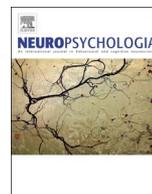




ELSEVIER

Contents lists available at ScienceDirect

## Neuropsychologia

journal homepage: [www.elsevier.com/locate/neuropsychologia](http://www.elsevier.com/locate/neuropsychologia)

## Native language shapes automatic neural processing of speech

Bastien Intartaglia<sup>a,b,\*</sup>, Travis White-Schwoch<sup>c</sup>, Christine Meunier<sup>b,d</sup>, Stéphane Roman<sup>e</sup>,  
Nina Kraus<sup>c,f,g</sup>, Daniele Schön<sup>a,b</sup><sup>a</sup> Aix-Marseille Université, INSERM, Institut de Neurosciences des Systèmes, Marseille, France<sup>b</sup> Brain and Language Research Institute, Labex BLRI, France<sup>c</sup> Auditory Neuroscience Laboratory and Department of Communication Sciences, Northwestern University, Evanston, IL, United States of America<sup>d</sup> Aix-Marseille Université, CNRS, LPL UMR 7309, 13100 Aix-en-Provence, France<sup>e</sup> La Timone Children's Hospital, ENT Unit, Marseille, France<sup>f</sup> Department of Neurobiology & Physiology, Northwestern University, Evanston, IL, United States of America<sup>g</sup> Department of Otolaryngology, Northwestern University, Chicago, IL, United States of America

## ARTICLE INFO

## Article history:

Received 12 October 2015

Received in revised form

30 May 2016

Accepted 31 May 2016

Available online 1 June 2016

## Keywords:

Experience-dependent plasticity

Auditory brainstem responses

Speech perception

## 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 attunement to native language. However, most 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 first formant when listening to the syllable of their native language as compared to non-native language. These results align with the hypothesis that language experience shapes sensory processing of speech and that this plasticity occurs as a function of what is meaningful to a listener.

© 2016 Elsevier Ltd. All rights reserved.

## 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 as poor 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-dependent reorganization of the phonemic inventory relies 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).

\* Correspondence to: Institut de Neurosciences des Systèmes, Inserm UMR1106, Aix-Marseille Université, Faculté de Médecine, 27, Boulevard Jean Moulin, 13005 Marseille, France.

E-mail address: [bastien.intartaglia@etu.univ-amu.fr](mailto:bastien.intartaglia@etu.univ-amu.fr) (B. Intartaglia).

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, an index of slower cortical activity occurring < 10 ms) amplitude drops significantly for non-native phonemes. Likewise, in a longitudinal study, Rivera-Gaxiola et al. (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 130 ms 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 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 rely heavily on the statistical distribution of speech sounds in the native language (Maye et al., 2002). In contrast, the top-down hypothesis suggests that learning low-level linguistic units involves higher-level units (i.e. words) (Fourtassi and Dupoux, 2014). 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 a commentary on Okada's article, Peelle (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 a 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, 2015).

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 the 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, focusing on faster neural activity occurring < 1 ms, 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 voice pitch (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 in subcortical systems (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 processing 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. This hypothesis is precisely tested by the interaction term of a factorial design: 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 the formants’ properties, we hypothesized that differences on neural responses would be maximal over high frequency components of the EEG spectrum (200–800 Hz) such as harmonics at formant frequencies but not at the fundamental frequency.

## 2. Materials and methods

### 2.1. Participants

Twenty-six (18 females and 8 males) adult native speakers of American English and 35 (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 (Wave V latency 5.41–5.97 ms in response to a click presented at 80 dB sound pressure level, rate of 31/s). To make sure that there were no different hearing thresholds, we compared wave V latencies and amplitudes between groups. Analyses showed that groups were matched for both measures (wave V latency:  $U=250$ ,  $p=0.633$ ; wave V amplitude:  $U=252$ ,  $p=0.664$ ). The two language groups were also matched in term of age (Americans:  $22 \pm 3$  years; French:  $23 \pm 3$  years;  $F(1,46)=0.19$ ,  $p=0.662$ ) and of musicianship (Americans:  $8 \pm 5$  years; French:  $5 \pm 6$  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.

