kingdom
Email: a.tierney@bbk.ac.uk
Phone: +44-020-7631-6368

Acknowledgements

We would like to thank Dr. Robert Slevc for sharing the Receptive Phonology test stimuli. This work was supported by the Wellcome Foundation grant #109719/Z/15/Z.

Abstract

Some people who attempt to learn a second language in adulthood meet with greater success than others. The causes driving these individual differences in second language learning skill continue to be debated. In particular, it remains controversial whether robust auditory perception can provide an advantage for non-native speech perception. Here, we tested English speech perception in native Japanese speakers through the use of frequency following responses, the evoked gamma band response, and behavioral measurements. Participants whose neural responses featured less timing jitter from trial to trial performed better on perception of English consonants than participants with more variable neural timing. Moreover, this neural metric predicted consonant perception to a greater extent than did age of arrival and length of residence in the UK, and neural jitter predicted independent variance in consonant perception after these demographic variables were accounted for. Thus, difficulties with auditory perception may be one source of problems learning second languages in adulthood.

Keywords: auditory, English, FFR, Japanese, speech

1. Introduction

Speaking and understanding a second language is a vital skill in an increasingly globalized world. However, learning a second language poses difficulties that surpass those experienced in learning a first language. Native Japanese speakers, for example, struggle to discriminate English /l/ and /r/ (Goto, 1971; Miyawaki et al., 1975). Nevertheless, the difficulties which non-native speech perception presents can be overcome. Native Japanese speakers, for example, through experience (MacKain et al., 1981; Flege et al., 1995; Ingvalson et al. 2011) and training (Logan et al., 1990; Lively et al. 1993, 1994; Bradlow et al., 1996, 1999; McCandliss et al., 2002; Iverson et al., 2005; Lim and Holt, 2011) can learn to perceive and produce the distinction between /l/ and /r/ with near-native accuracy. However, there are large individual differences in the degree to which non-native speech sound categories can be successfully acquired: some people achieve approximately native perception and production, while others produce heavily accented speech and struggle to perceive non-native speech even after extensive training (Wong and Perrachione, 2007; Golestani and Zatorre, 2009; Perrachione et al., 2011; Hanulíková et al., 2012; Kempe et al., 2012, 2015). Understanding the source of these individual differences would be an important step towards the development of tools to boost non-native speech perception.

Learning a non-native speech sound category requires highly precise perception of durational, pitch, and spectral information. One possible source of difficulties with non-native speech perception, therefore, is imprecise auditory perception. Supporting this theory, individual differences in non-native speech perception have been linked to non-verbal auditory perception skills, including amplitude envelope discrimination (Kempe et al., 2012), frequency discrimination (Lengeris and Hazan, 2010), pitch perception (Wong and Perrachione, 2007; Perrachione et al., 2011), and spectral discrimination (Kempe et al., 2015). However, electrophysiology research has supported a speech-specific source for non-native speech perception difficulties. Díaz et al. (2008, 2015), for example, found that non-native speech perception ability was linked to neural discrimination of speech sounds but not non-verbal sounds differing in duration or frequency.

This link between speech sound discrimination and individual differences in non-native speech perception has been replicated across languages (Garcia-Sierra et al., 2011; Jakoby et al., 2011; Zhang et al., 2009).

Here we examine the link between non-native speech sound perception and auditory processing in Japanese adults learning English as a second language using frequency-following responses (FFRs), an electrophysiological response which reproduces the frequencies present in the evoking sound and reflects early auditory processing in the brainstem and cortex (Coffey et al., 2016). The FFR features high test-retest reliability (Hornickel et al. 2012) and reflects neural origins in the brainstem and cortex (Coffey et al., 2016), making it an excellent measure of the robustness of early auditory processing. The precision of FFRs has been linked to individual differences in the development of language skills in children (Hornickel and Kraus, 2013; White-Schwoch et al., 2015), but it remains unknown how FFR precision relates to second language acquisition. Recently, Krizman et al., (2012) reported that bilingual FFRs more robustly encoded the fundamental frequency (F0) of synthesized speech. Here, therefore, we predicted that non-native speech perception ability would relate to F0 phase-locking. Given that impaired gamma-rate phase-locking has also been shown to characterize children with language impairment (Heim et al., 2011), we additionally investigated relationships between gamma phase-locking and non-native speech perception.

2. Methods

2.1 Participants

Participants were 25 native Japanese speakers (13 female, aged 19 to 35 (M = 29.3, SD = 4.5)) with English learning experience at secondary school level or above in Japan. Participants were required to have arrived in the UK after the age of 18 and to have been resident there for at least 1 month at the time of testing. Secondary inclusion criteria included normal audiometric thresholds (<= 25 dB HL for octaves from 250 to 8000Hz) and lack of diagnosis of a language impairment. Participants received a mean (sd) score of 7.6 (4.1) on the Musical Experience portion of the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014), indicating low levels of musical training. Mean age of arrival in the UK was 27.8 (4.9) years, and mean duration of residence in the UK was 2.6 (3.1) years. The Ethics Committee in the Department of Psychological Sciences at Birkbeck, University of London approved all experimental procedures. Informed consent was obtained from all participants. Participants were compensated £14 for their participation.

