Directional asymmetries reveal a universal bias in adult vowel perception [1]

Matthew Masapollo [2,3]

Linda Polka [2,1]

Monika Molnar [2,1, a]

School of Communication Sciences and Disorders, McGill University, 2001 McGill College, 8th Floor, Montreal, QC, H3A 1G1, Canada

[a] *BCBL. Basque Center on Cognition, Brain and Language. Donostia, Spain*

Lucie Ménard [2]

Département de Linguistique, Université du Québec à Montréal, Pavillon J.-A. De sève, DS-4425, 320, Sainte-Catherine Est, Montréal, QC, H2X 1L7, Canada

Running title: Universal bias in adult vowel perception

Abstract

Research on cross-language vowel perception in both infants and adults has shown 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 whose adjacent formants are close in frequency, which concentrates acoustic energy into a narrower spectral region). An alternative, but not mutually exclusive, account is that such effects reflect an experience-dependent bias favoring prototypical instances of native-language vowel categories. To disentangle the effects of focalization and prototypicality, we first identified a certain location in phonetic space where vowels were consistently categorized as /u/ by both Canadian-English and Canadian-French listeners, but that nevertheless varied in their stimulus goodness (i.e., the best Canadian-French /u/ exemplars were more focal compared to the best Canadian-English /u/ exemplars). In subsequent AX discrimination tests, both Canadian-English and Canadian-French listeners performed better at discriminating changes from less to more focal /u/'s compared to the reverse, regardless of variation in prototypicality. These findings demonstrate a universal bias favoring vowels with greater formant convergence that operates independently of biases related to language-specific prototype categorization.

PACS numbers: 43.71.Hw, 43.71.Es, 43.71.An

I. Introduction

Many aspects of human cognition are shaped by experience, while others appear to be governed by universal principles. A specific example comes from speech perception, where the fundamental perceptual operation is the mapping of an input acoustic signal onto phonological units. Considerable evidence, spanning some 60 years of research, indicates that this mapping process is influenced by both the intrinsic acoustic-phonetic properties of speech sounds and the structure of the language-specific phoneme inventory (see Cutler, 2012, for a review). The current work examines the nature of this complex interplay between universal and experiential factors in the context of an important phenomenon in speech perception: directional asymmetries in vowel discrimination.

Research on cross-language vowel perception in both infants and adults has shown that for many between-category vowel contrasts, discrimination is easier when the same pair of vowels is presented in one direction compared to the reverse direction (see Polka & Bohn, 2003, 2011, for reviews). For example, Polka and her colleagues found that both German-learning and English-learning infants performed better at discriminating the change from German /y/ to /u/, compared to the reverse change from /u/ to /y/ (Polka & Werker, 1994; Polka & Bohn, 1996). These directional asymmetries have been reported in numerous infant vowel discrimination studies using a wide range of vowel contrasts from across phonetic space in several behavioral paradigms (i.e., habituation, operant conditioning). Figure 1a shows some of the vowel contrasts that have been examined in infant vowel discrimination studies; each contrasting vowel pair is connected with an arrow indicating the direction of change that was easier to discriminate.

-- Insert Figure 1 about here --

Figure 1: (a) Schematic illustration of articulatory-acoustic vowel space based on the first two formants. Vowel contrasts showing asymmetries in infant vowel perception are plotted; arrows point in the direction that is easier to discriminate. The light gray area highlighted in this space corresponds to the region covered by the four vowel series synthesized for use in the present study. The dark grey area highlighted in this space corresponds to the region covered by the vowel series synthesized for use in Kuhl (1991). (b.) Magnified view of the vowel stimuli presented to English- and French-speaking listeners for identification and goodness ratings (Experiment 1). The formant frequencies were equally spaced on a psychophysical basis (on the bark scale; Zwicker & Terhardt, 1980). Embedded within this space are the tokens that were consistently identified as exemplars of English /i/ or French /y/ (F1=275 Hz and 300 Hz; F2=1753 Hz to 2202 Hz) and as exemplars of English /u/ or French /u/ (F1=275 Hz and 300 Hz; F2= 548 Hz to 979 Hz; outlined in black). (c) The six /u/ tokens used in the discrimination task (Experiment 2) are outlined in black, and labeled u_1 thru u_6. Stimuli u_1, u_2, and u_3 formed the less focal vowel set, and stimuli u_4, u_5, and u_6 formed the more focal set. The arrow points in the direction that NRV predicts will be easier to discriminate (see text for explanation). (d) The formant frequency values (in Hz) for the /i/ vowel stimuli used in Kuhl (1991). P is the "prototype" vowel and NP is the "non-prototype" as specified by Kuhl (1991). The gray area highlighted in this space corresponds to vowel stimuli on a common vector shared between P and NP. Kuhl (1991) reported directional asymmetries in infant discrimination of the stimuli along this common vector; the arrow points in the direction that was easier for infants to discriminate.

Critically, these asymmetries cannot be explained by reference to native vowel inventories because they emerge in infants' discrimination of both native and non-native (foreign language) vowel contrasts. These effects are also not consistent with a bias related to simple acoustic dimensions, such as pitch, amplitude or duration, as these variables were well-controlled in the test stimuli used across studies. However, these asymmetries in general, with few exceptions, could be predicted by considering the relative position of each vowel within articulatory/acoustic vowel space (defined by F1-F2). More precisely, infants tend to perform better at discriminating a change from a relatively less to a relatively more peripheral vowel, compared to the same change presented in the reverse direction. Although these findings were initially interpreted as a peripheral vowel bias, as explained below, with further research it became clear that this bias is related to formant convergence patterns that involve more than just F1 and F2.

Polka and Bohn (2003) initially proposed that this early vowel bias plays an important role in the development of vowel perception by establishing stable referents that help young infants attend to and differentiate vowels during the period when they are learning phonetic categories. Broadly consistent with this idea, more recent studies have shown that linguistic experience fine-tunes this initial vowel bias to optimize access to native-language vowel categories during speech processing (Polka & Werker, 1994; Polka & Bohn, 2011; Pons, Albareda-Castellot, Sebastián-Gallés, 2012; Dufour, Brunelliere, & Nguyen, 2013; Tyler, Best, Faber & Levitt, 2014). For example, with respect to German /u-/y/, monolingual English-speaking adults continue to show the same asymmetry as English- and German-learning infants, while German-speaking adults show symmetric (and near perfect) discrimination of this contrast (Polka & Bohn, 2011). A similar pattern of developmental change emerged when Danish-speaking adults and Danish-learning infants were tested on a native contrast (i.e., Danish /e/-/ø/) and a non-native contrast (i.e., British English /æ/-/ɯ; Polka & Bohn, 2011), and when [redacted]panish- and Catalan-learning infants were tested on discrimination of Catalan /i-e/ (Pons *et al*., 2012). While a developmental shift from a universal to a language-specific pattern of vowel perception is expected for non-native contrasts, it is important to note that asymmetries in adults' perception of some *native* contrasts have also been observed (Repp, Healy, & Crowder, 1979; Cowan & Morse, 1986; Repp & Crowder, 1990).

On the basis of these findings, Polka and Bohn (2011) more recently formulated the Natural Referent Vowel (NRV) Framework. According to this framework, young infants from across cultures come to the task of language acquisition universally biased toward certain vowels which act as natural reference points (or perceptual attractors) within

phonetic space. As infants accrue experience listening to a specific language, the perceptual vowel space is fine-tuned to align with the regularities of the native-language vowel system (Polka & Werker, 1994; Polka & Bohn, 2011). Nevertheless, this initial bias will continue to operate in adult language users, emerging most clearly in the perception of non-native vowel contrasts.