**Table 1.**

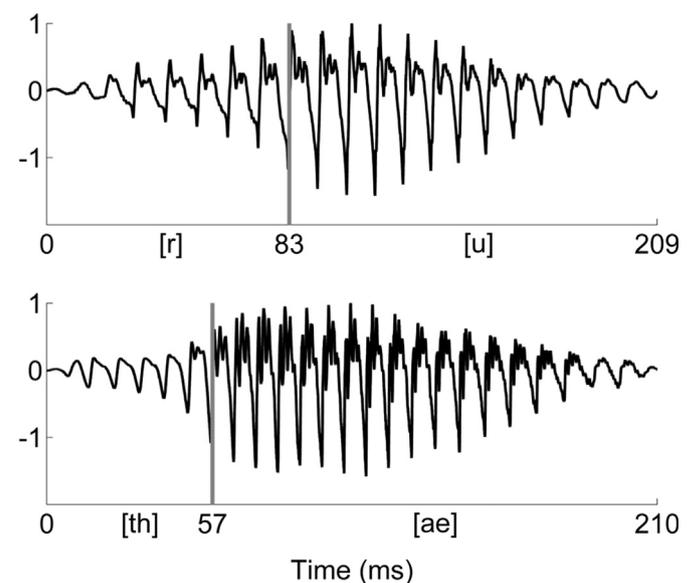
Native and non-native language proficiency. Mean, standard deviations, and significance values for the French and American groups’ self-rated proficiency of their oral and reading comprehension, and oral and written expression.

	Skill	French	American	p-Value
<b>Native language</b>	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
<b>Non-native language</b>	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

### 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 [ɾy] (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., [ɾ] 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 Fig. 1).

The syllables were matched in duration (209 ms for [ru] stimulus and 210 ms for [thae] stimulus). Frequencies of interest included the fundamental frequency ( $F_0$ ) and the second to sixth harmonics ( $H_2$ – $H_6$ ). These latter were chosen because they encompass the first formant range ( $F_1$ ) which is an acoustically relevant cue in discriminating vowels (see Table 2 for more details). Since the stimuli were natural speech sounds, the  $F_0$  and its subsequent harmonics  $H_2$ – $H_6$  vary slightly over time and were thus estimated by computing a fast Fourier transform (FFT) on three time regions (whole stimuli, consonant and vowel). These



**Fig. 1.** Waveforms of the two stimuli (normalized amplitudes). Top, waveform of the French stimulus [ru]. Bottom, waveform of the English stimulus [thae]. Vertical grey lines indicate the boundaries between the consonants and vowels, as established according to the spectral changes by an experienced phonetician.

**Table 2.**

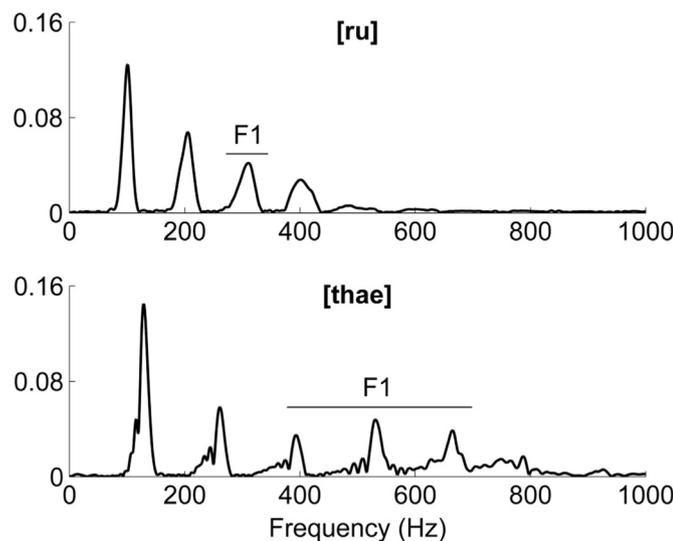
Frequencies of interest (FOI) for each time region and stimulus. Values represent the frequency peaks of the fundamental frequency (F0), its second through sixth subsequent harmonics (H2–H6) and frequency range of the first formant (F1 range) for each time region and stimulus (in Hz). Note that because stimuli are natural, FOI vary slightly over time and harmonics are not exactly whole integers of the F0.

Stimulus	FOI	Whole	Consonant	Vowel
[ru]	F0	101	93	105
	H2	205	188	210
	H3	311	279	316
	H4	401	373	421
	H5	487	461	528
	H6	590	549	632
	F1 range	N/A	N/A	242–298
[thae]	F0	129	118	125
	H2	261	235	258
	H3	393	355	386
	H4	531	472	528
	H5	665	590	665
	H6	787	711	788
	F1 range	N/A	N/A	420–705

time regions were established according to the spectral changes by an experienced phonetician. Importantly the ranges of the first two vowel formants ([u]:  $F1=242\text{--}298$  Hz,  $F2=1449\text{--}1732$  Hz; [ae]:  $F1=420\text{--}705$  Hz,  $F2=1408\text{--}1556$  Hz, from Praat: Boersma and Weenink, 2001) showed that for both stimuli F2 range exceeded the phase-locking limit of the auditory brainstem which is around 1.5 kHz and was thus not taken into consideration in further analyses (Aiken and Picton, 2008; Moushegian et al., 1973).