2.2 Behavioral Measures

English speech perception was tested using the Receptive Phonology Test (Slevc and Miyake, 2006). Each question in this test is designed to assess a phonological contrast in English with which Japanese subjects have difficulty. The test contains three main sections. In the *word* sub-test, participants see a list of 26 word pairs which differ in a single speech sound (e.g., "late/rate"). Participants then hear a list of words and are asked to indicate which of the two words they heard. In the *sentence* sub-test, participants see a list of 25 sentences, with one of the words replaced with a word differing in a single speech sound (e.g., "My sister loves to play with crowns/clowns.")

Participants then hear a list of sentences and are asked to circle the word that they heard. Finally, participants listen to a short story and are given a written version of the story that includes 42 underlined words. Participants are asked to circle any of the underlined words that are mispronounced.

Because the original version of the Receptive Phonology Test featured a speaker of American English, test materials were re-recorded by a native speaker of British English (Received Pronunciation) in soundproof room with a RODE NT1-A Condenser Microphone. 3 of the items from the original test were removed, as they feature speech sound contrasts which do not exist in British Received Pronunciation. Audio recordings were presented to participants using Sennheiser HD 25-1-II headphones. See Table 1 for a list of all of the speech sound contrasts included in the test.

2.3 Electrophysiology

2.3.1 Stimuli

Participants were presented with two 170-ms synthesized speech sounds [la] and [ra]. These syllables were synthesized using a Klatt synthesizer, as implemented in Praat (Boersma and Weenink, 2016). The two syllables differed only during the first 70 ms, during which each had a unique frequency trajectory for the third formant (F3). For [la], F3 was steady at 3400 Hz from 0 to 30 ms, then decreased linearly to 2530 Hz by 70 ms. For [ra], F3 was steady at 1601 Hz from 0 to 30 ms, then increased to 2530 Hz by 70 ms. All other stimulus characteristics were identical across stimuli. F1 was steady at 478 Hz from 0 to 30 ms then increased to 705 Hz by 70 ms. F2 was steady at 1088 Hz from 0 to 30 ms then decreased to 1035 Hz by 70 ms. From 70 to 170 ms F1, F2, and F3 were steady at 705, 1035, and 2530 Hz, respectively. F0 and F4 were constant throughout the stimulus at 100 Hz and 3850 Hz. A cosine off ramp with a duration of 20 ms was used to avoid transients. Figure 1 displays waveforms and spectrograms for the two stimuli.

2.3.2 Recording parameters

During electrophysiological testing participants sat in a comfortable chair in a soundproof booth with negligible ambient noise and read a book of their choice. Stimuli were presented through Etymotic ER-II earphones in alternating polarity at 80 +/- 1 dB SPL to both ears with an inter-onset interval of 251 ms. 6300 trials were collected for each stimulus, and stimuli were presented in blocks (i.e. all [ra] trials were collected in a single block). Electrophysiological data were recorded in LabView 2.0 (National Instruments, Austin, TX) using a BioSEMI Active2 system via the ActiABR module with a sample rate of 16,384 Hz and an online bandpass filter (100-3000 Hz, 20 dB/decade). The active electrode was placed at Cz, the grounding electrodes CMS and DRL were placed on the forehead at FP1 and FP2, and the reference electrodes were placed on the earlobes. Earlobe references were not electrically linked during data collection. Offset voltage for all electrodes was kept below 50 mV.

2.3.3 Data reduction

Electrophysiological data reduction was conducted in Matlab R2016a. Offline amplification was applied in the frequency domain for 3 decades below 100 Hz with a 20 dB rolloff per decade. The data was organized into epochs 40 ms before through 210 ms after the onset of the stimulus and baseline corrected. To ensure against contamination by electrical noise a second-order IIR notch filter with a Q-factor of 100 was used with center frequencies of 50, 150, 250, 350, 450, and 550 Hz. A bandpass filter (0.1–2000 Hz, 12dB/oct) was then applied to the continuous EEG recording, and epochs exceeding +/-100 µV were rejected as artifacts. The first 2,500 artifact-free responses to each stimulus polarity then were selected for further analysis.

2.3.4 Data analysis (>70 Hz)

To investigate the precision of neural sound encoding we calculated inter-trial phase-locking. This measure involves calculating the phase consistency at a particular frequency across trials and, therefore, no averaging is necessary. This procedure provides information similar to spectral analysis of average waveforms, but with a higher signal-to-noise ratio and less susceptibility to artifact (Zhu et al. 2013).

All electrophysiological data analysis was conducted in Matlab 2016a. Parameters for FFR analysis were used for frequencies >70 Hz, in accordance with the standards of previous research on speech FFRs (Bidelman and Krishnan, 2009, Parbery-Clark et al., 2009). For FFR analysis (> 70 Hz), phase-locking was calculated within 40-ms windows that were applied repeatedly across the epoch with a 1 ms step size. First, for each trial, a Hanning windowed fast Fourier transform was calculated. Second, for each frequency, the resulting vector was transformed into a unit vector. Third, all of the unit vectors were averaged. The length of the resulting vector—ranging from 0 (no phase consistency) to 1 (perfect phase consistency)—was then calculated as a measure of cross-trial phase consistency. Phase locking factors for [la] and [ra] were averaged together to form a global estimate of an individual's inter-trial phase locking.

This time-frequency data was then averaged in the following manner. First, data were collapsed across the entire response (10 to 170 ms). Phase-locking at the fundamental frequency (100 Hz) and the second through sixth harmonics was measured by extracting the maximum phase-locking value in a 40-Hz bin centered on each frequency. (Harmonics above 600 Hz were not consistently represented in every single participant and were therefore excluded.) Phase-locking at the harmonics was averaged together to form a general measurement of harmonic encoding. In addition, phase-locking was measured separately in the response to the consonant (10 to 80 ms) and the response to the vowel (80-170 ms).

