Native language shapes automatic neural processing of speech

Bastien Intartaglia, Travis White-Schwoch, Christine Meunier, Stéphane Roman, Nina Kraus, Daniele Schön

Elsevier Editorial System(tm) for Neuropsychologia

Manuscript Draft

Manuscript Number:

Title: Native language shapes automatic neural processing of speech

Article Type: Research Paper

Section/Category: Language

Keywords: Experience-dependent plasticity; Auditory brainstem responses; Speech perception

Corresponding Author: Mr. Bastien Intartaglia,

Corresponding Author's Institution: Institut de Neurosciences des Systèmes

First Author: Bastien Intartaglia

Order of Authors: Bastien Intartaglia; Travis White-Schwoch; Christine Meunier; Stéphane Roman; Nina Kraus; Daniele Schön

Abstract: The development of the phoneme inventory is driven by the acoustic-phonetic properties of one's native language. Neural representation of speech is known to be shaped by language experience, as indexed by cortical responses, and recent studies suggest that subcortical processing also exhibits this early attunement to native language. However, the majority of work to date has focused on the differences between tonal and non-tonal languages that use pitch variations to convey phonemic categories. The aim of this cross-language study is to determine whether subcortical encoding of speech sounds is sensitive to language experience by comparing native speakers of two non-tonal languages (French and English). We hypothesized that neural representations would be more robust and fine-grained for speech sounds that belong to the native phonemic inventory of the listener, and especially for the dimensions that are phonetically relevant to the listener such as high frequency components. We recorded neural responses of American English and French native speakers, listening to syllables of both languages. Results showed that, independently of the stimulus, American participants exhibited greater neural representation of the fundamental frequency compared to French participants, consistent with the importance of the fundamental frequency to convey stress patterns in English. Furthermore, participants showed more robust encoding and more precise spectral representations of the harmonics when listening to the syllable of their native language as compared to non-native language. These results are consistent with the hypothesis that language experience shapes early sensory processing of speech and that this plasticity occurs as a function of what is behaviorally-relevant to a listener.

Title : Native language shapes automatic neural processing of speech

Authors : Bastien Intartaglia (**a, b, c**), Travis White-Schwoch (**d**), Christine Meunier (**c, e**), Stéphane Roman (**f**), Nina Kraus (**d, g, h**), Daniele Schön (**a, b, c**)

a *Aix-Marseille Université, INS, Marseille, France*
b *INSERM, U1106, Marseille, France*
c *Brain and Language Research Institute, Labex BLRI*
d *Auditory Neuroscience Laboratory and Department of Communication Sciences, Northwestern University, Evanston, Illinois, United States of America*
e *Aix-Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France*
f *La Timone Children's Hospital, ENT Unit, Marseille, France*
g *Department of Neurobiology & Physiology, Northwestern University, Evanston, Illinois, United States of America*
h *Department of Otolaryngology, Northwestern University, Chicago, Illinois, United States of America*

Abstract

The development of the phoneme inventory is driven by the acoustic-phonetic properties of one's native language. Neural representation of speech is known to be shaped by language experience, as indexed by cortical responses, and recent studies suggest that subcortical processing also exhibits this early attunement to native language. However, the majority of work to date has focused on the differences between tonal and non-tonal languages that use pitch variations to convey phonemic categories. The aim of this cross-language study is to determine whether subcortical encoding of speech sounds is sensitive to language experience by comparing native speakers of two non-tonal languages (French and English). We hypothesized that neural representations would be more robust and fine-grained for speech sounds that belong to the native phonemic inventory of the listener, and especially for the dimensions that are phonetically relevant to the listener such as high frequency components. We recorded neural responses of American English and French native speakers, listening to natural syllables of both languages. Results showed that, independently of the stimulus, American participants exhibited greater neural representation of the fundamental frequency compared to French participants, consistent with the importance of the fundamental frequency to convey stress patterns in English. Furthermore, participants showed more robust encoding and more precise spectral representations of the harmonics when listening to the syllable of their native language as compared to non-native language. These results are consistent with the hypothesis that language experience shapes early sensory processing of speech and that this plasticity occurs as a function of what is behaviorally-relevant to a listener.

Keywords

Experience-dependent plasticity; Auditory brainstem responses; Speech perception

1. Introduction

While the number of consonants and vowels across world's languages is large, each language only uses a few dozen basic units. The development of this specific phoneme inventory during childhood is language dependent, meaning that it is driven by the acoustic-phonetic properties of a listener's native language. During the first months of life, infants are able to discriminate speech sounds that are not used in their native language but with growing exposure to their mother tongue, this ability declines, to finally disappear in adulthood (Werker and Tees, 2002). For example, in their cross-linguistic and longitudinal study, Werker and Tees (2002) showed that at 6-8 months of age, English infants' ability to discriminate Hindi or Salish speech contrasts is as good as native infants of the same age. Yet by 10-12 months of age, their performance drops drastically and remains poor as English-speaking adults. This decline is not restricted to Western languages: it has been also observed in Eastern languages, such as Japanese. For Japanese adults, the perceptual distinction of two acoustically close (but distinct) phonemes /r/ and /l/, which are not distinct in Japanese, is impossible (Iverson et al., 2003; Zhang et al., 2005). It is worth noting that this language-dependant reorganization of the phonemic inventory stems on two concomitant and opposite developmental patterns. Indeed, the infant's ability to discriminate foreign speech sounds decreases, while at the same time the ability to discriminate native speech sounds improves (Cheour et al., 1998; Kuhl et al., 2006, 1992; Rivera-Gaxiola et al., 2005).