It is important to make clear that, while the two stimuli have different spectral features (see FFT in Fig. 2 and Table 2), the aim of the study was to determine whether there would be an effect of language expertise on stimulus processing. This is reflected in the interaction term of the language expertise and stimulus factors. As such, a main effect of stimulus would not be particularly meaningful here because it would simply reflect the sensitivity of the neural system to acoustic differences in the 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 ( $\mu\text{V-ABR}$ , Microvitae Technologies).



**Fig. 2.** Fast Fourier transforms computed on the whole stimuli (normalized amplitudes). Top, French stimulus [ru]. Bottom, English stimulus [thae]. Horizontal lines indicate the range of the first formant.

### 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 ( $\mu\text{V-ABR}$ , Microvitae Technologies) with three Ag-AgCl scalp electrodes in a vertical montage (Cz active, forehead ground, and right earlobe reference). Electrode impedances were kept  $< 5$  k $\Omega$ . Six-thousand sweeps were collected for each stimulus (two blocks of 3000 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  $\pm 30$   $\mu\text{V}$  were rejected as artifacts and the responses were baseline-corrected to the pre-stimulus period ( $-30\text{--}0$  ms). Neural responses were then averaged over a  $-30\text{--}229$  ms window for [ru] stimulus and  $-30\text{--}230$  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

The spectral energy of the neural responses was analyzed by computing Fast Fourier transform on three time regions of the response (whole response, consonant 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), consonant (10–93 and 10–67 ms for the French and English stimuli respectively) and vowel (93–220 and 67–220 ms for the French and English stimuli respectively). For each time region, the maximum spectral amplitudes of the fundamental frequency (F0) and its second to sixth harmonics (H2–H6) were extracted in a bandwidth of 40 Hz surrounding the peak in the stimulus fast Fourier transform (e.g. for a peak at 105 Hz, values were extracted between 85 and 125 Hz). For the whole response and the consonant, the five values extracted from H2 to H6 were then averaged to form a global measure of harmonics' representation. For the vowel, since our hypothesis was that differences on neural responses would be maximal over frequencies that are relevant to discriminate vowels, harmonics were separated in two groups. Harmonics falling within F1 range  $\pm 20$  Hz were averaged to form a global measure of F1 spectral representation, while harmonics falling outside F1 range were averaged to form a global measure of non-formant spectral representation. For the vowel [u] the F1 range  $\pm 20$  Hz included only the third harmonic (H3), and the remaining harmonics were considered as non-formant harmonics (H2, H4–H6). For the vowel [ae] the F1 range  $\pm 20$  Hz included the third to fifth harmonics (H3–H5), and the remaining harmonics were considered as non-formant frequencies (H2, H6).

#### 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 the three time regions (whole response, consonant 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). For each time region, the maximum phase-coherence values of the fundamental frequency ( $F_0$ ) and its second to sixth harmonics ( $H_2$ – $H_6$ ) were extracted in a bandwidth of 40 Hz surrounding the frequencies of interest. Then, the same procedure described above for spectral analyses was used for the whole, the consonant and the vowel time regions.

## 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)  $\times$  stimulus ([ru] vs. [thae]) comparisons for spectral representation and inter-trial phase-coherence. *Post-hoc* tests were used when appropriate (Fisher LSD).

