Brain and Language 194 (2019) 77–83

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Short communication

Effects of formant proximity and stimulus prototypicality on the neural discrimination of vowels: Evidence from the auditory frequency-following response

T. Christina Zhao[a,*], Matthew Masapollo[b], Linda Polka[c], Lucie Ménard[d], Patricia K. Kuhl[a]

[a] Institute for Learning and Brain Sciences, Portage Bay Building, Box 357988, University of Washington, Seattle, WA 98195-7988, United States
[b] Department of Speech, Language and Hearing Sciences, Boston University, 635 Commonwealth Avenue, Boston, MA 02215, United States
[c] School of Communication Sciences and Disorders, McGill University, 2001 McGill College, 8th Floor, Montreal, Quebec H3A 1G1, Canada
[d] Department of Linguistics, University of Quebec at Montreal, 320 Sainte-Catherine East, Montreal, QC H2X 1L7, Canada

ARTICLE INFO

Keywords:
Speech perception
Focal vowels
Natural Referent Vowel framework
Native Language Magnet model
Auditory frequency-following response

ABSTRACT

Cross-language speech perception experiments indicate that for many vowel contrasts, discrimination is easier when the same pair of vowels is presented in one direction compared to the reverse direction. According to one account, these directional asymmetries reflect a *universal* bias favoring "focal" vowels (i.e., vowels with prominent spectral peaks formed by the convergence of adjacent formants). An alternative account is that such effects reflect an *experience-dependent* bias favoring prototypical exemplars of native-language vowel categories. Here, we tested the predictions of these accounts by recording the auditory frequency-following response in English-speaking listeners to two synthetic variants of the vowel /u/ that differed in the proximity of their first and second formants and prototypicality, with stimuli arranged in oddball and reversed-oddball blocks. Participants showed evidence of neural discrimination when the more-focal/less-prototypic /u/ served as the deviant stimulus, but not when the less-focal/more-prototypic /u/ served as the deviant, consistent with the focalization account.

1. Introduction

The fundamental goal of research in the field of speech perception is to explicate the mechanisms and processes by which listeners map the input acoustic signal onto phonological units, such as features, phonemes, syllables, and words. Much of the research in this field has been concerned with characterizing the mapping from the acoustic signal to phonemes; that is, the consonants and vowels that combine to form the words of language (see Holt & Lotto, 2010, for a review). Within this overarching agenda, researchers have often focused on addressing how and when the discrimination and categorization of consonants and vowels change with specific linguistic experience (Cutler, 2012; Kuhl et al., 2008). This emphasis on exploring what is *language-specific* as opposed to what is *language-universal* in the perception of phonetic information derived in large part from research with human infants, human adults, and non-human primates revealing that early linguistic experience profoundly alters perception by decreasing discrimination sensitivity near native phonetic category prototypes and increasing sensitivity near boundaries between categories (Feldman, Griffiths, & Morgan, 2009; Guenther, Husain, Cohen, & Shinn-Cunningham, 1999; Guenther, Nieto-Castanon, Ghosh, & Tourville, 2004; Kuhl, 1991; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

More recent efforts, however, have shown that there are also universal perceptual biases in place early in development that guide and constrain how listeners from diverse linguistic backgrounds decode the acoustic signal (Nam & Polka, 2016; Polka & Bohn, 2003, 2011). In the domain of vowel perception, it has become increasingly clear that listeners (both adult and infant) are universally biased toward vowels that fall closer to the periphery of acoustic vowel space (as defined by the first [F_1] and second formants [F_2]). This universal vowel bias, which is the focus of the current research, is often demonstrated in phonetic discrimination tasks as a directional asymmetry: significantly better discrimination performance is observed when changes occur from a relatively-less to a relatively-more peripheral vowel than changes in the reverse direction. This perceptual pattern is summarized in Fig. 1A, which shows many of the vowel contrasts that have been documented

* Corresponding author at: Institute for Learning and Brain Sciences, University of Washington, Portage Bay Building, Box 357988, Seattle, WA 98195-7988, United States.

E-mail addresses: zhaotc@uw.edu (T.C. Zhao), mmasapol@bu.edu (M. Masapollo), linda.polka@mcgill.ca (L. Polka), menard.lucie@uqam.ca (L. Ménard), pkkuhl@uw.edu (P.K. Kuhl).

https://doi.org/10.1016/j.bandl.2019.05.002
Received 26 January 2019; Received in revised form 14 May 2019; Accepted 16 May 2019