Electrophysiological studies confirm the hypothesis that children develop neural representations that become attuned to the processing of their native language to the detriment of foreign languages (Kuhl, 2004; Mehler et al., 1994; Ortiz-Mantilla et al., 2013). For instance, Cheour et al. (1998) found that from 6 to 12 months of age, mismatch negativity (MMN) amplitude drops significantly for non-native phonemes. Likewise, in a longitudinal study, Rivera-Gaxiola and colleagues (2005) found that discriminatory event-related potentials (ERP) to non-native contrasts were present at 7 months of age, but were largely reduced by 11 months of age, while at the same time, the responsiveness to native language contrasts increased over time. In adults, Dehaene-Lambertz et al. (2000) showed that acoustically-close French phonemes elicit an MMN peaking around 130ms in native speakers (French), whereas the MMN is reduced or absent in non-native speakers (Japanese) who are unable to discriminate these phonemes. Recently, Raizada et al. (2010) compared the patterns of activation of English and Japanese participants listening to two acoustically-close English syllables, /ra/ and /la/. The separability of brain metabolic patterns predicted subject's behavioral ability to discriminate the two syllables. Altogether these studies show that the neural attunement to native language takes place early during development and continues throughout the life-span, shaping the auditory system to become more efficient in processing phonemes that belong to the native language. Interestingly, Jeng et al. (2011) have shown that this pattern of developmental change is also present at the subcortical level. Compared to Mandarin-speaking infants, adults have stronger subcortical pitch encoding, a lexically-relevant feature to discriminate Mandarin syllables.

Two hypotheses have been proposed to explain this early attunement to native language. The bottom-up hypothesis assumes that infants extract discrete units from continuous speech through statistical learning. For instance, infants' ability to discriminate phonemes seems to heavily rely on statistical distribution of speech sounds in the native language (Maye et al., 2002). By contrast, the top-down hypothesis suggests that learning low-level linguistic units involves higher-level units (i.e. words). According to this view, the English infant would learn to discriminate two similar phonemes (/æ/ and /e/), because they are relevant to discriminate two different words (bad vs. bed).

For many decades, the bottom-up hypothesis of speech processing was predominant, conveying the idea that, as speech sound is processed along the auditory pathway, neural structures' sensitivity to the acoustic content decreases while the sensitivity to abstract features (syllables, words, intelligibility) increases (Okada et al., 2010). In their commentary on Okada's article, Peelle et al. (2010) proposed a hierarchical model of speech processing that starts from Heschl's gyrus

exhibiting high acoustic sensitivity and gradually shows higher acoustic invariance in anterior and posterior temporal regions.

An alternative view to the bottom-up and top-down hypothesis is a more interactive and dynamic model based on interplay between high and low levels of speech representation. Due to the high acoustical variability of real-life speech tokens, phonemic categories exhibit a certain degree of overlap (Hillenbrand et al., 1995), therefore, the bottom-up hypothesis is not sufficient to explain the whole development of phonemic inventory. Indeed, computational studies show that top-down influences are needed to refine phonemes categories with a high degree of accuracy (Fourtassi and Dupoux, 2014). Moreover, Lew-Williams and Saffran (2012) showed that previous exposure to specific word lengths (bi- or tri-syllabic words) influences infants' ability to segment fluent speech. In other words, prior linguistic knowledge builds expectations that influences speech processing in a top-down manner. In the Reverse Hierarchy Theory (RHT), Ahissar and Hochstein (2004) postulate that perceptual learning starts at high-level cortical areas. Then, through long-term exposure to a given context, plasticity would gradually reach lower-level areas, via top-down dynamics. The RHT was originally proposed for visual perception, but has been recently extended to auditory perception (Gutschalk et al., 2008; Suga, 2008). For example, electrical stimulation of the primary auditory cortex modulates activity in subcortical auditory structures such as the inferior colliculus (Gao and Suga, 2000) and the cochlea (Perrot et al., 2006). Together, these studies support the hypothesis that refinement of neuronal representations to native speech sounds is a result of continuous interactions between primary and associative auditory structures and subcortical auditory structures (Kraus and Chandrasekaran, 2010; Tzounopoulos and Kraus, 2009) and are consistent with an emerging view of the auditory system as a distributed, but integrated, circuit (Kraus and White-Schwoch, in press).

The anatomical organization of the auditory system supports these top-down and bottom-up interactions. Peripheral auditory structures such as the cochlea send neural firings from the auditory nerve to the auditory cortex via a series of brainstem nuclei. In addition, central auditory structures such as the primary auditory cortex and associative cortices send back top-down projections to periphery (Kral and Eggermont, 2007). Thus, the neural representation of speech sounds is the result of bottom-up mechanisms that can be modulated via descending cortico-fugal system acting on subcortical structures. According to this interactive model, the auditory midbrain, where afferent and efferent projections converge, presents an excellent model to study the effects of language experience on speech processing.

Research on language-dependent brain plasticity in the subcortical auditory system is an emerging area of study. Krishnan et al. (2005) compared auditory brainstem responses evoked by Mandarin tones in native speakers of Chinese Mandarin and native speakers of American English. They found that Chinese participants have a more robust and faithful representation of the fine pitch variations of Mandarin tones as compared to American participants. Indeed, in Mandarin Chinese, dynamic variations in pitch voice (i.e. the fundamental frequency) provide a major acoustic cue to discriminate two monosyllabic words. For instance, the syllable /yi/ with high-rising pitch contour means "aunt", whereas /yi/ with a high-falling pitch contour means "easy". In contrast, pitch variations in non-tonal languages (e.g. English) are not lexically relevant to discriminate words or syllables; rather they convey supra-lexical information such as stress and intonation patterns (Krishnan and Gandour, 2014). However, in a subsequent study, Krishnan and colleagues used iterated rippled noise (IRN) to simulate Mandarin tones without any speech context, and found that Mandarin speakers exhibited better pitch representation at the subcortical level as compared to American speakers. Thus, these effects may not be necessarily language-specific (Krishnan et al., 2009). Similar to musicians, who, via intensive training, develop outstanding abilities to track the fundamental frequency (i.e. the pitch) of music sounds, Mandarin speakers develop, through long-term exposure to tonal speech sounds, excellent skills to process fine variations of the pitch at the subcortical level (Bidelman et al., 2011). Overall, since tonal languages use qualitatively different phonemic contrasts as compared to non-tonal languages (i.e. pitch contour), it remains unclear whether the differences described above are due to top-down influences of long-term phonemic

representations on subcortical functioning or to a more precise pitch tracking computation, independently of whether the stimulus is part of phonemic inventory of the language system.