## 3. Results

### 3.1. Spectral representation

#### 3.1.1. Whole stimuli

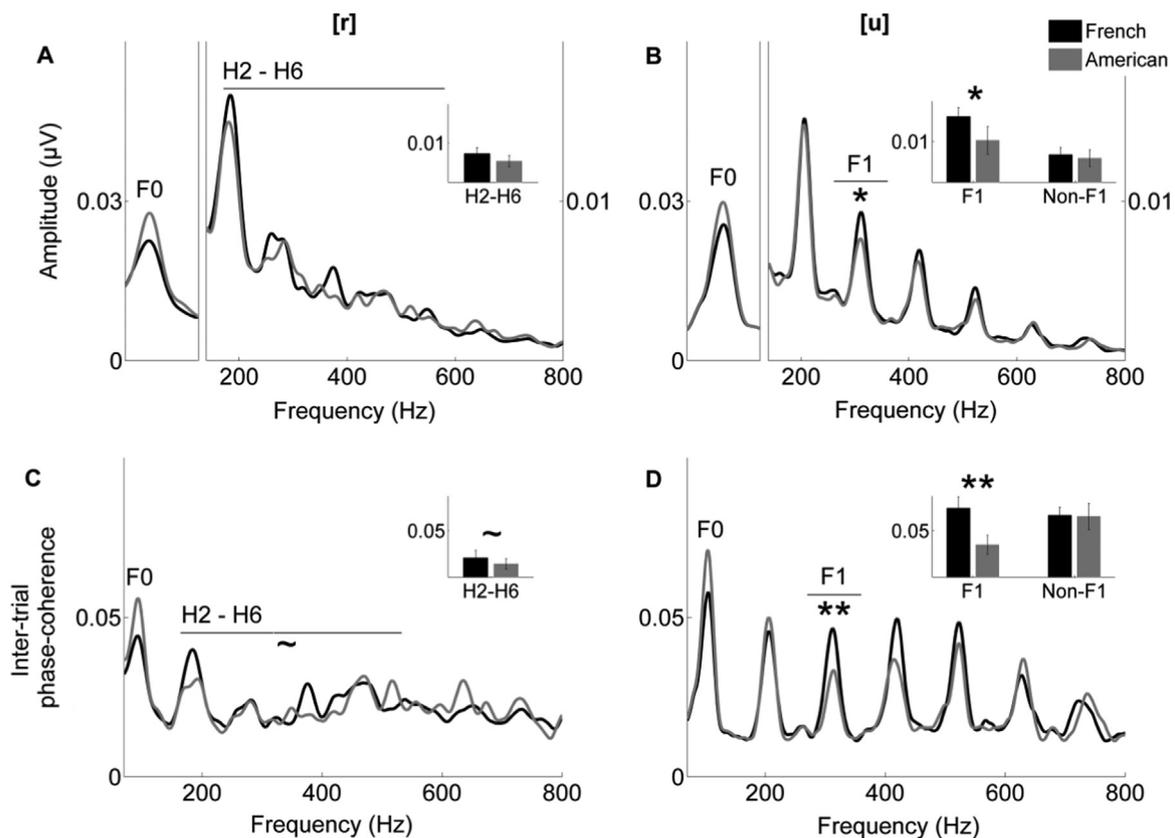
In the neural responses to the whole stimuli, American participants had larger spectral amplitudes in response to the fundamental frequency ( $F_0$ ) for both [ru] and [thae] (main effect of group:  $F(1,45)=3.580$ ,  $p=0.045$ ). This did not differ across stimuli (group  $\times$  stimulus interaction:  $F(1,45)=0.072$ ,  $p=0.789$ ).

Across both stimuli (thae/ru), the magnitude of the responses to the harmonics was equivalent for both groups ( $F(1,42)=0.057$ ,  $p=0.812$ ) suggesting that the main effect of group found for the  $F_0$  cannot be attributed to a global enhancement of stimuli representations in one group. Most importantly, stimuli neural representations of the harmonics ( $H_2$ – $H_6$ ) differed as a function of the native language (group  $\times$  stimulus interaction:  $F(1,42)=5.704$ ,  $p=0.021$ ).