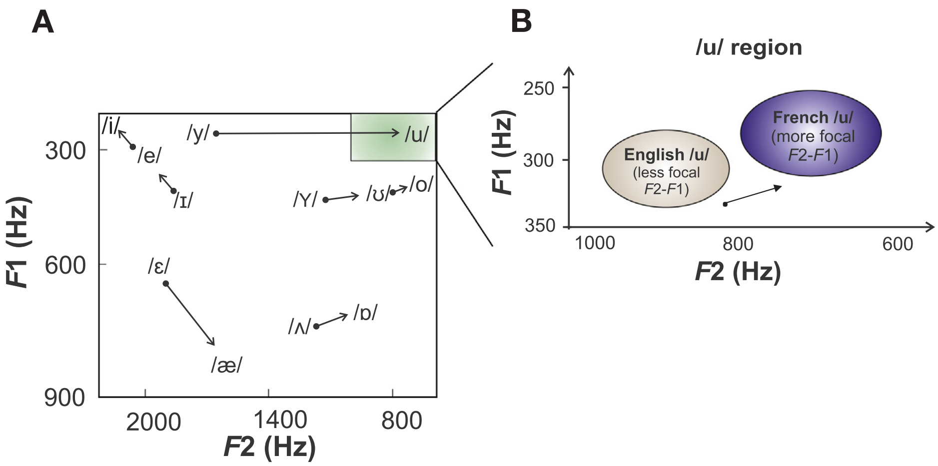

Fig. 1. (A) Schematic illustration of acoustic vowel space (defined by the first two formant frequencies [F_1 and F_2]; adapted from Polka and Bohn (2011)). Vowel contrasts reported to show directional asymmetries in studies of infant vowel perception are plotted (see Polka and Bohn (2003, Table 1, p. 225), for a list of studies these results are based on). Arrows indicate the direction of vowel change that is easier to discriminate. The green rectangle delimits the portion of acoustic space that corresponds to the acoustic realization of the vowel /u/ ("oo") across human languages. (B) Magnified view of the /u/ portion of acoustic space. The precise location in the acoustic space of the /u/ category in English and French is shown; the beige ellipse delimits the region corresponding to prototypic English /u/, and the blue ellipse delimits the region corresponding to prototypic French /u/. As the plot shows, French /u/ is more acoustically peripheral and more focal (between F_1 and F_2) than English /u/. The arrow points in the direction that has been found to be easier to discriminate by both English- and French-speaking adults (see text for explanation). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in infant vowel discrimination studies with arrows indicating the direction of change that was reported to be easier to discriminate (see Polka & Bohn, 2003, 2011, for the list of studies these results are based on). These findings have been reviewed and discussed extensively by Polka and Bohn (2003, 2011), and recently compiled in several meta-analyses (Polka, Ruan, & Masapollo, 2019; Tsuji & Cristia, 2017).

Current models and theories of speech perception provide insight into the potential mechanisms and processes underlying these directional asymmetries (Kuhl et al., 2008; Lahiri & Reetz, 2002; Polka & Bohn, 2011). The Natural Referent Vowel (NRV) framework, which is a model of early phonetic development (Polka & Bohn, 2011), has been used to guide a number of recent studies on vowel perception asymmetries (e.g., Masapollo, Polka, Molnar, & Ménard, 2017; Masapollo, Polka, & Ménard, 2017; Masapollo, Zhao, Franklin, & Morgan, 2019). In this model, directional asymmetries are argued to reflect a universal sensitivity to prominent spectral peaks formed by the convergence of adjacent formants. The basic idea is as follows: During vowel production, movements of the articulators, particularly those of the tongue, change the overall shape and configuration of the vocal tract. This change in shape is acoustically manifested in the speech signal as changes in formant values (see Stevens, 1989, 1998, for thorough reviews). Vowels that fall close to the periphery of acoustic vowel space are executed when the tongue body is in its most extreme posture and displacement (either front or back, high, or low) from a "neutral" (schwa-like) vocal tract configuration. In addition, some peripheral vowels (e.g., /u/, /y/) are implemented with a greater degree of lip compression and/or protrusion. These extreme vocalic articulations lead to acoustic signals in which formants merge close together in frequency (i.e., "focal" vowels). For example, F_2, F_3, and F_4 converge during the production of /i/ (that is the highest front vowel), and F_1 and F_2 converge during the production of /a/ (that is the lowest back vowel) as well as /u/ (that is the highest back vowel). When two neighboring formants merge close together in frequency there is a mutual reinforcement of their acoustic energy, such that the amplitude of one or both formants is enhanced. As a result, acoustic energy becomes concentrated into a narrow spectral region (see Kent & Read, 2002; Stevens, 1989, 1998, for discussion).[1] The NRV model proposes that this concentration of spectral energy gives rise to vowel sounds with well-defined spectral prominences that are easier for listeners to detect, encode, and retain in memory, which in turn, biases perception and leads to directional asymmetries during discrimination tasks (for discussion, see Schwartz, Abry, Boë, Ménard, & Vallée, 2005; Masapollo, Polka, Molnar, et al., 2017).

An alternative, but not mutually exclusive account, of asymmetries derives from the Native Language Magnet (NLM) model (Kuhl & Iverson, 1995; Kuhl et al., 2008), which is another prominent model of early phonetic development. This model argues instead that directional asymmetries may be induced by perceptual learning, using the statistical properties of the input speech, which biases listeners toward native language phonetic category prototypes (i.e., adult-defined "best" instances of a phonetic category; cf. Feldman et al., 2009). The best exemplars of a native phonetic category are said to "pull" similar auditory representations toward itself much as a magnet attracts iron (i.e., the "*perceptual magnet effect*"; Kuhl, 1991; see also, Miller & Eimas, 1996). This model thus predicts that listeners should display heightened sensitivity when discriminating a change from a less-prototypic to a more-prototypic vowel compared to the reverse (see Masapollo, Polka, Molnar, et al., 2017, for further discussion). Critically, however, these predictions have typically only been tested using *behavioral* methods (e.g., Schwartz & Escudier, 1989; Kuhl, 1991; Miller & Eimas, 1996; Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017). Relatively little *neural* data is available to evaluate these claims. We briefly review some recent findings on asymmetries from our lab group and others before presenting the present neuro-physiological study.