While recent progress has been made in interpreting asymmetries, we are still left with the question as to *why* perceivers are universally biased towards some vowels over others. As mentioned above, Polka and Bohn's (2003) initial hypothesis was that this bias favored vowels in the periphery of phonetic space. However, several asymmetries that were not predicted by this view revealed that the simple description of vowels within F1/F2 acoustic space was not adequate. Instead, researchers needed to consider additional formants (not just F1 and F2) and also the spectral proximity of formants to each other. NRV directly addresses this issue by postulating that asymmetries reflect a universal perceptual sensitivity to formant proximity (Polka & Bohn, 2011; see also, Schwartz Abry, Boë, Ménard, & Vallée, 2005). The fundamental idea is that extreme vocalic articulations give rise to acoustic signals with well-defined spectral prominences due to formant frequency convergence, or "focalization"; such articulations in turn lead to acoustic signals that exhibit increased spectral salience and acoustic-perceptual stability. To clarify, a good deal of research has shown that when spectrally adjacent formants move close together in frequency there is a mutual reinforcement of their acoustic energy, such that the amplitude of each formant increases. As a result, when formants converge acoustic energy becomes focused into a narrower spectral region (see Stevens, 1989; Kent & Read, 2002, for a discussion). Critically, a maximal degree of formant convergence is observed for vowels found at the periphery of phonetic space, which also have the most extreme vocal tract postures. For example, F2 and F3 and F4 converge (i.e., are close in frequency) for /i/ (which is the highest front vowel), and F1 and F2 are spectrally close to each other for /a/ (which is the lowest back vowel) as well as /u/ (which is the highest back vowel). Because the corner vowels (/i/ /u/ /a/) exhibit a maximal degree of formant convergence, they have been referred to as "focal vowels" in the speech literature (Schwartz *et al*., 1997).[6] Thus, the hypothesis is that more "focal" vowels have a privileged perceptual status due to their well-defined spectral prominences (i.e., they are easier to detect and encode in memory compared to vowels with less well-defined spectral prominences). For many, but not all, vowel contrasts a relatively more versus a relatively less peripheral location in the standard F1/F2 space aligns with differences in focalization. However, it is important to note that focalization and peripherality are conceptually not the same. Focalization takes into accounts the first 3-4 formants and their spectral position relative to each other, not just the vowel position within a simple F1/F2 space.

Indeed, there is some experimental evidence consistent with the claim that discrimination asymmetries are driven by focalization. Schwartz and Escudier (1989) tested French adult listeners on their ability to discriminate exemplars of French /e/

synthesized with the same F1, F2 and F4 values, but with different F3 values. Although all tokens were perceived as /e/, one vowel variant had an F3 frequency that was an equal psychophysical distance between F2 and F4, another had an F3 spectrally closer to F2 and yet another had an F3 closer to F4. Thus, these /e/ vowel tokens systematically differed in their degree of formant proximity. French adults showed directional asymmetries consistent with an effect of focalization. Specifically, discrimination was better when they heard a less focal /e/ followed by a more focal /e/, compared to when the same vowels were presented in the reverse order (i.e., more focal /e/ followed by less focal /e/). Although asymmetric patterns were reported in this study, these effects were interpreted as focalization effects only later when similar asymmetries emerged in infant vowel perception research (Schwartz, Abry, Boë, Ménard, & Vallée, 2005).

There are several reasons, however, to suggest that language experience could have played a role in shaping the perceptual asymmetry observed by Schwartz and Escudier (1989). In a seminal study by Kuhl (1991), it was reported that language experience affects listeners' perception of vowel stimuli from *within* a given phonetic category. She presented English adult listeners with a range of synthetic /i/ vowels that systematically varied in their first (F1) and second formants (F2). Listeners consistently perceived the stimuli in a particular part of vowel space as better exemplars of /i/, indicating that the category has a graded, internal structure. Interestingly, the stimuli perceived as prototypic category members also matched the average acoustic production values of /i/ (Peterson & Barney, 1952), suggesting that there may be a close correspondence between the mean stimulus values experienced in the input signal and the stored prototype (although, this outcome conflicts with other studies, which show that prototypicality judgments are often more peripheral compared the the average production values reported in a corpus distribution [see, e.g., Johnson, Wright, & Flemming, 1993; Lively & Pisoni, 1997; Diesch, Iverson, Kettermann, & Siebert, 1999; Whalen, Magen, Pouplier, Kang & Iskarous, 2004]).

On the basis of these findings, Kuhl (1991) identified a good exemplar as the prototype, and a poor exemplar as the non-prototype. She then synthesized 32 category variants that orbited the prototype and non-prototype in equal psychophysical steps (Figure 1d presents a F1/F2 plot of Kuhl's [1991] /i/ stimulus array). Using a change detection paradigm, she found that both English-speaking adults and English-learning infants (at 6-months of age) showed greater discrimination of the non-prototype from its variants, compared to discrimination of the prototype from its variants. In sharp contrast, rhesus macaques were found to discriminate the prototype and non-prototype variants of /i/ to the same degree. Kuhl (1991) reasoned that it might have been harder for the human adult and infant listeners to discriminate the prototype from its variants, than the non-prototype from its variants, because they (unlike the macaques) were organizing the stimuli into a category.

A second finding of the study by Kuhl (1991), reinforcing the interpretation that human vowel perception is influenced by category learning, was obtained in the discrimination task with infants. A subset of the /i/ stimuli were shared between the orbitals surrounding the prototype and non-prototype. An analysis of infants' discrimination of these stimuli revealed a directional asymmetry, such that they performed better at discriminating the change going in the direction from the non-prototypic to the prototypic /i/, compared to the same change presented in the reverse direction.

On the basis of these and other compatible findings, Kuhl (1993) proposed the Native Language Magnet (NLM) Theory, which argues that category learning influences perceptual patterns, such that listeners become biased toward native prototypes. The basic idea is that language-experience gives rise to prototypes that are perceptual “hot spots” that act like “perceptual magnets” and warp the perceptual space around the prototype. This in turn facilitates access to native vowel categories by increasing internal category cohesion, and maximizing the perceptual distance between category edges (see also, Feldman, Griffiths, & Morgan, 2009). The prototype’s magnetic properties can be demonstrated experimentally in two ways: 1.) As increased generalization (reduced discrimination) for vocalic exemplars close to the prototype in psychophysical space; and 2.) As a directional asymmetry in the discrimination of prototypic and non-prototypic exemplars that fall within a given vowel category; namely, poorer discrimination of a prototypic to non-prototypic change compared to a non-prototypic to prototypic change.