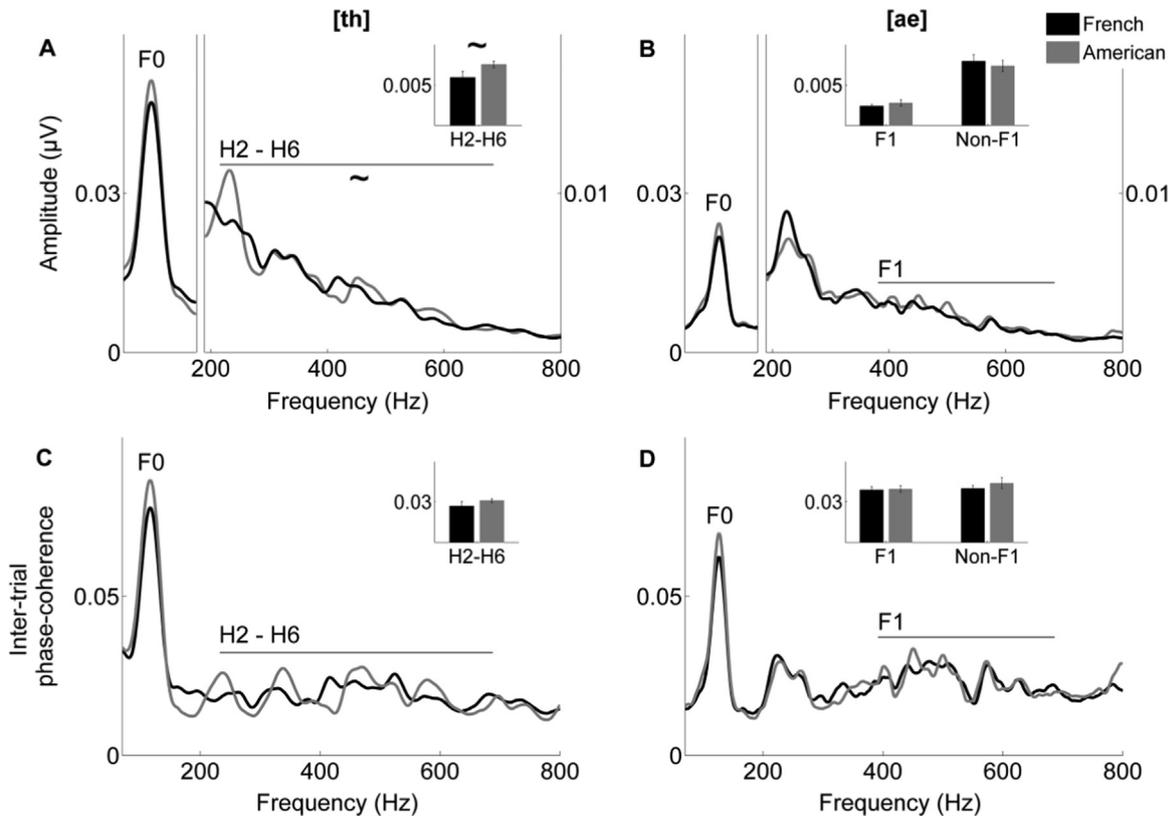
#### 3.1.2. Consonants

Spectral representation of the  $F_0$  did not show group differences for both stimuli (main effect of group:  $F(1,44)=0.262$ ,  $p=0.611$ ; group  $\times$  stimulus interaction:  $F(1,44)=0.022$ ,  $p=0.884$ ).

When turning to the harmonics ( $H_2$ – $H_6$ ), while the two groups did not differ overall (main effect of group:  $F(1,44)=0.440$ ,  $p=0.511$ ), stimuli neural representation did differ as function of the native language (group  $\times$  stimulus interaction:  $F(1,44)=7.128$ ,  $p=0.011$ ). Indeed, *post-hoc* tests revealed that American participants had a stronger spectral representation of the harmonics of the American consonant ([th]:  $p=0.056$ ), while this was not the



**Fig. 3.** Spectral representation (top) and inter-trial phase-coherence (bottom) of the French consonant [r] (left panels) and vowel [u] (right panels). A,C. The results show that for the French consonant [r], French (black) and American (gray) participants have equivalent representation of the  $F_0$  and its higher harmonics ( $H_2$ – $H_6$ ), while phase-coherence analyses show that French participants tend to have more robust representation of the higher harmonics only. B,D. For the French vowel [u], French participants exhibit an advantage over American participant with a greater and more robust representation of the harmonics at formant frequencies ( $F_1$ ). \* $p < 0.01$ , \*\* $p < 0.005$ . Error bars represent  $\pm 1$ SE.



**Fig. 4.** Spectral representation (top) and inter-trial phase-coherence (bottom) of the American consonant [th] (left panels) and vowel [ae] (right panels). A,C. The results show that, while both groups have equivalent representations of the F0, American participants (gray) have a stronger spectral representations of the harmonics (H2–H6) for the American consonant [th]. B,D. For the American vowel [ae], both groups show equivalent representations of the F0, and harmonics at formant and non-formant frequencies. Error bars represent  $\pm 1SE$ .

case for the French consonant ([r]:  $p=0.421$ , Figs. 3 and 4 A).

### 3.1.3. Vowels

Analysis of the F0 did not reveal any significant effect (main effect of group:  $F(1,44)=1.165$ ,  $p=0.286$ ; group  $\times$  stimulus interaction:  $F(1,44)=0.036$ ,  $p=0.850$ ).

Analysis of the harmonics at formant frequencies (F1) showed that vowel neural representations differed as function of the native language (group  $\times$  stimulus interaction:  $F(1,43)=4.295$ ,  $p=0.044$ ) with French showing an advantage over American participants in encoding F1-like harmonics of the French vowel [u]. *Post-hoc* tests revealed that for the French vowel [u] only, French participants had stronger spectral representations of F1 range harmonics frequencies ([u]:  $p=0.005$ , effect size = 0.002  $\mu V$ ; [ae]:  $p=0.998$ , Figs. 3 and 4 B).