Recent cross-language studies with adults provide critical data supporting the predictions of the NRV model (Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017). Using the Variable Linear Articulatory Model (Ménard, Schwartz, & Böe, 2004), Masapollo and his colleagues generated a broad array of vowel stimuli that varied in their first and second formant frequencies, to create a two-dimensional stimulus grid with the frequencies equally spaced on a psychophysical scale. These stimuli were then presented to native, monolingual English and French listeners for phonetic identification and goodness ratings. The results showed that although all of the members of the stimulus grid were consistently identified as intelligible instances of the vowel /u/ by listeners in both language groups, the best French /u/ exemplars had a higher degree of formant convergence (between F_1 and F_2) than did the best English /u/ exemplars. Thus, the best French /u/ was a more focal vowel than the best English /u/. In subsequent AX discrimination tests, subjects from both language groups performed better at discriminating changes from instances of the less-

[1] The peripheral vowels (/i/, /a/, and /u/) have been referred to as the "focal" vowels in the speech literature because they exhibit maximal degrees of formant convergence (Schwartz & Escudier, 2005). However, focalization is not all-or-nothing. Rather, it is a graded property that gives rise to salience differences across vowel space.

focal/English-prototypic /u/ to instances of the more-focal/French-prototypic /u/, compared with the reverse direction (shown in Fig. 1B). Moreover, the magnitude of the asymmetry did not interact with native language, demonstrating a universal bias favoring vowel sounds with a greater degree of formant convergence that operates independently of experience-dependent biases related to language-specific prototype categorization (contra Kuhl, 1991; Miller & Eimas, 1996).

The effects of formant proximity on vowel discrimination documented at the behavioral level by Masapollo *et al.* are presumed to be due to enhanced cognitive and neural encoding of vowels with more well-defined spectral peaks due to formant convergence. Several neurophysiological studies employing electroencephalography (EEG) methods provide corroborating data showing asymmetrical discriminative neural responses in adults during vowel processing (Dufour, Brunellière, & Nguyen, 2013; Molnar, Polka, Baum, & Steinhauer, 2014). Using an oddball paradigm, Dufour et al. (2013) examined the mismatch negativity (MMN), a cortical auditory-evoked response thought to index neural discrimination (Naatanen et al., 1997), in Southern French-speaking adults while listening to the non-native /o/-/u/ contrast. The results revealed asymmetric MMN responses that pattern as predicted by the NRV model; namely, discriminatory responses were heightened when /u/ served as the infrequent deviant stimulus, compared to when /o/ served as the deviant (i.e., F_1 and F_2 merge closer in frequency for /u/ than /o/). The first goal of the present study, therefore, was to expand on our previous behavioral studies and examine whether we can also observe neural directional asymmetries, consistent with the NRV framework, using /u/ vowel stimuli adapted from Masapollo, Polka, Molnar, et al. (2017).

Our second goal was to go beyond the examination of discriminatory MMN responses and investigate whether listeners show *asymmetrical neural encoding of frequencies corresponding to formants* when presented with more- versus less-focal variants of the vowel /u/. Toward this end, we examined the auditory frequency-following response (FFR) to a less-focal/English prototypic /u/ and a more-focal/French prototypic /u/. The FFR is a time-locked neural response to periodic sounds recorded from the scalp using EEG electrodes, and it is currently thought to arise from the summation of signals from both cortical and subcortical structures along the ascending auditory pathway (see Skoe, Krizman, Anderson, & Kraus, 2013; Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016; Tichko & Skoe, 2017; Bidelman, 2018). Typically, it is elicited during passive listening tasks in which participants do not attend to the stimuli or produce an overt judgment or other behavioral response. In the context of experimental speech perception studies, the FFR reflects pre-attentive neural tracking of sustained periodic information (i.e., fundamental frequency and higher harmonics in vowels). Critically, unlike cortical-evoked auditory responses, *the FFR can partially reflect the physical properties of the evoking stimulus (up to around 1000 Hz), and therefore can be used to assess the integrity with which formants are encoded in the brain.* Spectral analyses of the FFR to steady-state vowel stimuli show distinct peaks at the harmonics adjacent to the fundamental frequency and first formant (see, e.g., Krishnan, 2002; Bidelman, Moreno, & Alain, 2013). Frequencies above the first formant are not reliably reflected in the FFR and therefore are not commonly examined.

In the current research, EEG signals were recorded when English monolingual adults listened to a less-focal/English prototypic /u/ and a more-focal/French prototypic /u/, using an oddball/reversed-oddball paradigm. In the oddball condition, the standard stimulus was the less-focal/English prototypic /u/ and the deviant stimulus was the more-focal/French prototypic /u/. In the reversed oddball condition, the roles of the standard and deviant stimuli were switched. We hypothesized that if the neural encoding of vowels is sensitive to formant proximity (à la the NRV model), then there should be greater spectral energy in the *F1* frequency regions in the neural response for the more-focal/French prototypic /u/ compared to the less-focal/English prototypic /u/. Furthermore, this enhanced spectral encoding of the first formant for the more-focal variant may be stronger or occur only when it serves as the deviant stimulus, which would be consistent with the behavioral data reported in the literature (see Fig. 1B).

Although stimulus prototypicality did not appear to influence the directional asymmetry reported in Masapollo, Polka, Molnar, et al. (2017) behavioral data, it is still possible that long-term linguistic experience might influence the FFR (à la the NLM model). Indeed, recent FFR data show that the neural encoding of spectral and temporal properties of speech sounds is sensitive to language experience (Intartaglia et al., 2016; Zhao & Kuhl, 2018). With regard to spectral processing, Intartaglia et al. (2016) reported more robust F_1 encoding of naturally-spoken native vowels compared to non-native vowels. In view of these findings, it is possible that, for English listeners, the spectral representation of F_1 might instead be greater for the English prototypic /u/ compared to the French prototypic /u/, regardless of variation in formant proximity (between F_1 and F_2) and whether it serves as the standard or deviant stimulus.