The aim of this cross-language study is to determine how far subcortical encoding of speech sounds is sensitive to language experience. Comparing neural responses in native speakers of two non-tonal languages (American English and French), listening to syllables of both languages, gives us the opportunity to study language-dependent plasticity at the subcortical level without confounding factors such as those described above when comparing tonal and non-tonal languages.

We hypothesized that subcortical representations would be more robust and fine-grained for speech sounds that belong to the native phonemic inventory of the listener. In other words, American native listeners should exhibit a more robust and faithful subcortical representation of an American English syllable as compared to French native listeners. Conversely, French listeners should have better representation of a French syllable as compared to non-native listeners. To test this hypothesis we presented to French and American participants a French and an American English syllable, [ru] and [thae], respectively. Both phonemes are "illegal" in the non-native language, that is, both consonant and vowel of each syllable do not exist in the other language. This should maximize differences in long-term memory representations of these two syllables, thus increasing any potential top-down effect of language experience on auditory processing. Because the distinction of consonants and vowels mostly relies on formants properties, we hypothesized that differences on neural responses would be maximal over high frequency components of the EEG spectrum (200-500 Hz) and not at the fundamental frequency.

2. Materials and methods

2.1. Subjects

Twenty-six (18 females and 8 males) adult native speakers of American English and thirty-five (21 females and 14 males) native speakers of French, ranging in age from 18 to 36 years, participated in the study. American participants were recruited at Northwestern University (Chicago, USA) and French participants were recruited at Aix-Marseille University (Marseille, France). Inclusion criteria were a high-school level of education and click-evoked brainstem response latencies within lab-internal normal limits (5.41-5.97 ms, 80 dB sound pressure level, 31/s). The two language groups were matched in term of age (Americans : 22 ± 3 years ; French : 23 ± 3 years ; $F(1,46) = 0.19$, $p = 0.662$) and of musicianship (Americans : 8 ± 5 years; French : 5 ± 6 years of musical practice ; $F(1,46) = 0.87$, $p = 0.355$). The Northwestern University Institutional Review Board and INSERM approved all procedures. Participants gave their informed consent and were paid for their participation.

An ad-hoc questionnaire was used to measure language proficiency both in American English and French. On a scale from 0 (novice) to 10 (expert), participants self-rated their proficiency for oral and written expression, and oral and reading comprehension for both languages (English and French). All subjects reported high proficiency for their native language without any significant difference between groups. However, French participants reported significantly higher proficiency for English than American participants for French (see **Table 1**). This latter point will be further discussed in the discussion. One American participant was bilingual English-Spanish, and one French participant was bilingual French-Vietnamese.

	Skill	French	American	p-value
Native language	Understanding	10 (0)	9.95 (0.22)	0.35
	Reading	10 (0)	9.95 (0.22)	0.35
	Speaking	10 (0)	9.95 (0.22)	0.35
	Writing	10 (0)	9.90 (0.30)	0.06
Non-native language	Understanding	5.74 (2.71)	1.43 (3.09)	<0.001
	Reading	6.65 (2.44)	1.33 (2.88)	<0.001
	Speaking	5.13 (2.44)	1.28 (2.70)	<0.001
	Writing	5.65 (2.53)	0.86 (1.90)	<0.001

Table 1. Native and non-native language proficiency. Mean, standard deviations, and significance values for the French and American groups' self-rated proficiency of their oral and reading comprehension, and oral and written expression.

2.2. Stimuli

The two stimuli used were natural speech syllables, recorded in an anechoic chamber by an American English male speaker and a French male speaker. The French syllable [ʁy] (henceforth ru) and the English syllable [ðæ] (henceforth thae) were chosen because they are both « illegal » speech sounds in the other language, which means that both the consonant and the vowel do not exist in the other language (i.e., [ʁ] and [y] do not exist in English and [ð] and [æ] do not exist in French). Note that while the phoneme [r] exists in both English and French, its realization is very different across languages, with an uvular realization in French and a retroflex realization in English. This choice should maximize the differences between the two languages and should consequently maximize the expected effect of language experience on neural responses (see waveforms in **Figure 1**).

The syllables were matched in duration (209 ms for [ru] stimulus and 210 ms for [thae] stimulus). Both stimuli were natural speech sounds (see **Table 2** for the details on spectral content). While the stimuli have different spectral features, the aim of the study was to determine whether there is an effect of language expertise on stimulus processing (interaction) and not an effect of the stimulus itself.

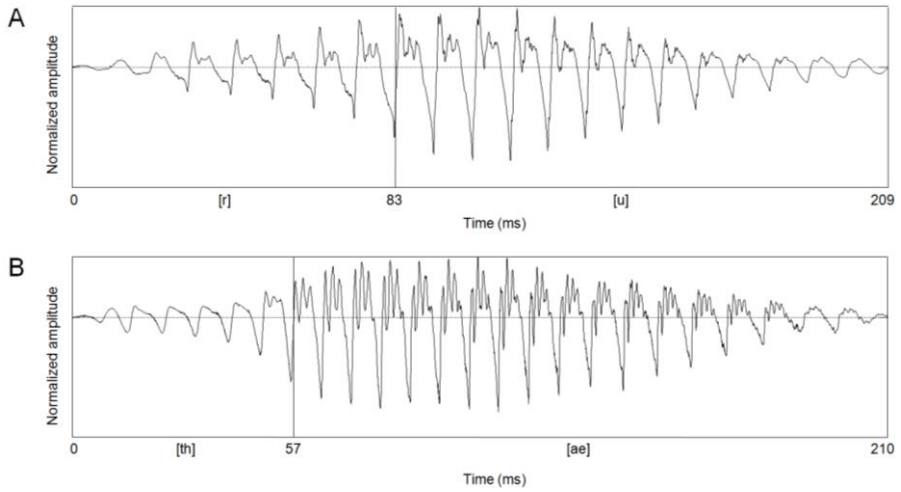

Figure 1. Panel A, waveform of the French stimulus [ru]. Panel B, waveform of the English stimulus [thae]. Vertical black lines indicate the boundaries between the consonants and vowels, as established according to the spectral changes by an experimented phonetician.