2.3.5 Data analysis (<70 Hz)

For lower-frequency analysis (< 70 Hz), phase-locking was calculated within 80-ms windows with a 1 ms step size. Visual inspection of the cross-subject average (see Figure 2) revealed an increase in phase-locking over baseline between 0 and 60 ms. Gamma phase-locking was quantified, therefore, as the average phase-locking within a window reaching from 0 to 60 ms and between 30 and 70 Hz.

2.3.6 Statistical analyses

Linear models of the behavioral and neural data were constructed using the lm() function with the software package 'R', and model comparisons were performed with the anova() function. For comparisons of correlations that shared one variable in common (Steiger, J.H. 1980), the r.test() function in the 'psych' package from 'R' was used.

3. Results

First we tested whether the ability to discriminate English consonants was related to our neural measures. Better performance (greater proportion correct items) on the consonant discrimination items of the Phonology Test was associated with greater phase-locking to F0 (R^2=.379, F(1,23)=14.03, p=.001) and with greater phase-locking within the gamma band (R^2=.21, F(1,23)=6.11, p=.021). Vowel errors were not associated with F0 phase locking (R^2=.053, F(1,23)=1.30, p=.27) or gamma phase locking (R^2=.000), and phase-locking to the harmonics (H2-H6) did not correlate with performance on consonant items (R^2=.03, F(1,23)=.78, p=.34) or vowel items (R^2=.025, F(1,23)=.059, p=.45). The correlation between phase-locking at F0 and consonant perception was significantly greater than the correlation with vowel perception (T=2.76, p=.011); similarly, the correlation between gamma phase-locking and consonant performance was significantly greater than the correlation with vowel perception (T=2.95, p =.007). The correlation between consonant perception and phase-locking at F0 was significantly greater than the correlation with phase-locking at the higher harmonics (T=2.81, p =.01). Figure 2 displays phase-locking for the cortical evoked response and FFR across all subjects. Figure 3 displays cortical and FFR phase-locking for good and poor perceivers of English consonants (top-bottom split). Figure 4 is a scatterplot displaying FFR phase-locking and cortical phase-locking versus consonant perception performance.

One possible explanation for this relationship between English speech perception and F0 phase-locking is that greater familiarity with English speech leads to enhanced encoding of neural responses to English speech sounds. If so, one would expect the relationship between English consonant perception and F0 phase-locking to be limited to the response to the consonant, which did not overlap with any Japanese speech sound. On the other hand, if our results reflect a more general relationship between precise auditory encoding and non-native speech perception, then English consonant perception should also relate to F0-phase-locking in the response to the vowel, which contained formant frequencies appropriate for a Japanese [a] (Nishi et al., 2008). We found that F0 phase-locking in the response to the consonant (10-80 ms) correlated with performance on consonant items (R^2 = 0.426, p = 0.001). F0 phase-locking in the response to the vowel (80-170) also correlated with performance on consonant items (R^2 = 0.260, p = 0.009). Moreover, the relationship between consonant perception and F0 phase-locking did not significantly differ between these two portions of the response (T=0.97, p = 0.34).

To further test whether confounding effects of language experience could explain our results, "Age Arrived in UK" and "Years in UK" were used to assess the extent of participants' experience with English. "Years in UK" was cube root-transformed to bring its distribution closer to normality (Shapiro-Wilk W=.89, p>.01 after transformation). Subjects who were older when they arrived in the UK made more consonant errors, although the correlation was only marginally significant (R^2=0.15, F(1,23)=4.02, p=.057). Age Arrived in UK also correlated negatively with F0 phase locking (R^2=.17, F(1,23)=4.77, p=.039), as well as gamma phase locking (R^2=.25, F(1,23)=7.71, p=.01). The number of years subjects had spent in the UK prior to testing was correlated with F0 phase locking (R^2=.31, F(1,23)=7.51, p=.004), but not with gamma phase locking (R^2=.014, F(1,23)=.337, p=.57).

To assess whether our neural measures predicted variance in phonological competence that could not be simply explained by experience, we fit two linear models: one with age of arrival in the UK and years residence in the UK predicting consonant performance (the "Experience Only" model), and another which also included the consistency of the neural response (F0 phase locking; the "Experience plus Neural model"). The two predictors in the Experience Only model together accounted for 25% of the variance on consonant performance. The Experience plus Neural model with F0 phase locking as a predictor performed significantly better than the Experience Only model (F(1,21)=5.43, p=.030), with the F0 phase-locking predictor accounting for an additional 15% of the variance for consonant performance. Including gamma phase locking as an additional predictor only accounted for an additional 1.5% of the variance, and this reduction in error was not significant (p=.50).

Finally, to investigate links between individual differences in low-frequency and high-frequency phase-locking, we compared phase-locking in the gamma band to phase-locking in the FFR at F0 and the harmonics. Gamma phase-locking was correlated with phase-locking at both F0 (R^2=.31, p=.004) and the harmonics (R^2=0.17, p=.039).