To address this theoretical tension, we focused our analysis on the power values of the FFR in the F_1 region. If formant convergence is at play, we expect English adults to display an interaction between vowel type and condition due to enhanced power in the F_1 region for the more-focal/French /u/ when it serves as the deviant compared to the standard. In contrast, if stimulus prototypicality is at play, we expect English adults to display, a main effect of vowel type due to enhanced power in the F_1 region for the more prototypic (but less focal) English /u/ irrespective of stimulus condition (deviant vs. standard).

2. Results

A 2 × 2 repeated measures analysis of variance (Vowel Type [less/focal/English /u/ vs. more focal/French /u/] × Condition [standard vs. deviant]) was performed on the power (mV2) values in the frequency region corresponding to F_1.[2] The results (shown in Fig. 3) revealed a significant main effect of Condition [$F(1,18) = 14.866$, $p = 0.001$, $\eta_p^2 = 0.452$], such that there were greater power values observed for the deviants (mean = 744.85; 95% CI [555.03 934.67]) compared to the standards (mean = 409.64; 95% CI [331.43 487.86]). There was also a significant interaction [$F(1,18) = 17.712$, $p = 0.001$, $\eta_p^2 = 0.496$]. The main effect of Vowel Type did not reach statistical significance [$\text{mean}_{\text{Less-focal/English prototypic /u/}} = 518.43$, $\text{mean}_{\text{More-focal/French prototypic /u/}} = 636.06$; $F(1,18) = 3.143$, $p = 0.093$, $\eta_p^2 = 0.143$].

We then conducted two simple effects tests (within the condition effect) to tease apart the Vowel Type × Condition interaction. A Bonferroni-adjusted alpha rate of 0.025 was used. These post-hoc analyses indicated that for the less-focal/English prototypic /u/, there was no difference in power values in the F_1 region when it served as standard [mean = 498.41, 95% CI [396.96 599.86]) compared to when it served as deviant [mean = 538.45, 95% CI [356.13 720.78]; $t(18) = -0.518$, $p = 0.611$). In contrast, for the more-focal French prototype /u/, power values in the F_1 region were lower when it served as standard [mean = 320.88, 95% CI [207.37 434.39]) than when it served as deviant [mean = 951.25, 95% CI [627.03 1275.46]; $t(18) = -4.576$, $p < 0.001$).

3. Discussion

In the current research, we investigated whether we can observe directional asymmetries in the neurophysiological correlates of vowel processing, and, if so, whether such directional effects are attributable to differences in formant proximity, as predicted by the NRV framework (Polka & Bohn, 2011), or long-term linguistic experience, as predicted by the NLM Model (Kuhl & Iverson, 1995; Kuhl et al., 2008; Kuhl,

[2] See *Supplementary Materials* for further details regarding analyses of frequency region corresponding to F_0.

1991). Specifically, we examined the auditory FFR in response to a less-focal/English prototypic /u/ and a more-focal/French prototypic /u/ in English-speaking adults, arranged in oddball and reversed-oddball blocks. Recent research by Masapollo and his colleagues (Masapollo, Polka, Molnar, et al., 2017; Masapollo, Polka, & Ménard, 2017; Masapollo et al., 2019) has shown evidence, at the behavioral level, that directional asymmetries are driven by a universal sensitivity to formant proximity that operates independently of language-specific prototype categorization (contra Kuhl, 1991). The present study extends this work by providing neurophysiological evidence that formant convergence influences the neural discrimination of vowels.

The present research focused on the FFR, which synchronizes with and reflects the acoustical information in the F_0 and lower harmonics of vowel stimuli (e.g., Krishnan, 2002; Skoe & Kraus, 2010; Bidelman et al., 2013; Bidelman, 2018). This allowed us to assess whether there is more robust neural encoding in the frequency range of the F_1 region for more focal versus more prototypical variants of the native vowel category /u/. Consistent with NRV, we found that English listeners exhibited *enhanced* power at the frequencies corresponding to F_1 when listening to the more-focal/French prototypic /u/, but only when it served as the deviant stimulus (Figs. 2 and 3). In contrast, the neural encoding of F_1 was similar in response to the less-focal/English prototypic /u/ regardless of whether it served as the standard or deviant stimulus. This finding is in agreement with the directional asymmetries reported in the behavioral data and provides corroborating evidence that listeners are biased toward vowel sounds with a high degree of formant convergence (Masapollo, Polka, Molnar, et al., 2017; Polka & Bohn, 2011; Schwartz et al., 2005).

Several recent studies have also shown deviant-related effects on the FFR using synthetic consonantal stimuli (e.g., Slabu, Grimm, & Escera, 2012; Skoe, Chandrasekaran, Spitzer, Wong, & Kraus, 2014; Shiga et al., 2015; reviewed in Escera, 2017). For example, Slabu et al. (2012) used a similar passive oddball/reversed oddball paradigm to investigate the neural discrimination of a synthetic /ba/-/wa/ contrast, which differed in the amplitude rise time during the initial consonantal portion of the acoustic signal. However, in contrast to the present findings and other related studies, these authors observed *attenuated* FFR deviants compared to FFR standards. It is not clear why the present findings with vowels do not line up with the deviant effects found with consonants. There may be fundamental differences in the subcortical processing of vowels and consonants given that consonants are characterized by transient acoustic cues (e.g., stop bursts), whereas vowels are characterized by steady-state spectral cues (e.g., formant trajectories).