Interestingly, analyses at non-formant frequencies did not reveal any significant difference (main effect of group:  $F(1,41)=0.659$ ,  $p=0.421$ ; group  $\times$  stimulus interaction:  $F(1,41)=0.008$ ,  $p=0.931$ ), showing that for the French vowel [u], French participants had a specific neural enhancement at harmonics within F1 range but not at harmonics outside F1 range.

## 3.2. Inter-trial phase-coherence

### 3.2.1. Whole stimuli

Analyses of inter-trial phase-coherence at the fundamental frequency (F0) did not reveal any significant effect (main effect of group:  $F(1,43)=0.550$ ,  $p=0.462$ ; group  $\times$  stimulus interaction:  $F(1,43)=0.187$ ,  $p=0.667$ ).

Robustness of neural responses at harmonics (H2–H6) was equivalent between group (main effect of group:  $F(1,40)=0.098$ ,  $p=0.755$ ). However, robustness of stimuli neural representations

of the harmonics (H2–H6) differed as a function of the native language (group  $\times$  stimulus interaction:  $F(1,40)=4.414$ ,  $p=0.042$ ).

### 3.2.2. Consonants

Analyses of inter-trial phase-coherence at the F0 did not reveal any significant effect (main effect of group:  $F(1,41)=0.021$ ,  $p=0.885$ ; group  $\times$  stimulus interaction,  $F(1,41)=0.002$ ,  $p=0.961$ ).

Inter-trial phase-coherence at harmonics (H2  $\times$  H6) was equivalent between group (main effect of group:  $F(1,43)=0.256$ ,  $p=0.615$ ). However, robustness of stimuli neural representations differed as function of the native language (group  $\times$  stimulus interaction:  $F(1,43)=5.179$ ,  $p=0.030$ ). This was mostly due to French participants having more robust encoding of the French consonant's harmonics compared to American participants, although this difference did not reach significant in *post-hoc* tests ([r]:  $p=0.102$ ; [th]:  $p=0.426$ , Figs. 3 and 4 C).

### 3.2.3. Vowels

Analyses of inter-trial phase-coherence at the F0 did not reveal any significant effect (main effect of group:  $F(1,42)=0.587$ ,  $p=0.448$ ; group  $\times$  stimulus interaction,  $F(1,42)=0.388$ ,  $p=0.537$ ).

Although inter-trial phase-coherence of the harmonics at formant frequencies (F1) was equivalent between groups (main effect of group:  $F(1,42)=2.763$ ,  $p=0.104$ ), the robustness of stimuli neural representations at F1 frequencies strongly differed as function of the native language ( $F(1,42)=10.567$ ,  $p=0.002$ ). *Post-hoc* tests revealed that for the French vowel, French participants had more robust neural representations of the harmonics at F1 frequencies while for the American vowel there was no significant difference ([u]:  $p=0.002$ , effect size = 0.016  $\mu V$ ; [ae]:  $p=0.592$ , Figs. 3 and 4 D).

Consistent with the results found for spectral representation,

analyses of inter-trial phase-coherence at non-formant frequencies did not reveal any significant effect (main effect of group:  $F(1,41)=0.614$ ,  $p=0.438$ ; group  $\times$  stimulus interaction:  $F(1,41)=1.924$ ,  $p=0.173$ ), showing that the advantage of French participants in encoding the French vowel [u] is specific to harmonics within  $F1$  range.

#### 4. Discussion

The goal of this study was to test the hypothesis that language experience shapes neural representations of speech sounds. More precisely, fast subcortical processing 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 ( $F0$ ) compared to French participants. Secondly, participants exhibited more robust encoding and more precise spectral representations when listening to phonemes of their native language as compared to non-native language, especially at frequencies that are linguistically-relevant (formants).

##### 4.1. Effect of language expertise on $F0$ representation

The subcortical representation of the fundamental frequency ( $F0$ ) 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  $F0$  (Bidelman et al., 2011, 2009; Wong et al., 2007), we carefully verified that American and French participants were matched in term of musicianship (see Section 2). Then, the global enhancement in subcortical encoding of the  $F0$  in the American group may be related to differences in language experience, particularly the role that  $F0$  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 ( $F0$ ) 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. In this experiment, 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  $F0$  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  $F0$  in French participants. Nonetheless, these group differences on the  $F0$  have to be interpreted with care because they were only visible on the spectral analyses of the whole stimuli and were absent in the phase coherence analyses of the whole stimuli and, most importantly, in the analyses separating consonants and vowels (the latter being the typical stress carrier in a word).