There are several reasons, however, to suspect that the asymmetry reported by Kuhl (1991) reflects something besides the proposed prototype magnet effect. The first concerns a methodological criticism of her study. Specifically, it has been argued that listeners might have shown enhanced discrimination of the non-prototype and its variants if the non-prototype was across a category boundary and was not consistently identified as /i/ (see, e.g., Lotto, Kluender & Holt, 1998). Since the listeners in Kuhl’s study were not explicitly asked to categorize the vowel stimuli, it was not certain whether the non-prototype was consistently identified as /i/. In fact, subsequent studies found that many listeners often identified the non-prototype as a different vowel (e.g., Iverson & Kuhl, 1995; Lively & Pisoni, 1997; Lotto *et al*., 1998, reported that many listeners identified the non-prototype as /e/), suggesting that the stimuli in Kuhl (1991) spanned different phonetic categories. This raises the possibility that the directional asymmetries reported by Kuhl (1991) may simply reflect a category boundary effect (cf. Iverson & Kuhl, 1995, 2000). A second reason to question Kuhl's (1991) interpretation is that the prototype was more focal (between F2 and F3) compared to the non-prototype. Thus, as in the case of Schwartz and Escudier (1989), the prototype/non-prototype asymmetry is also equivocal. Specifically, it could be attributed to prototypicality effects, category boundary effects, or focalization effects.

Thus, there are several reasons to examine further the nature of directional asymmetries in vowel perception. According to the NRV framework (Polka & Bohn, 2011), vowel perception is shaped by both universal biases and experiential influences. Specifically, it is hypothesized that a universal perceptual bias favoring more focal vowels is present in early infancy; this bias will then be modified by linguistic experience in mature adult perceivers. Attunement to native-language phonetic categories begins in infancy (Kuhl *et al.,* 1992; Polka & Werker, 1994), which eventually causes directional asymmetries to fade for native vowel contrasts but remain in place for non-native vowel contrasts (Polka & Bohn, 2011; Pons *et al.*, 2012). A central claim of the NRV framework is that asymmetries reflect a universal bias that is phonetically grounded in vowel acoustics. Critically, however, the existing data do not provide definitive evidence that adults show asymmetries rooted in perceptual sensitivity to differences in formant proximity, independent of prototype categorization processes. The NLM Theory (Kuhl, 1993; Kuhl *et al*., 2008), on the other hand, has largely ignored the role of universal biases, and instead proposes that asymmetries are a by-product of phonetic category learning induced by exposure to statistical distributions of vowels present in input speech.

To address this theoretical tension, we designed a study to examine whether adult listeners from different language backgrounds are indeed sensitive to differences in formant proximity while discriminating subtle acoustic differences that fall within a given vowel category, and, if so, whether that sensitivity operates independently of language-specific prototype categorization. It differs from earlier work by Schwartz and Escudier (1989) and Kuhl (1991) in teasing apart the effects of focalization *and* stimulus goodness on vowel discrimination, thus providing a strong cross-language test of the competing predictions of NRV and NLM.

Guided by previous experimental findings, we chose Canadian-English and Canadian-French as languages expected to differ in their prototype structure of the vowel /u/. Escudero and Polka's (2003) study of vowel production in Canadian-English and Canadian-French measured the spectral properties of five analogous vowel categories (/i, y, u, ɛ, æ/) produced by six native speakers in each of the two languages. They found that Canadian-French speakers consistently produced more extreme /u/ gestures (resulting in lower F1 and F2 values) than Canadian-English speakers, with the mean location of Canadian-French /u/ being more peripheral than that of Canadian-English /u/ in phonetic space (see also, Martin, 2002; MacLeod, Stoel-Gammon, & Wassink, 2009; Noiray, Cathiard, Ménard, & Abry, 2011). Thus, the acoustic structure of Canadian-French /u/ is more focal (between F1 and F2) than that of Canadian-English /u/. Given these quantitative language-specific differences in the acoustic structure of /u/, we hypothesized that we might also observe language-specific differences in the prototype structure of /u/, such that more focal variants will be more representative of the Canadian-French /u/ category. Recall that Kuhl (1991) found a close correspondence between the typicality of /i/

exemplars and the acoustics of average production (based on /i/ production estimates reported in Peterson and Barney [1952]; cf. Lively & Pisoni, 1997; Diesch *et al*., 1999).

The present study investigated whether directional asymmetries can be ascribed to (universal) vowel focality effects and/or to (language-specific) prototypicality effects. Experiment 1 was a perceptual study designed to determine the detailed nature of the internal structure of the /u/ category in Canadian-English and Canadian-French. This first experiment allowed us to determine whether we could identify the critical stimulus items for testing the competing predictions of NRV and NLM. Specifically, it was required that we obtain variants of /u/ that systematically differ in both their formant proximity (between F1 and F2) and category goodness ratings. Experiment 2 tested the predictions by asking Canadian-English and Canadian-French listeners to discriminate synthetic variants of /u/ that we had chosen on the basis of the results of Experiment 1.

II. Experiment 1: Vowel identification and goodness ratings

Previous phonetic descriptions summarized in the Introduction indicate that /u/ has different quantitative properties in Canadian-English and Canadian-French (Escudero & Polka, 2003; MacLeod *et al*., 2009; Noiray *et al*., 2011). However, it is unknown whether there is a convergence between production and perception measures of /u/ in both languages. Thus, Experiment 1 had two goals. The first goal was to confirm that Canadian-English and Canadian-French listeners' phonetic category *goodness* judgments are indeed linked to differences in formant convergence patterns between F1 and F2. The second goal was to identify /u/ vowel stimuli that could be used to disentangle the effects of formant convergence and category goodness on asymmetries in adult vowel discrimination.

These two goals were accomplished by synthesizing an array of vowel stimuli that spanned the high/back region of the vowel space to include prototypical instances of Canadian-English /u/ and of Canadian-French /u/ that also systemically differed in the proximity between F1 and F2 (with the frequency distance between F1 and F2 varied in equal psychophysical steps). Native, monolingual Canadian-English and Canadian-French listeners then completed a categorization and goodness rating task with this array of vowels. We hypothesized that the location of stimuli perceived as the best /u/ exemplars would fall in different regions of phonetic space for each language group with the best instances of French /u/ being more focal variants compared to the best instances of English /u/.

A. Materials and Methods

1. Subjects

Twenty-six adults were recruited in Montreal, including 13 native, monolingual Canadian-English listeners (mean age =24 years [*SD*=4.4]; 5 males) and 13 native, monolingual Canadian-French listeners (mean age =25 years [*SD*=4.8]; 7 males). To avoid potential dialectal variation effects (see, e.g., Frieda *et al*., 1999), only monolingual English listeners from Ontario[7] and monolingual French listeners from Quebec were included in the study. Subjects were paid for their participation. Subjects had to meet the following inclusion criteria: 1) report no speech or hearing disorders and no prior linguistic or phonetics training, 2) raised in a monolingual home and educated in a monolingual school in their respective language, 3) no experience learning a second language before 10 years of age, and 4) no experience conversing in a second language on a regular basis. Four additional subjects (two English; two French) were tested, but excluded from the analysis because their responses were highly irregular, and unlike their peers, failed to reveal a distinct region corresponding to "good" instances of their native-language /u/ category within the stimulus array (see details concerning the identification and goodness rating task below).

Subjects completed either the Canadian-English or Canadian-French version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumfeld, & Kaushanskaya, 2007). As part of this questionnaire, participants self-rated their speaking and listening proficiency in the language that was foreign to them (in English for the French monolinguals and in French for the English monolinguals) using a 1-10 rating scale (1=very poor, 10=excellent). The English subjects' median self-ratings of their French proficiency were 1.0 (*range* = 1-4) for speaking and 2.0 (*range* = 1-4) for understanding spoken French. The French subjects' median self-ratings of their English proficiency were 3.0 for speaking (*range* = 1-4) and 3.0 for understanding spoken English (*range* = 1-4). Median tests showed that the English proficiency ratings by French adults and French proficiency ratings by the English adults were not statistically different ($p > .05$) for either speaking or comprehension.