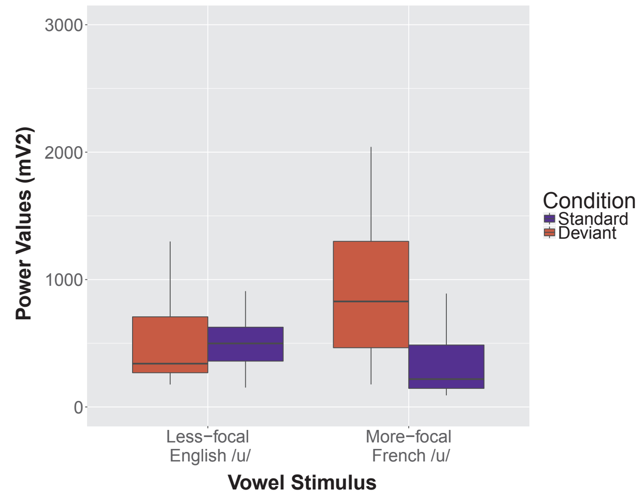

Fig. 3. Boxplots of the mean power (mV2) values at the frequency region corresponding to the first formant (F_1) for each stimulus (less-focal/English /u/ prototype vs. more-focal/French /u/ prototype) as a function of condition (standard vs. deviant).

Of particular relevance to the present study, Slabu et al. (2012) also reported directional asymmetries in the FFR discriminatory responses while participants listened to the /ba/-/wa/ contrast. Specifically, adult Catalan- and Spanish-speaking listeners showed evidence of neural discrimination of this contrast when /ba/ (which has a faster rise time) served as the deviant stimulus, but not when /wa/ served as the deviant stimulus. Although differences in formant proximity may account for

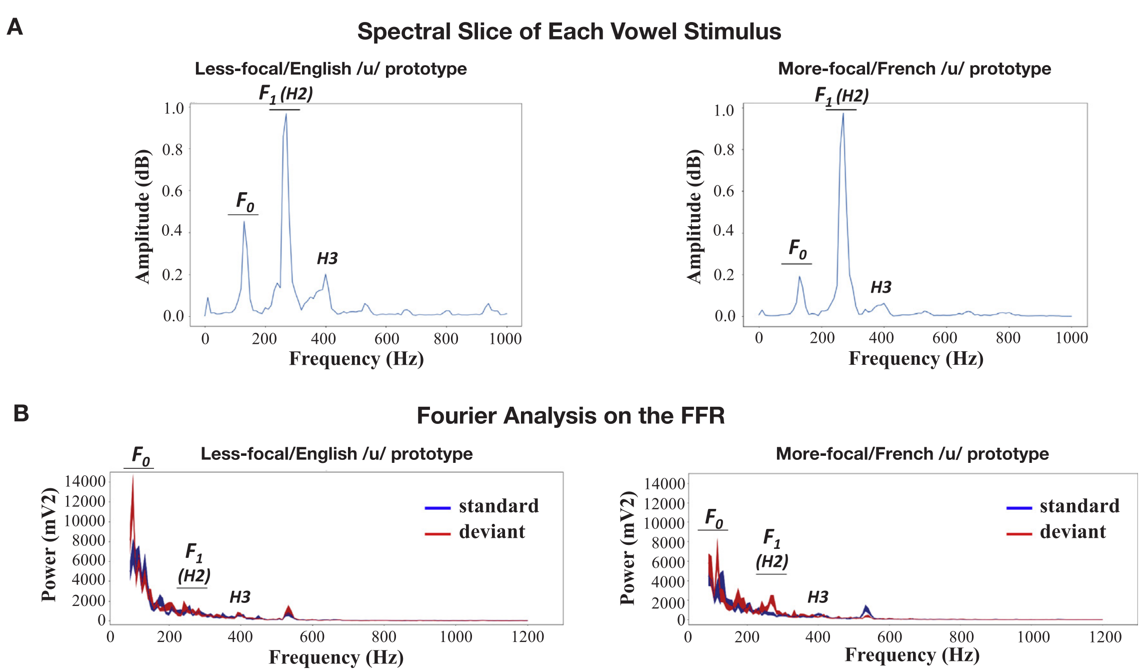

Fig. 2. (A) Fast Fourier transforms computed on each vowel stimulus. (B) Fourier analysis of the auditory FFR in response to each stimulus as a function of experimental condition (standard [blue] vs. deviant [red]). The grey lines in (A) and (B) delimit the range of frequencies corresponding to the fundamental frequency (F_0) and first formant (F_1; the second harmonic, H2) in the stimuli. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the asymmetries observed with vowels, it seems likely that other stimulus properties play a role in driving other types of asymmetries in consonant manner perception.

Slabu *et al.*'s findings can be related to those of other behavioral and neurophysiological studies. For example, Nam and Polka (2016; see also, Nam, 2014) found, using behavioral measures, that listeners (both adult and infant) performed better at discriminating a change from a fricative (e.g., /va/) to a stop (e.g., /ba/) compared to the reverse. These authors propose that this directional effect is attributable to differences in the rate of amplitude modulation (i.e., rise time) of the initial aperiodic noise in the consonantal portion of the acoustic signal. That is, listeners may be biased toward acoustic signals with more rapid rise times. This finding is compatible with the results of Slabu et al. (2012) given that the primary acoustic difference between /ba/ and /wa/ is in rise time.