4. Discussion

Here we examined English speech perception and neural sound encoding in twenty-five native speakers of Japanese who moved to the United Kingdom as adults. We found that English consonant perception was linked to the degree of phase-locking to the fundamental frequency of the frequency-following response (FFR) to sound and to phase-locking within the gamma band. Vowel perception, however, did not relate to neural phase-locking. The relationship between these neural metrics and English speech perception ability remained significant even after time in the UK and age of arrival were controlled for.

That FFR phase-locking relates to second language speech perception suggests that difficulties with auditory perception can interfere with the acquisition of non-native speech sound categories. On the other hand, we found that non-native vowel perception was not linked to FFR phase-locking, suggesting that vowel perception may depend less on the precision of auditory processing. These findings support previous behavioral research demonstrating relationships between non-native speech perception and auditory abilities including amplitude envelope discrimination (Kempe et al., 2012), frequency discrimination (Lengeris and Hazan, 2010), and spectral discrimination (Kempe et al., 2015). However, language learning is a complex process, and there are likely many ways in which foreign language learning can be disrupted. Only a portion of children with reading impairment, for example, display problems with auditory perception (Ramus et al., 2003), and the causes of adult language learning difficulty are likely to be similarly heterogenous. FFR phase-locking may be a useful metric to help identify people whose difficulties with non-native language perception stem from auditory impairments.

These findings support and extend previous work demonstrating links between the precision of neural sound encoding, language skill, and language experience. Krizman et al. (2015), for example, found that in Spanish-English bilinguals degree of bilingual experience was linked to the strength of fundamental frequency (F0) encoding in the FFR. Here we replicate this relationship in native speakers of Japanese learning English as a second language, and extend this finding by showing that this same neural metric can also explain individual differences in non-native speech perception, even after language experience is accounted for. Hornickel and Kraus (2013) demonstrated that the inter-trial consistency of the FFR is linked to individual differences in language skills in school-age children; here we show that precise neural encoding of sound is linked to successful adult language learning as well. Chandrasekaran et al. (2012) showed that the robustness of FFR pitch encoding can predict subsequent short-term learning of lexical tones; here we show that FFR phase-locking is linked to long-term language learning of non-tonal speech sounds.

What is the mechanism underlying this relationship between FFR phase-locking and non-native speech perception ability? One possibility is that FFR phase-locking reflects the precision of temporal perception. FFR phase-locking has been linked to the ability to precisely synchronize movements with sound onsets (Tierney and Kraus, 2013, 2016; Woodruff Carr et al., 2016). This suggests that precise tracking of sound timing relies upon consistent auditory neural timing, as synchronization places stringent demands upon the precision of auditory time perception (on the order of a few milliseconds; Repp, 2000). The ability to track sound timing is also vital for speech perception, as the temporal information contained in the speech envelope contains information relevant to speech sound discrimination (Rosen, 1992); in fact, discrimination of speech sounds is possible even if spectral information is greatly reduced (Shannon et al., 1995). Moreover, non-native speech perception may rely more upon temporal information than does native speech perception. For example, Japanese adults have a strong bias towards the use of temporal information such as closure duration and formant transition duration when distinguishing [la] and [ra], whereas native English speakers rely more heavily upon the frequency of the third formant (Iverson et al., 2005).

We replicate the finding of Krizman et al. (2012) that F0 encoding in the FFR is related to degree of bilingual experience but encoding of the harmonics is not. Moreover, we show that phase-locking at the F0 but not the harmonics is also linked to non-native speech perception ability. The specificity of this relationship was predicted based on these previous findings, but the underlying mechanism remains unclear. One possibility is that this result reflects a relationship between non-native speech perception ability and cortical auditory encoding. There is strong evidence that frequency-following responses at 250 Hz and above are generated within the auditory brainstem, as cooling the inferior colliculus in cats abolishes the scalp-recorded FFR (Smith et al. 1975) and patients with inferior colliculus lesions do not display an FFR (Sohmer et al. 1977). However, both of these studies included no stimuli below 250 Hz, and recent work has suggested that the FFR at 100 Hz is generated within multiple sources, including both cortical and subcortical regions (Coffey et al., 2016). Thus, the higher frequencies of the FFR may reflect a greater contribution from more peripheral areas such as the inferior colliculus, as generally the upper limit of phase-locking to sound is lower in more central structures (Joris et al., 2004). Our finding of a relationship between non-native speech perception ability and phase-locking within both the low-frequency FFR and the gamma band, therefore, may indicate that learning a second language in adulthood relies upon precise cortical but not subcortical auditory processing. This hypothesis cannot be properly evaluated by the current study; however, it could be tested by future work examining FFR phase-locking and non-native speech perception using MEG.

Previous work (Nagarajan et al., 1999; Heim et al. 2011) has demonstrated that children with language learning difficulties have less phase-locked gamma band onset responses to sounds presented with a short inter-stimulus interval (ISI). Here we find that degree of gamma phase-locking is linked to non-native speech perception. Given that our stimuli were presented with a short ISI, this could reflect an impaired ability to process rapidly presented sounds on the part of the participants who struggled to learn to perceive English. Future work could examine this hypothesis by examining links between non-native speech perception and gamma phase-locking to stimuli presented at different ISIs. This enhanced gamma phase-locking in participants better able to perceive English may also reflect greater recruitment of speech processing resources in response to synthesized English speech sounds in these participants, as gamma phase-locking has been shown to be greater for speech stimuli as compared to non-speech stimuli (Palva et al., 2002). This would be consistent with fMRI evidence showing that subjects who are better at learning novel speech sounds display more STG activity when passively listening to speech sounds (Archila-Suerte et al., 2016). Finally, gamma phase-locking has also been hypothesized to be an important component of speech perception in multi-time resolution models (Poeppel et al., 2008), in which phonetic information is carried within the gamma band and prosodic information is carried within the delta and theta bands. Greater gamma phase-locking in the participants who were better able to perceive English speech may, therefore, indicate more precise neural encoding of the timing of the speech envelope. This interpretation is supported by our finding that gamma phase-locking was correlated with FFR phase-locking.