2. Stimuli

The vowel stimuli used in Experiment 1 included 34 tokens, which are highlighted with gray shading on Figure 1b. These items were selected from the larger array of 128 vowel tokens shown in Figure 1b, which consisted of four series of isolated vowels varying in F1 (from 275 to 330 Hz) and F2 (from 476 to 2303 Hz) in equal psychophysical steps on the bark scale (Zwicker & Ternhardt, 1980).

The stimuli were synthesized using the Variable Linear Articulatory Model (Maeda, 1979, 1990; Boë, 1999; Ménard, Schwartz, & Boë, 2004), which generates high-quality acoustic vowel signals based on current understanding of the articulatory-acoustic relations of the human vocal tract. As described in detail in Ménard et al. (2004), the VLAM model is based on a statistical analysis of 519 cineradiographic images of a French speaker (Bothorel, Simon, Wioland, & Zerling, 1986). A Principal Component Analysis conducted on the midsagittal contours of the vocal tract revealed that 88% of the variance could be explained by the seven articulatory parameters. Each of these parameters is adjustable at a value in the range of ±3.5 SD of the mean values for this articulator in the cineradiographic images. The model can generate all vowels of the world's languages, without reference to the corpus used to extract the control parameters. Furthermore, in a study of perceptual categorization by French and English listeners, synthesized stimuli covering the model's maximal vowel space were generated (Ménard *et al*., 2004; Ménard, Davis, Boë, & Roy, 2009). It was shown that native French vowels as well as native English vowels could be reliably identified by participants.

Critically, VLAM also successfully simulates the interaction among converging formant frequencies in an ecologically-valid manner. The stimuli were identical in all other respects: f0, F3, F4, and F5 were 120 Hz, 2522 Hz, 3410 Hz, and 4159 Hz, respectively. These tokens emulated an adult male voice. Each stimulus was 400-ms in duration, and had the same intonation contour. For English listeners, the four vowel continua varied perceptually from /u/ (‘oo”) to /i/ (“ee”) as F2 values increased. For French listeners, the continua varied perceptually from /u/ (‘oo”) to /y/ (as in the French word “but”) to /i/ (“ee”) as the F2 values increased

As a preliminary step, pilot testing was conducted with the full array of 128 vowel tokens to select a smaller set of vowels that are identified only as /u/ by all English and all French listeners. Specifically, five monolingual Canadian-English and five monolingual Canadian-French listeners completed a phonetic identification and goodness rating task with the full array. They listened to each token three times in a randomized order and had to first decide whether the sounds they heard matched any of their native-language vowel categories by clicking on a word on a computer screen representing one of the target vowels. They then had to rate the quality (or category "goodness") of each vowel using a

1-5 rating scale (1=very poor, 5=very good).

Based on the results of this identification and goodness rating task, we then selected the 34 vowels corresponding to the tokens highlighted in gray in Figure 1b; the excluded vowels were judged as a non-match or a match with low goodness rating (below 2.4 on average). This final stimulus set included 22 back vowels targeting English /u/ and French /u/ vowel (F1=275Hz and 300Hz; F2= 548Hz to 979Hz), and 12 front vowels targeting English /i/ and French /i/ and /y/ (F1= 275Hz and 300Hz; F2=1753Hz to 2202Hz). Note that we also selected clear instances of /i/ and /y/ and synthesized two additional filler vowels (/o/ ["oh"] and /ə/ ["uh"]) to include in the stimulus set to provide some variation in vowel quality during the perception task. This also made it easier to assess whether subjects were successful in identifying vowel quality differences using key words.

-- Insert Figure 2 about here --

Figure 2: Composite /u/ ("oo") category goodness scores for each vowel in the stimulus array plotted for the English- (top) and French-speaking adults (bottom; Experiment 1). The relative differences in the median composite scores (which combines /u/ goodness ratings across all test trials; see text for explanation) are displayed as differences in the size of the corresponding circle (larger for higher median

composite score). The upper number within each circle is the median composite score for each stimulus, ranging from 0 (worst) - 50 (best). The lower number within each circle is the number of subjects who assigned that token the highest goodness score within the array. The six stimuli outlined in black are the ones that were selected to be used for the AX discrimination tests (Experiment 2). The arrows point in the direction that NLM predicts will be easier for each language group to discriminate (see text for explanation).

3. Procedure and Design

Listeners were presented with randomized sequences of the 36 vowel stimuli (the 34 initial tokens [highlighted in Figure 1b] and the 2 filler vowels) and were asked to identify each token in terms of their own native-language vowel categories, and then judge its category goodness using a 1-5 rating scale (1=very poor, 5=very good). Each test session included 360 trials (10 trials for each stimulus).

The experiment was conducted in a sound-treated laboratory room with participants facing a computer screen. The stimuli were presented over headphones at a comfortable loudness level of 65 dB in a self-paced manner; after each vowel presentation, subjects decided themselves how much time they needed to respond. The identification and goodness rating responses were collected using Praat (Boersma & Weenink, 2016). During stimulus presentation, words containing the target vowels and a rating scale appeared on the computer screen. By choosing and clicking on one of the words, the subjects decided what sound they heard (identification). Then, using a 5-point rating scale, they judged the quality of the sounds (category goodness).

For the English-speaking subjects, the following target words appeared on the screen: ‘boo’ (for the vowel /u/), ‘bee’ (for the vowel /i/), bowl’ (for the vowel /o/), ‘bug’ (for the vowel /ə/) and an ‘X’ to choose if the vowel could not be matched to any of the other vowel choices. The rating scale, presented at the bottom of the screen, consisted of five boxes labeled from 1 to 5. The scale endpoints were also labeled with “very poor exemplar (1)” and “very good exemplar (5).” For the French-speaking subjects, the following target words were presented on the screen: ‘boue’ (for the vowel /u/), ‘bu’ (for the vowel /y/), ‘bille’ (for the vowel /i/), ‘bas’ (for the vowel /a/) and ‘beau’ (for the vowel /o/) and option ‘X’ for no match.

B. Results

As mentioned above, in this study we were concerned only with the identification and goodness ratings of the 22 back vowels indicated in Figure 1b and 1c. Recall that subjects identified and rated each stimulus 10 times. The entire array of 22 back vowels was consistently identified as /u/ by both French and English adults; identification rate (across all subjects) was 100% for many tokens and was above 85% for all tokens. For each subject, a composite goodness score (hereafter referred to as goodness score) was computed for each vowel by adding up the ratings across every trial in which the token was categorized as /u/; trials in which the vowel was not identified as /u/ received a 0 score. Thus, goodness scores ranged from 0 (never identified as /u/) to 50 (identified as /u/ and assigned the highest rating [5] on all 10 trials). The median goodness scores for each /u/ token are plotted in Figure 2 with the results from the English adults on the top and the results from the French adults on the bottom. Each circle represents a vowel token. The circle size is scaled to represent the category goodness differences as indexed by the goodness score; the median goodness score for the group is shown in the center of each circle and below it is the number of subjects in the group for whom that token was rated as the best exemplar of the /u/ category (i.e., had the highest goodness score).