Stimulus	Language	Length	Fo	H1	H2	H3
ru	French	209	101	201	302	403
thae	English	210	121	243	364	486

Table 2. Duration (in ms) and frequency peaks (in Hz) of the fundamental frequency (Fo) and the first three harmonics for the two stimuli.

Each stimulus was presented monaurally to the right ear at 80 dB SPL at a rate of 3.8 Hz with alternating polarities through magnetically shielded insert earphones (ER-3A, Etymotic Research) using the stimulus presentation software Microvitae (µV-ABR, Microvitae Technologies).

2.3. Electrophysiological recording

During electrophysiological recordings, participants sat in a comfortable reclining chair in an electrically-shielded, sound-attenuated room and were instructed to watch a subtitled movie of their choice to maintain relaxation and prevent drowsiness. Brain responses were collected at 30 kHz sampling rate using Microvitae recording system (µV-ABR, Microvitae Technologies) with three Ag-AgCl scalp electrodes in a vertical montage (Cz active, forehead ground, and right earlobe reference). Electrode impedances were kept <5 KΩ. Six-thousand sweeps were collected for each stimulus (two blocks of three-thousand sweeps). The order of presentation of the two stimuli was counterbalanced across participants. One of the authors (BI) was in charge of data acquisition in both countries using the same portable EEG system. This prevents the possibility of having a bias due to different experimental setups, participant preparation, and instruction.

2.4. Data analysis

All analyses were performed using custom MATLAB scripts (MathWorks). First, electrophysiological recordings were bandpass filtered from 70 to 2000 Hz (12 dB/octave roll-off) using a Butterworth filter. Then, sweeps with activity exceeding ± 30 µV were rejected as artifacts and the responses were baseline-corrected to the pre-stimulus period (-30 to 0 ms). Neural responses were then averaged over a -30 to 229 ms window for [ru] stimulus and -30 to 230 ms for [thae] stimulus. The signal-to-noise ratio (SNR) was computed using the quotient of response root mean square (RMS) amplitude and pre-stimulus baseline RMS amplitude (see Skoe and Kraus, 2010). If the SNR was less or equal to 1.4 for one or both stimuli, the participant was excluded. This resulted in excluding 5 American and 8 French participants.

2.4.1. Spectral amplitude

Fast Fourier transform was performed on two time regions of the response (whole response and vowel). These time regions were defined on the basis of the stimuli by a phonetician also taking into account a 10 ms neural delay in the response: whole response (10-220 ms for both stimuli) and vowel (93-220 and 67-220 ms for the French and English stimuli respectively). Responses to the consonant alone were not analyzed due to both the difficulty to define a precise end of the consonantal features and to the short duration that implied poorer estimates of spectral features and phase coherence. The maximum spectral amplitudes were extracted in a bandwidth of 10% surrounding the frequencies of interest (e.g. for a peak at 103 Hz, values were extracted between 98 and 108 Hz). Frequencies of interest included the fundamental frequency (Fo) and its three subsequent integer harmonics H1-H3 (whole integer multiples of the Fo). For the harmonics, the

three maximum values were then averaged to form a global measure of the harmonics' representation.

2.4.2. Inter-trial phase-coherence

We used the same procedure described in Tierney and Kraus (2013). This technique measures the phase consistency across trials of each frequency component in the neural responses. To summarize, for both time regions (whole response and vowel) a fast Fourier transform was performed on each trial that resulted in two values, the amplitude and the phase for each frequency component. Since we were interested in phase variability across trials, only the phase values were kept. The vector's length of each frequency was computed using the Matlab toolbox CircStat Version 2012a (Berens, 2009). The length of the resultant vector represents the phase-coherence across trials for each frequency component. This measure ranges from 0 (no phase coherence) to 1 (perfect phase coherence). Finally, the maximum phase coherences were picked in a bandwidth of 10% surrounding the frequencies of interest (e.g. for a peak at 103 Hz, values were extracted between 98 and 108 Hz). For the harmonics, the three values were then averaged to form a global phase coherence measure of the harmonics.

2.5. Statistical analyses

All statistical analyses were performed using Statistica Version 7.1 (StatsSoft, Tulsa, OK). Repeated measure analyses of variance (RMANOVA) were used for group (American vs. French) x stimulus ([ru] vs. [thae]) comparisons for spectral representation and inter-trial phase-coherence. *Post-hoc* tests were used when appropriate (Fisher LSD). Because of their non-normal distribution phase-coherence values were Fisher z-transformed before statistical analyses (Mann-Whitney U test was used as a post-hoc test).

3. Results

3.1. Spectral representation

3.1.1. Fundamental frequency (Fo)

Neural responses were divided into two time regions, corresponding to the response to the entire stimulus and the response to the vowel only. Across both stimuli (thae/ru), American participants had larger spectral amplitudes in response to the fundamental frequency (main effect of group, $F(1,40) = 6.90$, $p = 0.012$). The Americans showed stronger representation of the fundamental frequency for both [ru] and [thae] (i.e. there was no group x stimulus interaction, $F(1,40) = 0.03$, $p = 0.861$).

Analyses of the vowel time region, confirmed that American participants had larger spectral amplitudes in response to the fundamental frequency (main effect of group, $F(1,38) = 2.20$, $p = 0.146$). This difference was equivalent across stimuli (no group x stimulus interaction, $F(1,38) = 0.007$, $p = 0.93$, **Figure 2**).