One limitation of this work is that it is difficult to rule out the possibility that the link between neural sound encoding and non-native speech perceptual ability is driven by experiential factors. Time spent in the United Kingdom, for example, was linked to both F0 phase-locking and English perception, a relationship which is likely contributing to the link between F0 phase-locking and speech perception performance. However, the relationship between neural sound encoding and non-native speech perception held even after time in the UK and age of arrival were controlled for, suggesting that this relationship partially reflects the dependence of successful non-native language learning on auditory skills. Moreover, the relationship between non-native speech perception and F0 phase-locking held both for the neural response to the consonant, which did not overlap with any Japanese speech sound category, and the response to the vowel, which contained formant frequencies similar to those of the Japanese [a] (Nishi et al. 2008). Nevertheless, in a retrospective study it is difficult to account for all possible confounding experiential factors. This limitation could be addressed in future work in which participants are tested prior to beginning study of a foreign language for the first time or through the use of very short-term training paradigms (Lim and Holt, 2011).

References

Archila-Suerte P, Bunta F, Hernandez A (2016) Speech sound learning depends on individuals' ability, not just experience. International Journal of Bilingualism 20: 231-253.

Bidelman G, Krishnan A (2009) Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. Journal of Neuroscience 29: 13165-13171.

Boersma P, Weenink D (2016) Praat: doing phonetics by computer [Computer program]. Version 6.0.22, retrieved from http://www.praat.org/

Bradlow A, Pisoni D, Akahane-Yamada R, Tohkura Y (1996) Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. JASA 101: 2299-2310.

Bradlow A, Akahane-Yamada R, Pisoni D, Tohkura Y (1999) Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. Perception and Psychophysics 61: 977-985.

Chandrasekaran B, Kraus N, Wong P (2012) Human inferior colliculus activity relates to individual differences in spoken language learning. J Neurophysiol 107: 1325-1336.

Coffey E, Herholz S, Chepesiuk A, Baillet S, Zatorre R (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. Nature Communications 7: 11070.

Díaz B, Baus C, Escera C, Costa A, Sebastián-Gallés N (2008) Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. PNAS 105: 16083-16088.

Díaz B, Mitterer H, Broersma M, Escera C, Sebastián-Gallés N (2015) Variability in L2 phonemic learning originates from speech-specific capabilities: an MMN study on late bilinguals. Bilingualism: Language and Cognition 19: 955-970.

Flege J, Takagi N, Mann V (1995) Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. JASA 99: 1161-1173.

Garcia-Sierra A, Rivera-Gaxiola M, Percaccio C, Conboy B, Romo H, Klarman L, Ortiz S, Kuhl P (2011) Bilingual language learning: an ERP study relating early brain responses to speech, language input, and later word production. Journal of Phonetics 39: 546-557.

Golestani N, Zatorre R (2009) Individual differences in the acquisition of second language phonology. Brain and Language 109: 55-67.

Goto H (1971) Auditory perception by normal Japanese adults of the sounds "L" and "R". Neuropsychologia 9: 317-323.

Hanulíková A, Dediu D, Fang Z, Basnaková J, Huettig F (2012) Individual differences in the acquisition of a complex L2 phonology: a training study. Language Learning 62: 79-109.

Heim S, Friedman J, Keil A, Benasich A (2011) Reduced oscillatory activity during rapid auditory processing as a correlate of language-learning impairment. Journal of Neurolinguistics 24: 538-555.

Hornickel J, Kraus N (2013) Unstable representation of sound: a biological marker of dyslexia. Journal of Neuroscience 33: 3500-3504.

Ingvalson E, McClelland J, Holt L (2011) Predicting native English-like performance by native Japanese speakers. J Phon 39: 571-584.

Iverson P, Hazan V, Bannister K (2005) Phonetic training with acoustic cue manipulations: a comparison of methods for teaching English /r/-/l/ to Japanese adults. JASA 118: 3267-3278.

Jakoby H, Goldstein A, Faust M (2011) Electrophysiological correlates of speech perception mechanisms and individual differences in second language attainment. Psychophysiology 48: 1517-1531.

Joris P, Schreiner C, Rees A (2004) Neural processing of modulated sounds. Physiol Rev 84: 541-577.

Kempe V, Thoresen J, Kirk N, Schaeffler F, Brooks P (2012) Individual differences in the discrimination of novel speech sounds: effects of sex, temporal processing, musical and cognitive abilities. PLoS ONE 7: e48623.

Kempe V, Bublitz D, Brooks P (2015) Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. British Journal of Psychology 106: 349-366.

Krizman J, Marian V, Shook A, Skoe E, Kraus N (2012) Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages. Proceedings of the National Academy of Sciences 109: 7877-7881.

Krizman J, Slater J, Skoe E, Marian V, Kraus N (2015) Neural processing of speech in children is influenced by extent of bilingual experience. Neuroscience Letters 585: 48-53.

Lengeris A, Hazan V (2010) The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. JASA 128: 3757-3568.