Important aspects of the expected language-specific patterns were confirmed. Non-parametric tests showed that there were significant differences in /u/ category goodness scores within each language group. Wilcoxon signed rank tests were conducted to examine the effects of F1 value on goodness scores. This test was implemented within each group, to first compare overall goodness scores for vowels across the two F1 series (275 Hz vs. 300 Hz) after collapsing across all F2 levels within each series, and then also at each level of F2. For English adults, overall goodness scores were significantly higher ($p<.01$) for vowels with higher F1 (300 Hz > 275 Hz); this effect was significant for the all vowels with F2 values of 790 Hz or higher. For French adults there was no overall effect of F1 on category goodness ($p>.05$); however, when F2 was 548 Hz category goodness was higher for the token with F1 of 275 Hz.

With respect to F2 frequencies, as shown in Figure 2, English adults assigned the highest goodness scores to tokens with the highest F2 frequencies, whereas the French adults assigned the highest goodness scores to tokens in the middle of each F2 series. Friedman ANOVAs were conducted to analyze rank differences in category goodness related to F2 frequency across subjects. Separate ANOVAs were conducted for each F1 series within each language group. All four ANOVAs revealed highly significant effects of F2 frequency on category goodness. For English adults, $F = 52.2$ ($p<.001$) for the F1=275 Hz series and $F = 83.3$ ($p<.001$) for the F1=300 Hz series. For French adults, $F= 205.4.2$ ($p<.001$) for the F1=275 Hz series and $F = 199.7$ ($p<.001$) for the F1=300 Hz series.

To probe the goodness scores in more detail, mean goodness scores were computed for F2 variants grouped into 3 subsets: high F2 tokens (979 Hz, 929 Hz, 881

Hz & 835 HZ), mid F2 tokens (790 Hz, 746 Hz, 704 Hz, and 663Hz) and low F2 tokens (625 Hz, 5895 Hz, and 548 Hz). The subsets scores were computed separately for each F1 series (F1=275 Hz vs. F1=300 Hz) for each language group. Mean scores on these subsets (high vs. mid, mid vs. low, and high vs. low) were then compared using Wilcoxon sign ranked tests (p<.01) to evaluate how goodness scores differed as F2 varied. Overall, the results confirmed the language differences shown in Figure 2. For English adults, goodness scores decreased as F2 decreased; significant differences were found across all 3 comparisons with high F2 subset > mid F2 subset > low F2 subset. This pattern was found for each F1 series. For French adults, the goodness scores were higher for the mid F2 subset compared to both the high F2 and low F2 subsets for the F1=275 series. For the F1=300 Hz series, the goodness scores for the mid and high F2 subsets did not differ and both were higher than the low F2 subset.

Using acoustic and category goodness measures, we then selected two sets of vowels that differ systematically in both their category goodness measures and F1 and F2 formant frequency convergence patterns. These sets are indicated in Figures 1c and 2. The more focal set (u_4, u_5, u_6) included 3 tokens from the F1=275 Hz series with a relatively small spectral distance between F1 and F2; these tokens were also among those assigned the highest goodness scores by French adults. The less focal set (u_1, u_2, u_3) included 3 tokens from the F1=300 Hz series in which there is a relatively larger spectral distance between F1 and F2 (compared to the more focal set); these tokens were among those assigned the highest goodness scores by English adults.

Cross-group differences in the category goodness scores assigned to the less focal and more focal vowel sets were evident at the group and individual level. For French adults, median goodness scores were higher for the more focal set (*median* = 40) than the less focal set (*median* = 34), whereas for English adults the opposite pattern emerged with higher scores for the less focal (*median* = 39.7) than the more focal set (*median* = 31.7). Wilcoxon sign ranks tests confirmed a statistically robust difference in goodness score in the English group (less focal > more focal); the opposite pattern observed in the French group (more > less focal) did not reach statistical significance.[8] Median tests also revealed significantly higher goodness scores for the French adults compared to the English adults for the more focal set, $\chi^2 = 2.52$, $p < .05$); the reverse pattern (English > French) was observed for the less focal set, but this difference was not statistically supported.

On the individual level, every English subject showed the same pattern with respect to median goodness scores (less focal > more focal), and 9 of the 13 subjects assigned their highest goodness score to a vowel in the less focal set. In the French group, 9 of 13 subjects showed the opposite pattern with respect to median goodness scores (more focal > less focal); 8 of 13 French adults also assigned their highest goodness score to a vowel in the more focal set. However, neither of the later proportions exceed chance predictions.

C. Discussion

The results of Experiment 1 are clear in showing that Canadian-English and Canadian-French listeners are highly sensitive to the distinct acoustic-phonetic properties that specify their native /u/ category. While listeners from both language groups identified all of the vowel stimuli as unambiguous members of their /u/ category, there was evidence that the differences in their long-term linguistic experiences altered which stimuli within the category were perceived as the best exemplars of /u/. The best /u/ exemplars for

Canadian-French listeners were more focal with respect to F1 and compared to the best /u/ exemplars for Canadian-English listeners.

Additionally, there was a close correspondence between the average formant values reported in previous cross-linguistic English-French vowel production studies and the way in which the /u/ category appears to be structured in perception across the two languages. Previous acoustic-phonetic analyses show that Canadian-French /u/ tends to be produced with significantly lower F1 and F2 values than English /u/ vowels (Escudero & Polka, 2003; MacLeod *et al*., 2009; Noiray *et al*., 2011). This pattern is in accord with the present perceptual data, which shows that vowels with lower F1 and F2 values are perceived as better /u/'s by French listeners than English listeners. These findings are not surprising given that these detailed differences must be learned by native speakers, because they have consequences for category boundaries in perception and because they must be accurately reproduced to achieve a native accent.

Finally, and of critical importance to the present study, these cross-language differences in perception provide a stimulus array that can be used to evaluate the rival (although not mutually exclusive) predictions of NLM and NRV. For French, focalization differences and category goodness differences are aligned (i.e., the more focal vowel set contains prototypic French /u/ variants). For English, however, these factors are not aligned (i.e., the less focal vowel set contains prototypic English /u/ variants). Thus, the less-focal/English-prototypic and more-focal/French-prototypic vowel sets define stimulus conditions that we can use to test and compare discrimination performance in English and French listeners in order to disentangle the effects of focalization and prototypicality on directional asymmetries in vowel discrimination.

III. Experiment 2: Vowel discrimination

As described above in the Introduction, the NLM model proposes that directional asymmetries reflect language-specific categorization processes, and therefore predicts that listeners should display enhanced sensitivity for discriminating a change from a relatively "poor" to relatively "good" native vowel category exemplar. On the other hand, the NRV framework proposes that universal perceptual biases and language experience interact to shape vowel perception in mature perceivers, and thus predicts that both factors may contribute to directional asymmetries in adult vowel discrimination. Universal biases are phonetically grounded in formant proximity and drive directional asymmetries in which perceivers show enhanced sensitivity for discriminating a change from a relatively less to a relatively more focal vowel. Experiment 2 was designed to address the merit of each perspective.

Toward this end, we tested whether there are directional asymmetries in Canadian-English and Canadian-French listeners' discrimination of within-category /u/ vowel pairs, using a standard same-different (AX) task. Based on the findings of Experiment 1, we were able to select a set of less-focal/English-prototypic /u/ vowels and more focal/ French-prototypic /u/ vowels, that allowed us to systematically evaluate the impact of these factors. For these select stimuli, NRV and NLM predict asymmetries in the *same* direction for French adults, but in *opposite* directions for the English adults. More precisely, if focalization effects alone drive directional asymmetries, then we should observe the same direction effect in both language groups, i.e., enhanced performance when discriminating a change from a relatively less to a relatively more focal /u/ vowel, compared to the reverse direction.