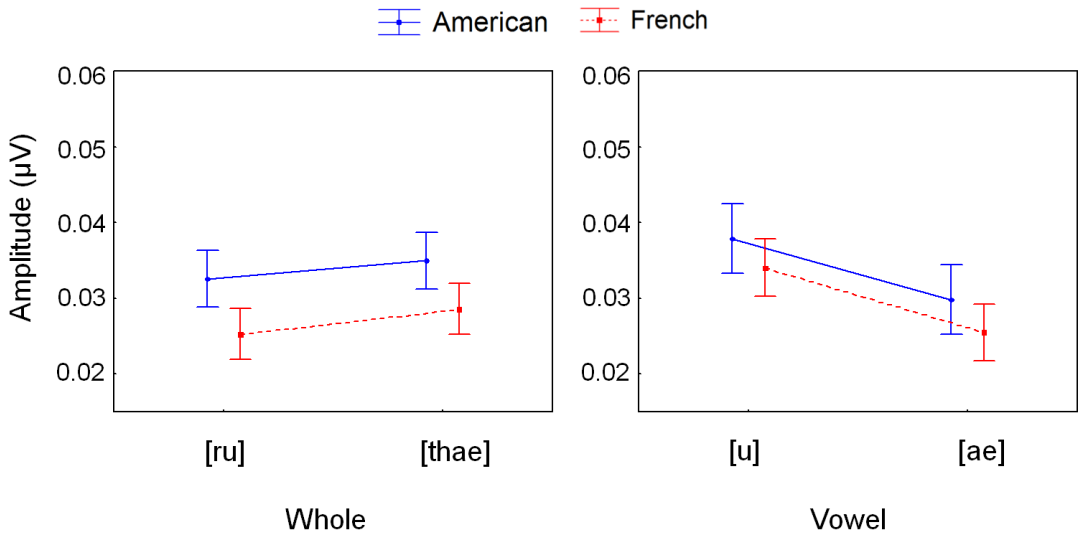

Figure 2. Spectral representation of the fundamental frequency (mean ± 95% confidence intervals). American participants demonstrated enhanced spectral encoding of the Fo in the neural responses to the whole stimulus and the vowel.

3.1.2. Harmonics

Across both stimuli (thae/ru), the magnitude of the responses to the harmonics was equivalent for both groups (main effect of group, $F(1,38) = 1.05$, $p = 0.312$). However, stimuli neural representations differed as a function of native language (group x stimulus interaction, $F(1,38) = 6.18$, $p = 0.017$, **Figure 3).** Post-hoc tests revealed more robust spectral representations of the harmonics of the stimulus [ru] for French compared to American native listeners ($p = 0.040$) whereas for the stimulus [thae] there was no significant difference ($p = 0.765$).

Analyses of the vowel time region did not show group differences for either stimuli (main effect of group, $F(1,38) = 0.91$, $p = 0.346$; group x stimulus interaction, $F(1,38) = 0.65$, $p = 0.425$).

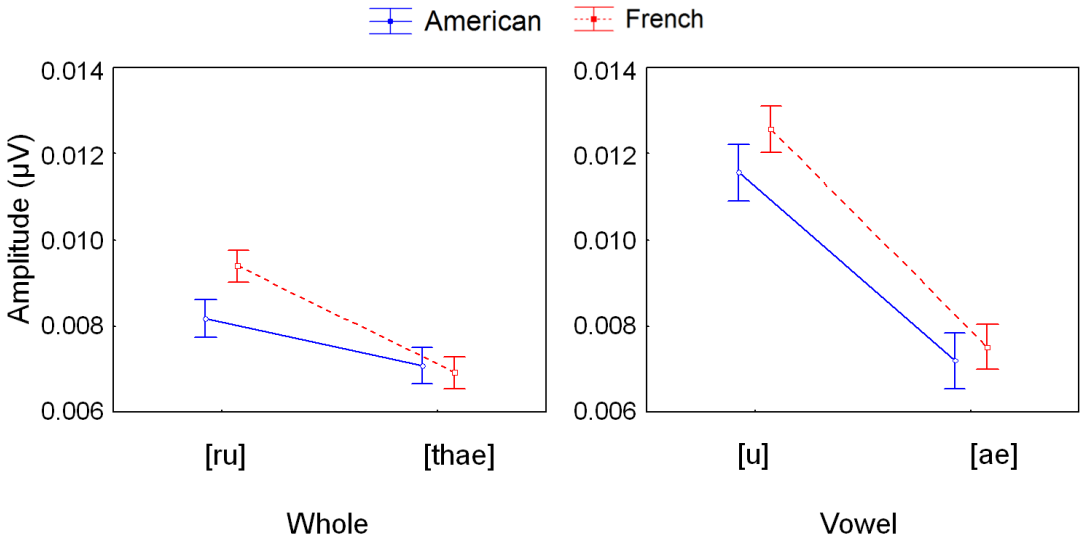

Figure 3. Spectral representation of the first three harmonics (mean H1-3 ± 95% confidence intervals). There was a group x stimulus interaction ($p = 0.017$) in the neural responses to the whole stimulus mainly driven by greater spectral encoding of the [ru] stimulus harmonics in French neural responses ($p < 0.05$).

3.2. Inter-trial phase-coherence

3.2.1. Fundamental frequency (Fo)

The analysis of the whole response showed that fundamental frequency phase-coherence was equivalent across group for either stimuli (main effect of group, $F(1,38) = 0.04$, $p = 0.832$; group x stimulus interaction, $F(1,38) = 0.77$, $p = 0.385$).

The analysis of the vowel did not reveal any significant effect (main effect of group, $F(1,38) = 0.40$, $p = 0.529$, group x stimulus interaction, $F(1,38) = 0.22$, $p = 0.643$).

3.2.2. Harmonics

The analysis of the whole response showed that harmonics phase-coherence was equivalent across group for either stimuli (main effect of group, $F(1,36) = 2.18$, $p = 0.149$; group x stimulus interaction, $F(1,36) = 0.256$, $p = 0.616$).

The analysis of the vowel revealed that the phase-coherence of the responses to the harmonics was equivalent for both groups (main effect of group, $F(1,39) = 0.65$, $p = 0.424$). However, inter-trial phase-coherence differed as a function of native language (group x stimulus interaction, $F(1,39) = 4.32$, $p = 0.044$, **Figure 4).** Post-hoc tests revealed that, for the French stimulus [ru], French native listeners showed more robust representations of stimulus' harmonics compared to American native listeners ($p = 0.054$), while group differences were not visible for the stimulus [thae] ($p = 0.830$).