##### 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 on the responses to the whole stimuli when considering harmonics higher than the fundamental frequency ( $H2-H6$ ). Analyses on the consonants and the vowels separately revealed two main findings. First, in both analyses of spectral density (the amplitude of the signal in a spectral representation) and inter-trial phase-coherence (the stability of phase over trials), spectral representations of the consonants were sensitive to the native language of the listeners, but only at harmonics higher than the fundamental frequency ( $H2-H6$ ). While differences are not striking when considering only one stimulus at a time, the analyses of the interaction between groups and stimuli is well designed to reveal subtle differences of language expertise on subcortical representations, independently of any possible effect driven by differences in the acoustic content of the stimuli. The results demonstrate a fine-grained subcortical representation of the consonant that belong to the phonemic inventory of the listener. Second, this sensitivity to the native language of the listeners was also present in the analyses of the vowels, but only when considering the harmonics at  $F1$  frequency range. The fact that the interaction was visible at  $F1$  frequencies, rather than on the fundamental frequency or other harmonics, is consistent with our prediction that language expertise shapes linguistically relevant neural responses.

Although the effect of the native language on stimulus processing was present for both stimuli, it was more robust with the stimulus [ru], with French participants showing a clear advantage processing the French vowel, while group differences for the stimulus [thae] were marginal. 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.

Another explanation for the lack of difference between the French and American participants for the stimulus [thae] could be related to the fact this speech sound did not evoke distinct responses to the harmonics, thus potentially blurring the difference between groups. Working with natural tokens is in our opinion the best way to study language dependant effects on subcortical representations, but it comes with the price of having more complex and time-varying harmonic structure as compared to synthetic sounds. In the present experiment, the [thae] stimulus has more varying harmonics over time as compared to the [ru] stimulus, which results in more spread out peaks on the Fourier transform. Because of the phase-locking property of subcortical structures, representations of these frequency components are therefore not stable over time resulting in a lower and more spread out peaks on the response FFT. The difference in pitch between the two stimuli could also explain a poorer harmonics representation for the [thae] stimulus. Compared to the [ru], the [thae] syllable has a higher fundamental frequency and harmonics (see Table 2). Given that neural phase-locking of subcortical structures becomes weaker with increasing frequency (Greenberg, 1980; Greenberg et al., 1987), higher and more varying harmonics for the syllable [thae] may induce less robust neural representations.

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, 2015). 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., (2009, 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). While in tonal languages, representation of pitch variations (i.e.  $F_0$ ) carrying lexical information is enhanced in native speakers, our results show that in non-tonal languages, spectral representations of formant frequencies ( $F_1$ ) that are relevant to the phonetic system are specifically strengthened in native speakers whereas other non-relevant spectral features are not affected by language experience.

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. These effects are likely to stem from a cascade of top-down processes that enhance linguistically-relevant, and prune non-relevant, information (Ahissar and Hochstein, 2004; Gilbert and Sigman, 2007; Kraus and White-Schwoch, 2015; Krishnan and Gandour, 2014; Tzounopoulos and Kraus, 2009). This means that the relevant auditory stimulus

features (e.g. first and second formants) may be 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 more robust neural representations (Fritz et al., 2010). Interestingly, the impact of attention on neural activity is evident throughout the auditory system (Perrot et al., 2006). Our results are in line with the recent views of the auditory system, describing auditory perception as an active process that involves interaction between cognitive and sensory levels, where top-down processes take place (Kraus and White-Schwoch, 2015).

Overall, our results can be explained both in terms of bottom-up and top-down processes. On the one hand, 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 hand, 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 account for more accurate subcortical representations of native language stimulus features.

Future studies are needed to explore language-dependent effects on subcortical processing by comparing possibly more distant languages in terms of phonetic repertoire. This may maximize the effects that we showed here for two Indo-European languages. It will also be important to recruit participants that have a minimal or absolutely no knowledge of the other language, which was not the case here for French participants. Finally, while natural speech sounds possibly maximize the language-dependent effects, it is not possible to carefully control them in terms of their acoustic features. Thus, a possible development of our work could also be the use of resynthesized natural tokens, carefully controlling the spectro-temporal variability.

## Acknowledgements

This work was supported by the Brain and Language Research Institute (BLRI, ANR-11-LABX-0036).