Alternatively, if language-specific categorization processes alone drive direction asymmetries, then we should observe asymmetries going in opposite directions across the two language groups with enhanced discrimination for a change from a relatively “poor” to a relatively “good” /u/ vowel exemplar. Importantly, these factors are not mutually exclusive and thus both may play a role. This is the perspective taken in the NRV framework, which proposes that linguistic experience acts to alter initial universal vowel biases. In this case, we hypothesized that we would observe a main effect of focalization, as well as an interaction with native language. Specifically, we predicted that both language groups will show an asymmetry due to focalization, i.e., better discrimination in the less to more focal direction. However, this asymmetry will be more robust for French listeners given that it is aligned with the predicted effects of prototype categorization, and weaker in English perceivers given that it is misaligned with the predicted effects of prototype categorization.

A. Materials and Methods

1. Subjects

Thirty adults were recruited in Montreal including 15 monolingual Canadian English speaker-listeners (mean age = 23 years [*SD*=5.1]; 6 males) and 15 monolingual Canadian French speaker-listeners (mean age = 36 years [*SD*=12.6]; 9 males). Subjects were paid for their participation. Eight additional subjects were tested, but excluded from the analysis for the following reasons: equipment failure (*n*=1); subjects did not follow task instructions(*n*=7)[9].

As in Experiment 1, English listeners were from Ontario and French listeners were from Quebec and met the same inclusion criteria as the subjects tested in Experiment 1.

Subjects also completed either the Canadian English or Canadian French versions of the LEAP-Q (Marian *et al*., 2007, which included the self-rating of their speaking and listening proficiency in the other language on the same 10-point scale described above in Experiment 1 [1=very poor, 10=excellent]). The English subjects' median self-ratings of French proficiency was 2.0 for speaking (*range* = 1-4) and 2.0 for understanding (*range* = 1-5).

The French subjects' median self-ratings of English proficiency were 5.5 for speaking (*range* = 2-7) and 7.0 for understanding (*range* = 2-8). Median split tests showed that French subjects rated their English proficiency significantly higher than the English subjects rated their French proficiency for both speaking (p<.005) and listening (p< .005).

Median split tests also showed no difference between English adult's ratings of French proficiency across Experiments 1 and 2 for either speaking or listening. However, French adults rating of English proficiency were significantly higher in Experiment 2 than Experiment 1 for both skills (p<.005), despite having applied the same inclusion criteria as Experiment 1.

2. Stimuli

A subset of six of the 22 stimuli from Experiment 1 were selected for use in this second experiment. The selected stimuli are marked and labeled in Figure 1c. As already described, these stimuli were of two types (3 tokens each): Less-focal/English-prototypic /u/ and more-focal/French-prototypic /u/.

3. Procedure and Design

Participants completed an AX discrimination task (e.g., Polka, 1992; Iverson & Kuhl, 1995). On each trial, participants heard two stimuli, and then judged whether they were the "same" or "different." A long ISI (1500 ms) was used to ensure that a "phonetic level” of analysis was being tapped (i.e., by exerting greater processing demands on auditory working memory and attention; see, e.g., Werker & Logan, 1985; Cowan & Morse, 1986; Repp & Crowder, 1990). Participants initiated a trial by pressing a response key, and then pressed one of two labeled buttons on a keyboard to indicate whether the second stimulus was the same as the first [A] or different from the first [X]. Prior to the start of the experiment, participants were informed that all of the sound changes that they would hear were subtle, and that they should respond to any differences that they heard in the stimuli. These instructions were meant to replicate the task demands employed in previous studies by Kuhl and her colleagues investigating the effects of category goodness on vowel discrimination in adulthood (Kuhl, 1991; Iverson & Kuhl, 1995, 2000).

Before the test trials started, participants completed six practice trials. During the test trials, participants heard every possible pairing of the six stimuli (including each stimulus being paired with itself), 5 times, in both presentation orders. There were 180 test trials in a session (150 different, 30 [acoustically identical] same). The experiment was broken up into five blocks. Each block had 36 trials, which consisted of one presentation of each possible pairing. No feedback was provided on either the practice or experimental trials. Participants took short breaks as needed.

The experiment was conducted in a sound-treated laboratory room with participants facing a computer screen and with a keyboard in front of them. The stimuli were delivered to both ears through insert earphones (an echo-attenuated plastic tube system terminating in a foam earplug) at 60 dB SPL, using Presentation (Neurobehavioral Systems Inc.).

B. Results

Our analysis focused on listeners' discrimination of the nine "cross-set" stimulus pairs that contained a vowel from the less focal set (stimuli u_1, u_2, u_3) and a vowel from the more focal set (stimuli u_4, u_5, u_6). To ensure that differences in discrimination performance did not reflect an inherent bias to respond "same" or 'different," each subject's performance on each cross-set stimulus pair was converted to an *A'* score.[10] *A'* is a non-parametric unbiased index of performance that ranges from .50 (chance) to 1.0 (perfect discrimination).[11] For each participant, an *A'* score was computed for each of the nine cross-set stimulus pairs with separate scores computed for each direction of change (less to more focal and more to less focal, e.g. u_1/u_4 and u_4/u_1). The mean *A'* score for each direction of vowel change is plotted for each language group in Figure 3. These scores were submitted to an ANOVA with order of vowel change (less to more focal vs. more to less focal) and stimulus pair (u_1/u_4 vs. u_1/u_5 vs. u_1/u_6 vs. u_2/u_4 vs. u_2/u_5 vs. u_2/u_6 vs. u_3/u_4 vs. u_3/u_5 vs. u_3/u_6) as within-subject factors, and native language (English vs. French) as a between-subjects factor. A significant effect of native language [$F(1, 28) = 6.492$, $p = .017$, $\eta^2_p = .188$] revealed overall higher sensitivity in French listeners (M=0.97, SE=.005) compared to English listeners (M=0.94, SE=.007). A robust effect of order of vowel change [$F(1, 28) = 17.098$, $p = .000$, $\eta^2_p = .393$] revealed more accurate discrimination in the less to more focal direction (M=0.97, SE=.004) compared to the more to less focal direction (M=0.94, SE=.008). There was also a significant effect of stimulus pair [$F(8, 224) = 3.344$, $p = .001$, $\eta^2_p = .107$] and an order x stimulus pair interaction [$F(8, 224) = 2.359$, $p = .019$, $\eta^2_p = .078$]. The latter results show that the direction effect, although observed for each cross-set stimulus pairing, was more robust for some stimulus pairs than others. All other interactions were non-significant.[12,13]

-- Insert Figure 3 about here --

Figure 3: Native language (English vs. French) by order of vowel change (less to more focal vs. more to less focal) summary for the cross-set discrimination analysis. Mean percent *A'* scores for all different vowel pairs contrasting a vowel from the less focal vowel set (stimuli u_1, u_2, u_3) with a vowel from the more focal vowel set (stimuli u_4, u_5, u_6) for the English and French listeners separately. When discrimination performance is at chance, *A'* = 0.5.