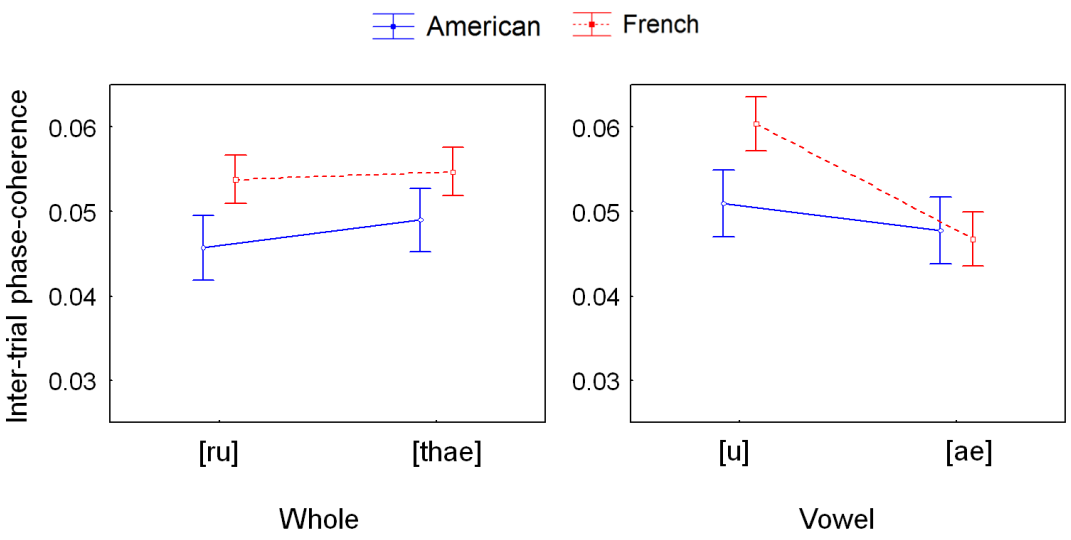

Figure 4. Inter-trial phase-coherence of the first three harmonics (mean H1-3 ± 95% confidence intervals). There was a group x stimulus interaction ($p = 0.044$) in the neural responses to the vowel, mostly driven by more robust encoding of the harmonics of the vowel [u] in the neural responses of French participants ($p = 0.054$).

4. Discussion

The goal of this study was to test the hypothesis that language experience shapes neural representations of speech sounds. More precisely, subcortical representation should be more accurate for speech sounds that belong to the native phonemic inventory of the listener. To test this hypothesis, we recorded speech-evoked subcortical responses in American and French native

speakers using an American English and a French syllable. Importantly, both consonants and vowels of these two syllables - [thae] and [ru] - do not exist in the other language, which means that for each participant language experience should be maximal with one sound and minimal with the other sound. Taken together, results are consistent with the hypothesis that language experience shapes the neural processing of speech, and that this plasticity occurs as a function of what is behaviorally-relevant to a listener. Importantly, the legacy of this linguistic experience was apparent during a passive listening task, suggesting that language experience sculpts automatic response properties of auditory nuclei. We will focus our discussion on two main findings. Firstly, independently of the stimulus type, American participants showed a greater subcortical representation of the fundamental frequency (Fo) compared to French participants. Secondly, participants exhibited more robust encoding and more precise spectral representations when listening to the syllable of their native language as compared to non-native language.

4.1. Effect of language expertise on Fo representation

The subcortical representation of the fundamental frequency (Fo) was enhanced in American, compared to French, participants across both stimuli. Since previous research has shown that musical expertise can have an effect on representation of the Fo, we carefully verified that American and French participants were matched in term of musicianship (see Materials and methods). Then, the global enhancement in subcortical encoding of the Fo in the American group may be related to differences in language experience, particularly the role that Fo cues play in American English as compared to French. Indeed, although English is not considered as a tonal language, it is nonetheless characterized by a large range of pitch dynamics akin to tonal languages (Duanmu, 2004). By contrast, French exhibits less variability in pitch (Fo) at the utterance level and is classified as a non-tonal and non-stress language at the word level (Braun et al., 2014; Vaissière, 1991). Moreover, pitch carries segmental information in American English while it does not in French, which could explain the subcortical strengthening of this stimulus feature in American listeners independently of whether the stimulus belongs to their native language or not.

Interestingly, in a cross-language experiment, Braun et al. (2014) tested whether the complexity of the pitch system in the native language modulates encoding of non-native tonal speech sounds. For instance, native speakers of a non-stress language (French), after learning associations between pictures and non-words distinguished only by their tonal contrasts, exhibited more difficulties remembering these associations than native speakers of a stress language (German). These results suggest that languages without stress at the word level (e.g. French) are less sensitive to tonal contrasts (i.e. pitch variations) as compared to stress languages. Dupoux et al. (1997) have shown that French listeners exhibited significantly more difficulties than Spanish listeners to discriminate words that differ only by their accent. This reduced sensitivity to stress patterns, referred to as "stress deafness", could result from the difficulty for French listeners to represent stress at the phonological level (Dupoux et al., 2008). Since Fo variations are a major marker of stress, the findings of a reduced sensitivity to stress patterns go well along with our findings of a poorer representation of Fo in French participants.

4.2. Specific neural enhancement of native-language sounds

Learning a language requires the ability to discriminate subtle differences in the phonemic inventory. This specialization for sounds of the native language may take place at the detriment of phonemes of other languages (Werker and Tees, 2002). Our results showed a significant interaction of language expertise and stimulus type when considering the first three harmonics of the neural responses. That the interaction was visible on the harmonics rather than on the fundamental frequency is consistent with our predictions because high-frequency spectral components are particularly relevant to define phonetic characteristics.

Importantly, this interaction was significant in both analyses of spectral density (i.e. the amplitude of the signal in a spectral representation) and inter-trial phase-coherence (the stability of phase over trials). More precisely, for spectral density this was visible on the response to the entire syllable duration, while for inter-trial phase-coherence this interaction occurred in response to the vowel only. The finding that differences in inter-trial phase-coherence are mostly driven by vowels is not so surprising. Indeed, while different acoustic features allow phoneme discrimination, subcortical responses are typically elicited either to periodic cues such as vowels or to stop consonants with a sharp onset. Thus, there may not be very strong or reliable phase-locking going on during fricative or liquid consonant time regions. This would explain a greater differences of inter-trial phase-coherence during the vowel compared to the entire syllable.