Lim S, Holt L (2011) Learning foreign sounds in an alien world: videogame training improves non-native speech categorization. Cognitive Science 35: 1390-1405.

Lively S, Logan J, Pisoni D (1993) Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. JASA 94: 1242-1255.

Lively S, Pisoni D, Yamada R, Tohkura Y, Yamada T (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. JASA 96: 2076-2087.

Logan J, Lively S, Pisoni D (1990) Training Japanese listeners to identify English /r/ and /l/: a first report. JASA 89: 875-886.

MacKain K, Best C, Strange W (1981) Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics 2: 369-390.

Marian V, Blumenfeld H, 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.

McCandliss B, Fiez J, Protopapas A, Conway M, McClelland J (2002) Success and failure in teaching the [r]-[r] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. Cognitive, Affective, and Behavioral Neuroscience 2: 89-108.

Miyawaki K, Strange W, Verbrugge R, Liberman A, Jenkins J (1975) An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. Perception and Psychophysics 18: 331-340.

Nagarajan S, Mahncke H, Salz T, Tallal P, Roberts T, Merzenich M (1999) Cortical auditory signal processing in poor readers. PNAS 96: 6483-6488.

Nishi K, Strange W, Akahane-Yamada R, Kubo R, Trent-Brown S (2008) Acoustic and perceptual similarity of Japanese and American English vowels. JASA 124: 576-588.

Palva S, Palva J, Shtyrov Y, Kujala T, Ilmoniemi R, Kaila K, Naatanen R (2002) Distinct gamma-band evoked responses to speech and non-speech sounds in humans. Journal of Neuroscience 22: 1-5.

Parbery-Clark A, Skoe E, Kraus N (2009) Musical experience limits the degradative effects of background noise on the neural processing of sound. Journal of Neuroscience 29: 14100-14107.

Perrachione T, Lee J, Ha L, Wong P (2011) Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. JASA 130: 461-472.

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

Ramus F, Rosen S, Dakin S, Day B, Castellote J, White S, Frith U (2003) Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. Brain 126: 841-865.

Repp B (2000) Compensation for subliminal timing perturbations in perceptual-motor synchronization. Psychol Res 63: 106-128.

Rosen S (1992) Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions: Biological Sciences 336: 367-373.

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

Skoe E, Kraus N (2010) Auditory brainstem response to complex sounds: a tutorial. Ear and Hearing 31: 302-324.

Slevc R, Miyake A (2006) Individual differences in second-language proficiency: does musical ability matter? Psychological Science 17: 675-681.

Steiger J (1980) Tests for comparing elements of a correlation matrix. Psychological Bulletin 87: 245-251.

Tierney A, Kraus N (2013) The ability to move to a beat is linked to the consistency of neural responses to sound. Journal of Neuroscience 33: 14981-14988.

Tierney A, Kraus N (2016) Getting back on the beat: links between auditory-motor integration and precise auditory processing at fast time scales. European Journal of Neuroscience 43: 782-791.

White-Schwoch T, Woodruff Carr K, Thompson EC, Anderson S, Nicol T, Bradlow AR, Zecker SG, Kraus N (2015) Auditory processing in noise: a preschool biomarker for literacy. PLoS Biology 13: e1002196.

Wong P, Perrachione T (2007) Learning pitch patterns in lexical identification by native English-speaking adults. Applied Psycholinguistics 28: 565-585.

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

Zhang Y, Kuhl P, Imada T, Iverson P, Pruitt J, Stevens E, Kawakatsu M, Tohkura Y, Nemoto I (2009) Neural signatures of phonetic learning in adulthood: a magnetoencephalography study. NeuroImage 46: 226-240.

Zhu L, Bharadwaj H, Xia J, Shinn-Cunningham B (2013) A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. JASA 134: 384-395.

Speech sound contrast	**Number of items**
consonants	38
b-v	4
f-h	6
l-r	14
n-ŋ	3
s-ʃ	3
s-θ	8
vowels	32
æ-ɛ	4
æ-ʌ	6
ɑː-ʌ	1
ɒ-ʊ	1
ɒ-ʌ	2
ʊ-ɔː	5
ɜː-ɑː	5
iː-ɪ	4
ɪ-ɛ	4

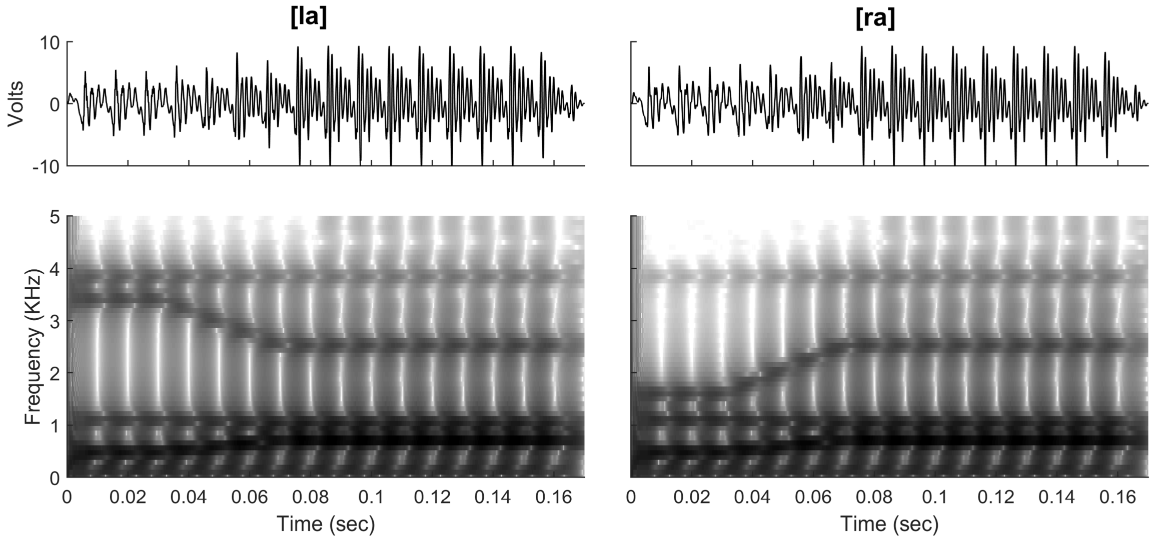
Table 1. Speech sound contrasts included in the receptive phonology test.