Table 1 shows the mean *A'* scores for each stimulus pairing and order of vowel change collapsed across language groups. To probe the order x stimulus interaction, simple effects of direction were analyzed via post-hoc LSD *t*-tests conducted on each stimulus pairing; the results are reported in Table 1. Significant directional asymmetries consistent with a focalization effect emerged for stimulus pairs u_1/u_4, u_3/u_4, u_3/u_5, and u_3/u_6. The same directional pattern was observed for the remaining stimulus pairs, but did not reach statistical significance. The direction effect (based on mean *A'* scores for each direction collapsed across stimulus pairs) was also evident in individual performance. Thirteen out of the fifteen English participants were better at detecting the less to more focal /u/ vowel change (p = .007, two-tailed binomial test), and eleven out of the fifteen French participants tested were better at detecting the less to more focal /u/ vowel change (p = .118, two-tailed binomial test). The few French participants who failed to display a direction effect were performing at ceiling. Overall, the results strongly support the predictions of NRV that listeners are universally attuned to formant convergence patterns, and that this bias influences listeners' perception of subtle acoustic differences that fall within the /u/ category.

-- Insert Table 1 about here --

TABLE I. Mean *A'* values by stimulus pair and presentation order for the cross-set discrimination analyses (statistical values are listed on the far right).

C. Discussion

In Experiment 2, we investigated whether directional asymmetries emerged when Canadian-English and Canadian-French listeners discriminated variants that fell within the /u/ category, and, if so, whether those asymmetries were better predicted by focalization, prototypicality, or both factors. Overall, the findings reveal a robust effect of focalization alone. When discriminating pairs of vowels drawn from across the less-focal/English-prototypic and more-focal/French-prototypic sets, both English and French listeners showed the same directional asymmetry – both were better at discriminating the less to more focal /u/ changes, compared to the reverse direction. This direction effect was more robust for some cross-set stimulus pairs than others. Importantly, this direction effect did not interact with native-language, and thus appears to not be influenced by language experience. These findings provide the first direct evidence of a directional asymmetry in adult vowel discrimination that can be attributed unequivocally to formant focalization. These data provide firm support for a central claim of the NRV framework; namely, that there is a universal perceptual bias favoring vowels with more focal spectral quality that can be observed in adult, as well as infant, listeners.

IV. General Discussion

A fundamental goal of research on cross-language speech perception is to explicate the role of both universal and experiential factors (see Cutler, 2012). In the domain of vowel perception, considerable research has focused on examining how listeners' response patterns are influenced by both the intrinsic acoustic-phonetic properties of vowels and phonetic category learning. Research to date has provided ample evidence that the perception of vowels is initially influenced by their relative degree of formant proximity

(Polka & Bohn, 2003, 2011), but that this initial vowel bias dynamically adjusts to reflect the structure of the native-language phonological system (Kuhl, 1991; Kuhl *et al*., 1992; Polka & Werker, 1994; Polka & Bohn, 2011; Pons *et al*., 2012).

In the current work, we examined further how universal biases and phonetic category learning might interact to shape adults' perception of subtle acoustic-phonetic differences that fall within a given vowel category. Specifically, we investigated the effects of formant proximity and prototypicality on Canadian-English and Canadian-French listeners' perception of a synthetic stimulus array whose members systematically varied in the proximity between their first and second formants. As described in the Introduction, we chose Canadian-English and Canadian-French as languages based on certain aspects of their vowel systems. Specifically, the internal structure of their native /u/ categories was expected to differ in ways that would be informative for teasing apart the competing predictions of NRV and NLM concerning asymmetries in within-category vowel discrimination.

In Experiment 1, Canadian-English and Canadian-French listeners' categorization responses to a range of synthetic vowels revealed a perceptual sensitivity to sub-phonemic properties of native vowel production. Although all members of the stimulus array were consistently identified as /u/ by listeners in both language groups, the best French /u/ exemplars tended to be more focal compared to the best English /u/ exemplars. This outcome represents a perceptual counterpart to Escudero and Polka's (2003) finding that Canadian-French /u/ is produced with a greater degree of formant convergence than Canadian-English /u/ (see also MacLeod *et al*., 2009). In particular, it shows that English and French listeners appear to be especially attuned to the fine-grained acoustic structure of their native /u/ category.

Experiment 2 then examined the influences of focality and prototypicality on Canadian-English and Canadian-French listeners' discrimination of a subset of the /u/ variants (categorized in Experiment 1), which systematically differed in both their degree of formant proximity and stimulus goodness ratings, using an AX task. The results were clear in demonstrating that listeners from both language groups showed a directional asymmetry in which their discrimination performance was better when discriminating a change from a less-focal/English-prototypic /u/ to more-focal/French-prototypic /u/, compared to the reverse. However, there was no evidence that stimulus prototypicality modulated this asymmetry.

A comparable asymmetry was observed by Escudier and Schwartz (1989) when they examined French adults' discrimination of less versus more focal variants of French /e/. However, interpretation of their data was equivocal due to the absence of data to address the role of prototypicality effects. The current results provide the first direct evidence that directional asymmetries can be modulated by differences in formant convergence exclusively and independently of language-specific prototype categorization. In a recent follow-up study, we replicated the effect of formant proximity in English and French adults using natural speech tokens of French /u/ and English /u/ produced by a bilingual English-French speaker of Canadian English and French (Masapollo, Polka & Menard, 2016). Thus, synthetically controlled stimuli are not required to measure the vowel perception bias documented in the present study.

These findings collectively support one of the basic tenets of the NRV framework; namely, that the perception of vowel sounds is influenced by a universal sensitivity to extreme articulatory postures, which are acoustically specified via formant convergence patterns (Polka & Bohn, 2011; see also Schwartz *et al*., 2005). The focalization-based perceptual bias documented here also provides critical data in support of the Dispersion-Focalization theory of vowel systems (Schwartz *et al*., 1997, 2005).

Importantly, although the present findings clearly demonstrate asymmetric patterns that can be attributed exclusively to a universal vowel bias, these findings do not preclude effects of language-specific categorization on directional asymmetries. The NRV framework proposes that both universal and experience-dependent, language-specific biases play a role in shaping vowel perception, and therefore, that both may contribute to asymmetries. It is possible that directional asymmetries involving the discrimination of relatively large within- and between-category differences reveal universal biases that shape the global organization of the vowel space, while asymmetries involving the discrimination of relatively small within-category differences reveal language-specific biases that shape the local internal structure of native vowel categories. Although the findings from Experiment 2 fail to show a modulation of focalization effects by language-specific categorization experience, this may be a limitation of our stimulus array. Specifically, it is possible that prototypicality effects may emerge only during the discrimination of vowel exemplars that fall very close to a prototype stimulus in psychophysical space. In fact,

Kuhl’s (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. Our stimulus array was not designed to assess asymmetries in the narrow region immediately surrounding each native prototype. To accomplish this would require a stimuli set that is carefully constructed to define equivalent and more fine-grained perceptual gradients around the Canadian-English and Canadian-French /u/ prototypes.

The failure to observe effects of prototypically on asymmetries in Experiment 2 may be due, at least in part, to slightly weaker differences in category goodness across the less and more focal sets for the French subjects compared to their English peers, which are probably linked to the Canadian-French listeners' overall enhanced sensitivity to the /u/ variants in comparison to their English peers. The later unexpected, but robust finding may be due to differences in the vowel systems of English and French. Canadian-French has a richer inventory of high vowels (/i y u/) than English (/i u/).[14] Thus, this high region of the vowel space is more dense in French than in English, which may explain French listeners' enhanced sensitivity to spectral differences in this part of the vowel space.