Although differences are also present for the stimulus [thae], the language by stimulus interaction was driven by the group differences for the stimulus [ru]. This could be due to the fact that French subjects are necessarily more familiar with English speech sounds than are American subjects with French speech sounds, since English is mandatory in France at school from the age of 12. Indeed, French subjects reported a better proficiency for English than American for French (See Material and Methods, **Table 1**). According to Song et al. (2008), subcortical plasticity can occur even in adults after short-term auditory training (see also Carcagno and Plack, 2011; Chandrasekaran et al., 2012). These results suggest that exposure to English language in French schools could induce subcortical plasticity effects that may reduce group differences for the English syllable. Put differently, the need for French listeners to distinguish French and English could accentuate the contrast in language-dependent processing.

Overall, this study reveals that language-dependent effects do not result in a global enhancement of subcortical encoding of native speech sounds, but rather in a strengthening of specific stimulus features that are linguistically-relevant (in this case, the high frequency components that are critical to discriminate vowels). This is in line with previous studies on experience-dependent plasticity that emphasize that it occurs along dimensions that are behaviorally-relevant to an individual (Kraus and White-Schwoch, in press). For instance, long-term musical practice strengthens subcortical encoding of specific stimulus features of music sounds (Lee et al., 2009) and speech sounds (Parbery-Clark et al., 2012) that may also have behavioral relevance. Even more relevant to the goal of this study, Krishnan et al. (2005) have shown that long-term language experience does not induce an overall enhancement of stimulus processing, but rather a specific strengthening of stimulus features that are linguistically-relevant (see also Strait et al., 2009).

The mechanisms behind these language-dependent effects likely relate both to bottom-up and top-down hypotheses.

From birth and even before, the infant is exposed to sounds, some of which are more frequent and relevant than others. More precisely, native speech sounds are more prevalent in the infant environment than non-native speech sounds. Thus, development of speech perception is, at first, essentially a bottom-up mechanism whereby neural representations along the auditory pathway are shaped in response to the statistical distribution of stimuli in the external world. With growing exposure to the mother tongue, the infant develops high-level representations of linguistic units (phonemes) that become resistant to the inherent variability present in regular speech—that is, different utterances of the same syllable [ru] are categorized as the same syllable. Once lexical representations become stable, one can make predictions about the upcoming words and syllables (e.g. Nina plays the guitar and I play the pia.....[no]). These effects are likely to stem from a cascade of top-down processes that enhance linguistically-relevant, and prune non-relevant, information. This means that the relevant auditory stimulus features (e.g. first and second formants) maybe also anticipated. One direct consequence of anticipating upcoming events is that attention can be more efficiently directed towards these specific stimulus features and this will in turn render neural representations more robust (Fritz et al., 2010). Interestingly, the impact of attention on neural activity is not limited to cortical structures, but can indeed extend to peripheral auditory structures (Perrot et al., 2006).

Overall, our results can be explained both in terms of bottom-up and top-down processes. Long-term exposure to the mother tongue implies that a phoneme of the native language will be heard a huge amount of times. On the other side, the existence in each language of a limited phonemic inventory and lexicon allows one to predict auditory events and focus attention to relevant stimulus features. These two explanations may possibly account for more accurate subcortical representations of native language stimulus features.

References

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8, 457–464. doi:10.1016/j.tics.2004.08.011

Berens, P., 2009. CircStat: A MATLAB Toolbox for Circular Statistics. J. Stat. Softw. 31, 1–21.

Bidelman, G.M., Gandour, J.T., Krishnan, A., 2011. Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. Brain Cogn. 77, 1–10. doi:10.1016/j.bandc.2011.07.006

Braun, B., Galts, T., Kabak, B., 2014. Lexical encoding of L2 tones: The role of L1 stress, pitch accent and intonation. Second Lang. Res. 30, 323–350.

Carcagno, S., Plack, C.J., 2011. Subcortical Plasticity Following Perceptual Learning in a Pitch Discrimination Task. J. Assoc. Res. Otolaryngol. 12, 89–100. doi:10.1007/s10162-010-0236-1

Chandrasekaran, B., Kraus, N., Wong, P.C.M., 2012. Human inferior colliculus activity relates to individual differences in spoken language learning. J. Neurophysiol. 107, 1325–1336. doi:10.1152/jn.00923.2011

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., Näätänen, R., 1998. Development of language-specific phoneme representations in the infant brain. Nat. Neurosci. 1, 351–353.

Dehaene-Lambertz, G., Dupoux, E., Gout, A., 2000. Electrophysiological Correlates of Phonological Processing: A Cross-linguistic Study. J. Cogn. Neurosci. 12, 635–647. doi:10.1162/089892900562390

Duanmu, S., 2004. Tone and non-tone languages: An alternative to language typology and parameters. Lang. Linguist. 5, 891–923.

Dupoux, E., Pallier, C., Sebastian, N., Mehler, J., 1997. A destressing “deafness” in French? J. Mem. Lang. 36, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., Peperkamp, S., 2008. Persistent stress “deafness”: The case of French learners of Spanish. Cognition 106, 682–706. doi:10.1016/j.cognition.2007.04.001

Fourtassi and Dupoux, 2014. A Rudimentary Lexicon and Semantics help Bootstrap Phoneme Acquisition 191–200.

Fritz, J.B., David, S.V., Radtke-Schuller, S., Yin, P., Shamma, S.A., 2010. Adaptive, behaviorally gated, persistent encoding of task-relevant auditory information in ferret frontal cortex. Nat. Neurosci. 13, 1011–1019. doi:10.1038/nn.2598

Gao, E., Suga, N., 2000. Experience-dependent plasticity in the auditory cortex and the inferior colliculus of bats: Role of the corticofugal system. Proc. Natl. Acad. Sci. 97, 8081–8086. doi:10.1073/pnas.97.14.8081

Gutschalk, A., Micheyl, C., Oxenham, A.J., 2008. Neural Correlates of Auditory Perceptual Awareness under Informational Masking. PLoS Biol 6, e138. doi:10.1371/journal.pbio.0060138

Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K., 1995. Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. 97, 3099–3111. doi:10.1121/1.411872

Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. ’ich, Kettermann, A., Siebert, C., 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition 87, B47–B57. doi:10.1016/S0010-0277(02)00198-1

Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., Lin, C.-D., 2011. Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. Ear Hear. 32, 699–707.