Comparable findings demonstrating differential processing of rise time cues have also been reported in investigations of cortical auditory-evoked potentials in English-speaking adults. Gage, Poeppel, Roberts, and Hickok (1998) found that the M100 component occurred at a shorter latency and at a higher amplitude in response to stops (/b/, /t/) compared to fricatives and nasals (e.g., /m/, /f/). In later research, Thomson, Goswani, and Baldeweg (2009) reported comparable findings using non-speech (e.g., tonal) stimuli differing in their rate of amplitude modulation. In that case, the N1 and MMN components were affected by rise time differences.

We now turn to the theoretical implications of the present findings for the NLM model (Kuhl & Iverson, 1995; Kuhl et al., 2008; Kuhl, 1991). Although effects of language experience on phonetic perception have been clearly demonstrated in previous behavioral (e.g., Kuhl, 1991; Kuhl et al., 1992), FFR (e.g., Intartaglia et al., 2016; Zhao & Kuhl, 2018), and functional brain-imaging (e.g., Guenther et al., 2004) experiments, they were not captured in the present FFR experiment. According to one recent FFR study (Intartaglia et al., 2016), long-term linguistic experience did not lead to a global enhancement in spectral processing for native syllables compared to non-native syllables. Rather, a specific strengthening of the neural representations for linguistically relevant acoustic-phonetic features was observed (i.e., enhanced F_1 encoding, but not F_0 encoding for the vocalic portion of a syllable). On the basis of these findings, one would predict that the spectral representation of F_1 would be stronger in the FFR component while listening to the more prototypic native vowel sound (i.e., less-focal/English /u/). Thus, the present findings run contra to the NLM hypothesis (Kuhl, 1991) that directional asymmetries derive from an experience-dependent bias favoring prototypical speech sounds.

However, it is important to note that the present FFR study differed from previous behavioral studies in that discriminative responses were elicited in a *passive* listening condition while attention was directed to another task (i.e., watching a silent video). On the one hand, this suggests that formant proximity appears to influence vowel perception even under task conditions that do not require a lot of attentional resources. On the other hand, it is possible that additional attentional resources, such as those required to make overt judgments about category-goodness, may be required to elicit an effect of prototypicality on discrimination.

We also note that, the lack of an NLM effect may reflect the nature of our stimuli or it may be a weaker effect that is more challenging to measure compared to the focal vowel bias. It is still possible that a "perceptual magnet" effect might emerge during the discrimination of vowel exemplars that fall very close to the native-language prototype in psychophysical space. In fact, Kuhl (1991) findings with adults showed larger NLM effects for vowels very close to the prototype stimulus and smaller effects for vowels further from the prototype. To address this issue, we are currently testing English- and French-speaking adults on a range of /u/ stimuli carefully constructed to define equivalent and more fine-grained perceptual gradients around the English and French /u/ prototypes, using behavioral methods (Liu, Polka, Masapollo, & Ménard, in preparation).

Finally, the focalization-based perceptual bias documented here at the neural level is also compatible with other phonetic theories, such as Stevens (1989), Lindblom and Engstrand (1989) and Schwartz et al. (2005), which seek to explicate the phonetic and cognitive factors that shape and constrain vowel inventories across human languages. These models commonly propose that the focal vowels, /i/, /a/, and /u/, are nearly universal across phonological systems because they are maximally acoustically distinct from one another and intrinsically focal. The NRV framework (Polka & Bohn, 2011) further argues, on the basis of data on asymmetries in infant vowel perception, that this perceptual bias favoring the salient spectral peaks of focal vowels may guide the development of vowel perception. Specifically, it is postulated that the acoustic-phonetic salience of the focal vowels may make them easier for infants to perceive, which in turn, may perceptually enhance the contrast between relatively more- versus less-focal vowels.

In summary, the present study contributes to our theoretical understanding of how the neurophysiological correlates of vowel perception are influenced by intrinsic acoustic-phonetic properties of vowel sounds and the structure of the language-specific phoneme inventory. The results extend the existing literature and provide additional and stronger evidence to support the NRV framework (Polka & Bohn, 2011), which posits that focal vowels have a perceptual priority for listeners and that this bias operates independently of language-specific phonetic prototypic categorization processes (see also, Schwartz et al., 2005; Masapollo, Polka, Molnar, et al., 2017). Our results specifically indicate that this "focal" vowel bias reflects more robust neural encoding of harmonics higher than the fundamental frequency (i.e. in the F_1 frequency region). To further investigate how universal and language-specific aspects of phonetic perception interact, studies will examine whether there is neural enhancement of focal vowels in monolingual English- and French-learning infants early in speech development.

4. Material and methods

4.1. Participants

Twenty students (10 males, aged 18–22 years, mean age = 19.50 years [*SD* = 1.09]) at the University of Washington participated. All were native, monolingual speakers of American English. Language background and proficiency was assessed using the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007). The experiment took approximately 1.5 h, and the participants received course credit.

4.2. Stimuli

The stimuli were two computer-synthesized vowel sounds, a less-focal/English prototypic /u/ and a more-focal/French-prototypic /u/, adapted from Masapollo, Polka, Molnar, et al. (2017) previous behavioral experiments. The vowel tokens that received the highest perceptual goodness ratings by English listeners and French listeners (Masapollo et al., 2017, Experiment 1) were selected from the array and adapted for use in the present experiment. Both vowel tokens were resynthesized using the Variable Linear Articulatory Model (Ménard et al., 2004) with the same parameter values from Masapollo, Polka, Molnar, et al. (2017). In the present study, each vowel was 100-ms long with a 50-ms onset/offset ramp and had a mean F_0 (or first harmonic, H1) of 130 Hz. Acoustic and neural FFR waveforms for each vowel are shown in the Supplementary Materials (see Fig. S1). The second harmonic (H2) that falls within the F_1 region were nearly identical in energy in both sounds as shown in the power spectrum of the two stimuli in Fig. 2A. The complete details of the speech synthesis parameters can also be found in the Supplementary Materials.