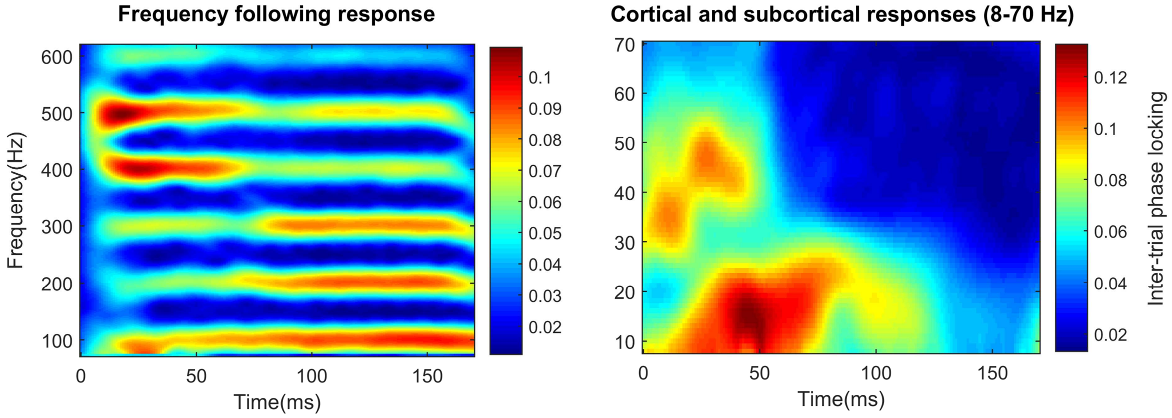
Figure 1. Waveforms (top) and spectrograms (bottom) of synthesized speech stimuli. The [la] and [ra] stimuli differed only in the first 70 ms, and were identical thereafter.

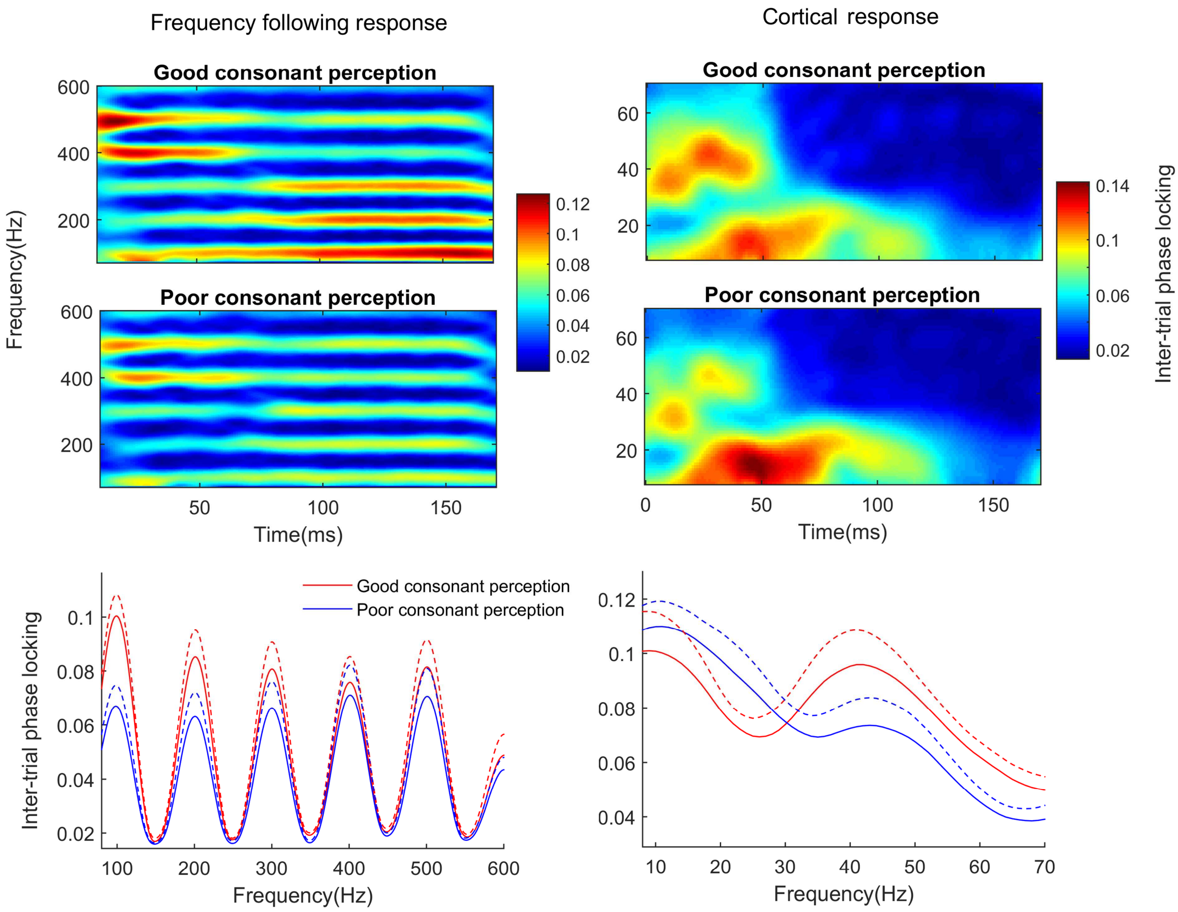
Figure 2. (Left) Time-frequency plot of inter-trial phase locking across all subjects for the frequency following response (71-600 Hz). (Right) Time-frequency plot of inter-trial phase locking across all subjects for the cortical response (8-70 Hz).