## References

- Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464. <http://dx.doi.org/10.1016/j.tics.2004.08.011>.
- Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47. <http://dx.doi.org/10.1016/j.heares.2008.08.004>.
- Berens, P., 2009. CircStat: a MATLAB toolbox for circular statistics. *J. Stat. Softw.* 31, 1–21.
- Bidelman, G.M., Gandour, J.T., Krishnan, A., 2009. Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J. Cogn. Neurosci.* 23, 425–434. <http://dx.doi.org/10.1162/jocn.2009.21362>.
- 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. <http://dx.doi.org/10.1016/j.bandc.2011.07.006>.
- Boersma, P., Weenink, D., 2001. Praat, A System for Doing Phonetics by Computer.
- 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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/10.1162/0898929000562390>.

- 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. <http://dx.doi.org/10.1016/j.cognition.2007.04.001>.
- Fourtassi and Dupoux, 2014. A rudimentary lexicon and semantics help bootstrap phoneme acquisition. In: *Proceedings of the Eighteenth Conference of Computational Language Learning*. pp. 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. <http://dx.doi.org/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. <http://dx.doi.org/10.1073/pnas.97.14.8081>.
- Gilbert, C.D., Sigman, M., 2007. Brain states: top-down influences in sensory processing. *Neuron* 54, 677–696. <http://dx.doi.org/10.1016/j.neuron.2007.05.019>.
- Greenberg, S., Marsh, J.T., Brown, W.S., Smith, J.C., 1987. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear. Res.* 25, 91–114. [http://dx.doi.org/10.1016/0378-5955\(87\)90083-9](http://dx.doi.org/10.1016/0378-5955(87)90083-9).
- Greenberg, S., 1980. WPP, No. 52: Temporal Neural Coding of Pitch and Vowel Quality. Dep. Linguist. UCLA.
- Gutschalk, A., Micheyl, C., Oxenham, A.J., 2008. Neural correlates of auditory perceptual awareness under informational masking. *PLoS Biol.* 6, e138. <http://dx.doi.org/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. <http://dx.doi.org/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. [http://dx.doi.org/10.1016/S0010-0277\(02\)00198-1](http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/10.1038/nrn2882>.
- Kraus, N., White-Schwoch, T., 2015. Unraveling the biology of auditory learning: a cognitive-sensorimotor-reward framework. *Trends Cogn. Sci.* 19, 642–654. <http://dx.doi.org/10.1016/j.tics.2015.08.017>.
- Krishnan, A., Gandour, J.T., 2014. Language experience shapes processing of pitch relevant information in the human brainstem and auditory cortex: Electrophysiological evidence. *Aust. Acoust. Soc.* 42, 166–178.
- 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. <http://dx.doi.org/10.1016/j.cogbrainres.2005.05.004>.
- 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. <http://dx.doi.org/10.1097/WNR.0b013e3283263000>.
- Kuhl, P.K., 2004. Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. <http://dx.doi.org/10.1038/nrn1533>.
- 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.
- 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. <http://dx.doi.org/10.1111/j.1467-7687.2006.00468.x>.
- Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective subcortical enhancement of musical intervals in musicians. *J. Neurosci.* 29, 5832–5840. <http://dx.doi.org/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. <http://dx.doi.org/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. [http://dx.doi.org/10.1016/S0010-0277\(01\)00157-3](http://dx.doi.org/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. [http://dx.doi.org/10.1016/0959-4388\(94\)90068-X](http://dx.doi.org/10.1016/0959-4388(94)90068-X).
- Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalogr. Clin. Neurophysiol.* 35, 665–667. [http://dx.doi.org/10.1016/0013-4694\(73\)90223-X](http://dx.doi.org/10.1016/0013-4694(73)90223-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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/10.3389/fnagi.2012.00030>.
- Peelle, J.E., 2010. Hierarchical processing for speech in human auditory cortex and beyond. *Front. Hum. Neurosci.* . <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/10.1097/AUD.0b013e3181c8b272>.
- 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. <http://dx.doi.org/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. <http://dx.doi.org/10.1111/j.1460-9568.2009.06617.x>.
- Suga, N., 2008. Role of corticofugal feedback in hearing. *J. Comp. Physiol. A* 194, 169–183. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/10.1016/j.neuron.2009.05.002>.
- Vaissière, J., 1991. Rhythm, accentuation and final lengthening in French. pp. 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.* 25, 121–133. [http://dx.doi.org/10.1016/S0163-6383\(02\)00093-0](http://dx.doi.org/10.1016/S0163-6383(02)00093-0) (25th Anniversary Special Issue ).
- Wong, P.C.M., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. <http://dx.doi.org/10.1038/nn1872>.
- 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. <http://dx.doi.org/10.1016/j.neuroimage.2005.02.040>.