4.3. Experimental design and procedure

During an experimental session, participants were consented and first completed the LEAP-Q questionnaire (Marian et al., 2007). For the auditory FFR recording portion, a standard 3-electrode set-up was used: Cz electrode on a 10–20 system, ground electrode on the forehead and reference electrode on the right ear lobe. Impedance of all electrodes were kept under 10 kΩ. Note that even though the threshold was set to be slightly higher than other studies, the recordings yielded great data quality (see accepted trial information below in Data Processing). All stimuli were delivered by Stim2 software (Version 4.0, Audio CPT), sent from a Dell Optiplex 755 computer to the Stim Audio System, and then to a monaural insert earphone to the right ear at 80 dB. Both stimulus presentation hardware and software are part of the Neuroscan system from Compumedics, Inc. (Victoria, Australia). Participants listened to the sounds passively while sitting comfortably in a reclining chair and watching a silent video of their choice in a sound-treated booth without subtitles. All experimental procedures were approved by the University of Washington's Institution Review Board (IRB).

During the electrophysiological recordings, the stimuli were arranged into oddball and reversed oddball blocks (3000 per block). In three of the blocks, the less-focal/English prototypic /u/ served as standard (80%, 2400 stimuli) while the more-focal/French prototypic /u/ served as deviant (20%, 600 stimuli). In the other three, the standard and deviant were reversed. The sequence of the blocks alternated. The inter-stimulus-intervals (offset to onset) were ~50-ms (jittered between 34 ms and 67 ms) and the stimuli were alternating in polarity (see, e.g., Aiken & Picton, 2008; Easwar et al., 2015). Total recording time was approximately 50 min. Continuous EEG was amplified using the SynAmps2™ system and recorded using the Scan (Version 4.5) software at a sampling rate of 20 kHz (Neuroscan, Compumedics).

4.4. Data processing

All data analyses were performed using EEGLAB software (Delorme & Makeig, 2004) in the MATLAB environment (MathWorks Inc. [Natick, MA]). During the initial preprocessing stage, the EEG data were first offline referenced to the reference channel. Next, the data were low-pass filtered at 2000 Hz and high-pass filtered at 80 Hz. Epochs (50-ms before stimulus onset to 150-ms after stimulus onset) were then extracted, averaged and baseline corrected (−50 ms to 0 ms served as baseline) for Standards and Deviants for each condition (less-focal/English prototypic /u/ as standard or more-focal/French prototypic /u/ as standard). Trials with voltage values exceeding ± 35 μV were rejected. The same number of standard trials were selected to match the number of deviant trials for calculation of the FFR. One subject had a significant number of trials rejected (> 20%) and their data was therefore excluded from further analysis. For all the subjects included in the analysis (N = 19), the average number of accepted trials was 1726.92 ± 80.19 (out of 1800) for each condition in each language.

FFRs were calculated for each participant as a function of Language (English /u/ vs. French /u/) and Condition (as standard vs. as deviant) (see Supplementary Materials, Fig. S1). The spectral power of the FFR was then analyzed by computing a fast Fourier transform, the results of which are shown in Fig. 2B. The power values in the frequency band (50 Hz) centered around the second harmonic (reflecting F_1) (*H2*: 270 Hz) were averaged to reflect the tracking of stimuli in those frequency ranges.

5. Statement of significance to the neurobiology of language

The current research contributes to our theoretical understanding of how the neurophysiological correlates of speech perception are influenced by intrinsic acoustic-phonetic properties of vowel sounds and the structure of the language-specific phoneme inventory. Results indicate that the neural discrimination of vowels is influenced by the proximity of adjacent formant frequencies, independent of variation in language-specific stimulus prototypicality.

Declaration of Competing Interest

None.

Acknowledgements

The research reported here was supported by the Ready Mind Project at the University of Washington's Institute for Learning and Brain Sciences. Matthew Masapollo was also supported by NIH grant R01 DC002852 (Frank H. Guenther, principal investigator) during the preparation and revision of the manuscript. We are grateful to Gloria Lam at the University of Washington for assistance with participant recruitment and data-collection. This work benefited from helpful discussions with, or comments from, Frank H. Guenther, Bharath Chandrasekaran, Pascale Tremblay, Caroline A. Niziolek, Tyler K. Perrachione, and Sung-Joo Lim.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2019.05.002.

