Neuropsychologia 89 (2016) 57–65

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

Native language shapes automatic neural processing of speech

Bastien Intartaglia[a,b,*], Travis White-Schwoch[c], Christine Meunier[b,d], Stéphane Roman[e], Nina Kraus[c,f,g], Daniele Schön[a,b]

[a] Aix-Marseille Université, INSERM, Institut de Neurosciences des Systèmes, Marseille, France
[b] Brain and Language Research Institute, Labex BLRI, France
[c] Auditory Neuroscience Laboratory and Department of Communication Sciences, Northwestern University, Evanston, IL, United States of America
[d] Aix-Marseille Université, CNRS, LPL UMR 7309, 13100 Aix-en-Provence, France
[e] La Timone Children's Hospital, ENT Unit, Marseille, France
[f] Department of Neurobiology & Physiology, Northwestern University, Evanston, IL, United States of America
[g] 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.

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 (B. Intartaglia).

http://dx.doi.org/10.1016/j.neuropsychologia.2016.05.033
0028-3932/© 2016 Elsevier Ltd. All rights reserved.

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