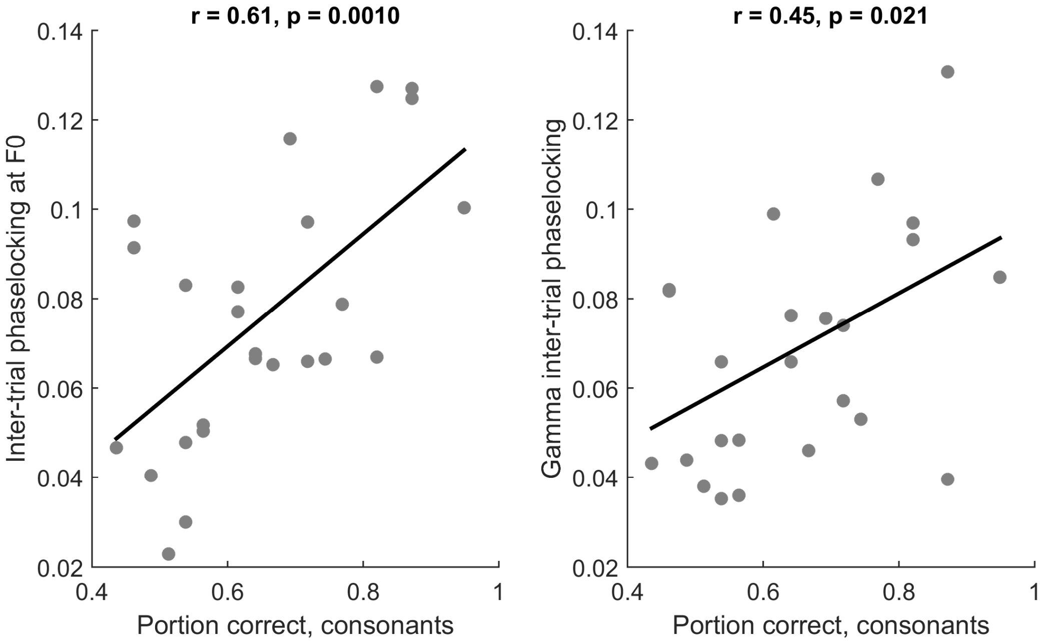
Figure 3. (Left, top) Time-frequency plot of inter-trial phase locking for the frequency following response for participants with good versus poor perception of English consonants. Participants were divided into top and bottom halves based on performance on the consonant portions of the receptive phonology test. (Right, top) Time-frequency plot of inter-trial phase locking for the cortical response for good versus poor consonant perceivers. (Left, bottom) Inter-trial phase locking in the frequency following response as a function of frequency across the entire response (10-170 ms) for good (red) versus poor (blue) consonant perceivers. Error bars are one standard error of the mean. (Right, bottom) Inter-trial phase locking in the frequency following response as a function of frequency across the first 60 ms of the response for good versus poor consonant perceivers.

Figure 4. (Left) Scatterplot displaying performance on the consonant portions of the receptive phonology test (displayed as portion correct) versus inter-trial phaselocking at the fundamental frequency during the entirety of the frequency following response. (Right) Scatterplot displaying consonant perception versus inter-trial phaselocking within the gamma band (31-70 Hz) during the first 60 ms of the cortical response. R-values and p-values are derived from Pearson correlations.


[la]
[ra]
Volts
10
0
-10
Frequency (KHz)
5
4
3
2
1
0
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
Time (sec)


Frequency following response
Frequency(Hz)
600
500
400
300
200
100
0
50
100
150
Time(ms)
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Cortical and subcortical responses (8-70 Hz)
70
60
50
40
30
20
10
0
50
100
150
Time(ms)
0.12
0.1
0.08
0.06
0.04
0.02
Inter-trial phase locking


Frequency following response
Cortical response
Good consonant perception
Poor consonant perception
Good consonant perception
Poor consonant perception
Frequency(Hz)
Time(ms)
Time(ms)
Inter-trial phase locking
Good consonant perception
Poor consonant perception
Inter-trial phase locking
Frequency(Hz)
Frequency(Hz)


r = 0.61, p = 0.0010
Inter-trial phaselocking at F0
Portion correct, consonants
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0.4
0.6
0.8
1
r = 0.45, p = 0.021
Gamma inter-trial phaselocking
Portion correct, consonants