References

Aiken, S. K., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research, 245*, 35–47.
Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage, 175*, 56–69.
Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage, 79*, 201–212.
Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications, 7*, 1–11.
Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* Cambridge, MA: MIT Press.
Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21.
Dufour, S., Brunellière, A., & Nguyen, N. (2013). To what extent do we hear phonemic contrasts in a non-native regional variety? Tracking the dynamics of perceptual processing with EEG. *Journal of Psycholinguistic Research, 42*, 161–173.
Easwar, V., Beamish, L., Aiken, S., Choi, J. M., Scollie, S., & Purcell, D. (2015). Sensitivity of envelope following responses to vowel polarity. *Hearing Research, 320*, 38–50.
Escera, C. (2017). The role of the auditory brain stem in regularity encoding and deviance detection. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.). *The frequency-following response: A window into human communication* (pp. 101–121). New York: Springer Nature.
Feldman, N. H., Griffiths, T. L., & Morgan, J. L. (2009). The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psychological Review, 116*(4), 752–782.
Gage, N., Poeppel, D., Roberts, T. P., & Hickok, G. (1998). Auditory evoked M100 reflects onset acoustics of speech sounds. *Brain Research, 814*(1), 236–239.
Guenther, F. H., Husain, F. T., Cohen, M. A., & Shinn-Cunningham, B. G. (1999). Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America, 106*, 2900–2912.
Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., & Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research, 47*(1), 46–57.
Holt, L. L., & Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception & Psychophysics, 72*, 1218–1227.
Intartaglia, B., White-Schwoch, T., Meunier, C., Roman, S., Kraus, N., & Schön, D. (2016). Native language shapes automatic neural processing of speech. *Neuropsychologia, 89*, 57–65.
Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech.* Stamford, CT: Singular/Thomson Learning.
Krishnan, A. (2002). Human frequency-following responses: Representation of steady-state synthetic vowels. *Hearing Research, 166*, 192–201.
Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories: Monkeys do not. *Perception & Psychophysics, 50*, 93–107.
Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B, 363*, 979–1000.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect”. In W. Strange (Ed.). *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore, MD: York Press.
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.
Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven, & N. Warner (Vol. Eds.), *Laboratory Phonology: Vol. 7*, (pp. 637–676). Berlin: Mouton.
Lindblom, B., & Engstrand, O. (1989). In what sense is speech quantal? *Journal of Phonetics, 17*, 107–121.
Liu, Y.Y., Polka, L., Masapollo, M., & Ménard, L. (in preparation). Disentangling the roles of formant proximity and stimulus prototypicality on asymmetries in vowel perception.
Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067).
Masapollo, M., Polka, L., & Ménard, L. (2017b). A universal bias in adult vowel perception – By ear or by eye. *Cognition, 166*, 358–370.
Masapollo, M., Polka, L., Molnar, M., & Ménard, L. (2017a). Asymmetries in vowel perception reveal a universal bias in adult vowel perception. *Journal of the Acoustical Society of America, 141*, 2857–2869.
Masapollo, M., Zhao, T. C., Franklin, L., & Morgan, J. L. (2019). Asymmetric discrimination of non-speech tonal analogues of vowels. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0000603.
Ménard, L., Schwartz, J.-L., & Böe, L.-J. (2004). The role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *Journal of Speech, Language, and Hearing Research, 47*(5), 1059–1080.
Miller, J. L., & Eimas, P. D. (1996). Internal structure of voicing categories in early infancy. *Perception and Psychophysics, 38*, 1157–1167.
Molnar, Polka, Baum, & Steinhauer (2014). Learning two languages from birth shapes pre-attentive processing of vowel categories: Electrophysiological correlates of vowel discrimination in monolinguals and simultaneous bilinguals. *Bilingualism: Language and Cognition, 17*(3), 526–541.
Naatanen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., et al. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 1997*(385), 432–434.
Nam, Y. J. (2014). *The role of acoustic-phonetic bias in consonant manner perception*Ph.D. Thesis. McGill University.
Nam, Y., & Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition, 155*, 56–66.
Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication, 41*, 221–231.
Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics, 39*, 467–478.
Polka, L., Ruan, Y., & Masapollo, M. (2019). Understanding vowel perception biases – it's time to take a meta-analytic approach. In A. M. Nyvad, M. Hejna, A. Hojen, A. B. Jesperson, & M. Hjortshoj (Eds.). *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn* (pp. 561–582). .
Schwartz, J.-L., Abry, C., Boë, L. J., Ménard, L., & Vallée, N. (2005). Asymmetries in vowel perception, in the context of the Dispersion- Focalization Theory. *Speech Communication, 45*, 425–434.
Schwartz, J.-L., & Escudier, P. (1989). A strong evidence for the existence of a large- scale integrated spectral representation in vowel perception. *Speech Communication, 8*, 235–259.
Shiga, T., Althen, H., Cornella, M., Zarnowiec, K., Yabe, K., & Escera, C. (2015). Deviance-related responses along the auditory hierarchy: Combined FFR, MLR and MMN evidence. *PLoS One, 10*(9), e0136794.
Skoe, E., Chandrasekaran, B., Spitzer, E. R., Wong, P. C., & Kraus, N. (2014). Human brainstem plasticity: The interaction of stimulus probability and auditory learning. *Neurobiology of Learning and Memory, 109*, 82–93.
Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear & Hearing, 31*(3), 302–324.
Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2013). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex, 25*(6), 1415–1426.
Slabu, L., Grimm, S., & Escera, C. (2012). Novelty detection in the human auditory brainstem. *The Journal of Neuroscience, 32*(4), 1447–1452.
Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3–46.
Stevens, K. N. (1998). *Acoustic phonetics.* Cambridge, MA: MIT Press.
Thomson, J. M., Goswani, U., & Baldeweg, T. (2009). The ERP signature of sound rise time changes. *Brain Research, 1254*, 74–83.
Tichko, P., & Skoe, E. (2017). Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. *Hearing Research, 348*, 1–15.
Tsuji, S., & Cristia, A. (2017). Which acoustic and phonological factors shape infants' vowel discrimination? Exploiting natural variation in InPhonDB. *Proceedings of Interspeech,* 2108–2112.
Zhao, C. T., & Kuhl, P. K. (2018). Linguistic effect on speech perception observed at the brainstem. *Proceedings of the National Academy of Sciences, 115*(35), 8716–8721.