Kral, A., Eggermont, J.J., 2007. What's to lose and what's to learn: Development under auditory deprivation, cochlear implants and limits of cortical plasticity. Brain Res. Rev. 56, 259–269. doi:10.1016/j.brainresrev.2007.07.021

Kraus, N., Chandrasekaran, B., 2010. Music training for the development of auditory skills. Nat. Rev. Neurosci. 11, 599–605. doi:10.1038/nrn2882

Kraus, N., White-Schwoch, T., in press. Unraveling the biology of auditory learning: A cognitive-sensorimotor-reward framework. Trends Cogn. Sci.

Krishnan, A., Gandour, J.T., 2014. Language experience shapes processing of pitch relevant information in the human brainstem and auditory cortex: Electrophysiological evidence. Acoust. Aust. Aust. Acoust. Soc. 42, 166–178.

Krishnan, A., Gandour, J.T., Bidelman, G.M., Swaminathan, J., 2009. Experience-dependent neural representation of dynamic pitch in the brainstem: NeuroReport 20, 408–413. doi:10.1097/WNR.0b013e3283263000

Krishnan, A., Xu, Y., Gandour, J., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Cogn. Brain Res. 25, 161–168. doi:10.1016/j.cogbrainres.2005.05.004

Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. Nat. Rev. Neurosci. 5, 831–843. doi:10.1038/nrn1533

Kuhl, P.K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., Iverson, P., 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev. Sci. 9, F13–F21. doi:10.1111/j.1467-7687.2006.00468.x

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.

Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective Subcortical Enhancement of Musical Intervals in Musicians. J. Neurosci. 29, 5832–5840. doi:10.1523/JNEUROSCI.6133-08.2009

Lew-Williams, C., Saffran, J.R., 2012. All words are not created equal: Expectations about word length guide infant statistical learning. Cognition 122, 241–246. doi:10.1016/j.cognition.2011.10.007

Maye, J., Werker, J.F., Gerken, L., 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, B101–B111. doi:10.1016/S0010-0277(01)00157-3

Mehler, J., Dupoux, E., Pallier, C., Dehaene-Lambertz, G., 1994. Cross-linguistic approaches to speech processing. Curr. Opin. Neurobiol. 4, 171–176. doi:10.1016/0959-4388(94)90068-X

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.-H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical Organization of Human Auditory Cortex: Evidence from Acoustic Invariance in the Response to Intelligible Speech. Cereb. Cortex 20, 2486–2495. doi:10.1093/cercor/bhp318

Ortiz-Mantilla, S., Hämäläinen, J.A., Musacchia, G., Benasich, A.A., 2013. Enhancement of Gamma Oscillations Indicates Preferential Processing of Native over Foreign Phonemic Contrasts in Infants. J. Neurosci. 33, 18746–18754. doi:10.1523/JNEUROSCI.3260-13.2013

Parbery-Clark, A., Anderson, S., Hittner, E., Kraus, N., 2012. Musical experience strengthens the neural representation of sounds important for communication in middle-aged adults. Front. Aging Neurosci. 4. doi:10.3389/fnagi.2012.00030

Peelle, J.E., 2010. Hierarchical processing for speech in human auditory cortex and beyond. Front. Hum. Neurosci. doi:10.3389/fnhum.2010.00051

Perrot, X., Ryvlin, P., Isnard, J., Guénot, M., Catenoix, H., Fischer, C., Mauguière, F., Collet, L., 2006. Evidence for Corticofugal Modulation of Peripheral Auditory Activity in Humans. Cereb. Cortex 16, 941–948. doi:10.1093/cercor/bhj035

Raizada, R.D.S., Tsao, F.-M., Liu, H.-M., Kuhl, P.K., 2010. Quantifying the Adequacy of Neural Representations for a Cross-Language Phonetic Discrimination Task: Prediction of Individual Differences. Cereb. Cortex 20, 1–12. doi:10.1093/cercor/bhp076

Rivera-Gaxiola, M., Silva-Pereyra, J., Kuhl, P.K., 2005. Brain potentials to native and non-native speech contrasts in 7- and 11-month-old American infants. Dev. Sci. 8, 162–172. doi:10.1111/j.1467-7687.2005.00403.x

Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31, 302–324. doi:10.1097/AUD.0b013e3181cdb272

Song, J.H., Skoe, E., Wong, P.C.M., Kraus, N., 2008. Plasticity in the Adult Human Auditory Brainstem following Short-term Linguistic Training. J. Cogn. Neurosci. 20, 1892–1902. doi:10.1162/jocn.2008.20131

Strait, D.L., Kraus, N., Skoe, E., Ashley, R., 2009. Musical experience and neural efficiency - effects of training on subcortical processing of vocal expressions of emotion. Eur. J. Neurosci. 29, 661–668. doi:10.1111/j.1460-9568.2009.06617.x

Suga, N., 2008. Role of corticofugal feedback in hearing. J. Comp. Physiol. A 194, 169–183. doi:10.1007/s00359-007-0274-2

Tierney, A., Kraus, N., 2013. The Ability to Move to a Beat Is Linked to the Consistency of Neural Responses to Sound. J. Neurosci. 33, 14981–14988. doi:10.1523/JNEUROSCI.0612-13.2013

Tzounopoulos, T., Kraus, N., 2009. Learning to Encode Timing: Mechanisms of Plasticity in the Auditory Brainstem. Neuron 62, 463–469. doi:10.1016/j.neuron.2009.05.002

Vaissière, J., 1991. Rhythm, accentuation and final lengthening in French 108–120.

Werker, J.F., Tees, R.C., 2002. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behav. Dev., 25th Anniversary Special Issue 25, 121–133. doi:10.1016/S0163-6383(02)00093-0

Zhang, Y., Kuhl, P.K., Imada, T., Kotani, M., Tohkura, Y. 'ichi, 2005. Effects of language experience: Neural commitment to language-specific auditory patterns. NeuroImage 26, 703–720. doi:10.1016/j.neuroimage.2005.02.040