An alternative, but not mutually exclusive, explanation for the Canadian-French listeners' enhanced sensitivity to the vowel differences in Experiment 2 concerns their exposure/experience with Canadian-English. Even though the French subjects were functionally monolingual, they had undoubtedly still experienced passive exposure to spoken English, and also heard English-accented French, throughout their lives in Quebec. In contrast, the English participants were university students from Ontario studying in Montreal, and therefore, probably had much less passive experience listening to French than the French participants had listening to English. We cannot exclude the possibility that this difference in linguistic experience contributed to the French listeners' perception of the /u/ variants.

To reduce experience-related differences in overall sensitivity to /u/ vowel differences it would be informative to test Canadian- and/or European-French adults with more limited exposure to English. Additionally, we could examine how Spanish listeners discriminate the /u/ stimuli implemented in the present study. Like English, Spanish only maintains a two-way high vowel contrast (/i-u/), but, like French, the Spanish /u/ is more peripheral (and therefore, more focal between F1 and F2) than English /u/. Accordingly, we would predict that both Spanish listeners and English listeners will show the same focalization-based asymmetry for the cross-set vowel pairs, and a prototype-based asymmetry in opposite directions for English and Spanish perceivers for the within set vowel pairs.

It is also important to recognize that in the broader research context the recent focus on vowel perception asymmetries is not new. Vowel perception asymmetries (referred to as order effects) reported in several earlier studies with adults led Cowan and Morse (1986) to propose the "neutralization hypothesis," which was tested further by Repp & Crowder (1990) and more recently by Karypidis (2007). According to this hypothesis, when listeners hold vowel sounds in memory, the representation of vowel quality drifts toward the neutral vowel [ə]. In this view, vowel order effects can be explained as a memory effect in which some vowel qualities are more subject to decay than others. Overall, this earlier work provided only weak support for the neutralization hypothesis. As well, the AX task implemented in this work often utilized a short ISI which favors auditory (trace) memory coding over phonetic (context coding), or did not use a bias-free measure of discrimination. These task differences make it difficult to interpret many of these earlier finding with respect to the focalization hypothesis.

The neutralization hypothesis was ultimately considered an inadequate account for vowel order effects by the researchers who proposed it. Nevertheless, their findings and others (e.g., Macmilllan, Goldberg, & Braida, 1988) clearly revealed that, along with memory factors, stimulus salience is also a critical factor driving directional asymmetries in vowel perception and other domains as well (e.g., Rosch. 1975). Given that asymmetries provide a way to tag perceptual salience, our research, and research motivated by NLM, has focused on identify stimulus properties that make some vowels more perceptually salient than others and allow them to serve as anchor or referent points in vowel perception when perceivers are engaged in a phonetic processing task.

According to the NRV framework, it is hypothesized that focalization biases reflect phonetic processes rather than general auditory processes. Thus, we do not expect to observe asymmetries to emerge in all test conditions. However, it is not entirely clear what task demands are needed to elicit a phonetic mode and/or restrict access to auditory memory traces. Several previous studies show that in an AX task, vowel order effects emerge or increase as ISI and stimulus predictability are increased whereas overall performance improves and asymmetries decrease when ISI and stimulus predictability are decreased (e.g., Werker & Logan, 1985; Cowan & Morse, 1986; Repp & Crowder, 1990; Polka & Bohn, 2011). For example, Polka and Bohn (2011) reported a directional asymmetry with a non-native vowel contrast when the ISI was 1500 ms, but not when the ISI was 500 ms. These task factors also likely interact with the magnitude of the stimulus difference presented. Given that overall performance levels were fairly high in Experiment 2, we would expect performance to be symmetric and at ceiling if the auditory memory access was enhanced by using a shorter ISI; however, different task effects may emerge for discrimination of more subtle vowel differences. Future research that can clarify the role of both memory aspects and stimulus salience in modulating directional asymmetries linked to focalization and categorization will advance our understanding of the mechanisms involved in vowel processing.

Finally, the present findings have important implications for theoretical models of developmental speech perception, which are centrally concerned with the language-specific reorganization of initial perceptual capacities and biases. While there is no doubt that linguistic experience has a profound and early impact on vowel perception (Kuhl *et al*., 1992; Polka & Werker, 1994; Polka & Bohn, 2011; Pons *et al*., 2012), our overall finding that directional asymmetries are driven by a universal sensitivity to formant proximity, independent of language-specific prototype categorization, demonstrates that there are constraints on the extent to which initial vowel biases can be modified through linguistic experience. Recent evidence indicates that this may also be the case in the perception of consonants. Specifically, Nam & Polka (2016) found that young infants from across different linguistic communities are initially biased toward stop consonants over fricative consonants, and that this bias continues to operate within adult language users (see also, Nam, 2015). At the present time, developmental theories, such as NLM (Kuhl *et al*., 2008), have focused almost exclusively on explicating how native-language experience alters speech perception, but such models will ultimately need to provide a more complete account of how universality and language-specificity fit together.

Acknowledgements

This research was supported by NSERC Discovery Grant 105397 to Linda Polka and NSERC Discovery Grant 312395 to Lucie Ménard. We are grateful to Anush Nersisyan for assistance with subject recruitment and data-collection, and Vincent Gracco for the generous use of facilities at the Center for Research on Brain, Language, and Music at McGill University. Special thanks to Lauren Franklin for her help with writing coding in R. This work benefited from helpful discussions with, or comments from, Susan Rvachew, Carol Fowler, Meghan Clayards, Shari Baum, Patricia Kuhl, and James Morgan. Experiment 1 was conducted as part of the third author's doctoral dissertation, and Experiment 2 was conducted as part of the first author's doctoral dissertation.

Endnotes

1. Portions of this work were presented in "Asymmetries in vowel perception: Effects of formant convergence and category goodness," Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, Scotland, UK, August 2015.

2. Also Centre for Research on Brain, Language, and Music, McGill University, 3640 de la Montagne, Montreal, Quebec, H3G 2A8, Canada.

3. Current address: Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer St., Providence, R.I., 02912, USA.

4. Author to whom correspondence should be addressed. Electronic mail: linda.polka@mcgill.ca

5. Current address: Basque Center on Cognition, Brain, and Language, Paseo Mikeletegi 69, 2nd Floor, 20009 Donostia-San Sebastian, Spain.

6. Specific vowels are referred to as "focal' vowels because they represent maximum formant convergence levels; however, focalization is not all or none, it is a graded effect that gives rise to salience differences across the vowel space. The effects of format convergence are also described in the Quantal Theory of Speech (Stevens, 1989); in this context, the corner vowels (/i/, /a/, /u/) are called "quantal" vowels.

7. Note that the English listeners in Experiments 1 and 2 were living in Montreal at the time of testing.

8. This analysis had less power due to one tie in the ranks.

9. These 7 subjects were recruited by an ad in a free newspaper whereas almost all other subjects in both experiments were university students. These subjects did not follow

directions and pressed buttons that they were not instructed to use. Their data was removed because we were not confident t that they understood or were fully cooperative in the task.

10. The same results were obtained for all analyses when percent-correct scores were used as the dependent variable.

11. The following formula was used: *A'* = .5 + (H-FA)(1+H-FA)/[4H(1-FA)], where H = proportion of hits and FA = proportion of false alarms (Grier, 1971). The false alarm rate was the combined error rate observed on same trials involving each vowel within the stimulus pair.

12. A robust direction effect was observed when separate ANOVAs were conducted on *A'* scores for each language group.

13. Reaction times (RTs) collected in Experiment 2 also provided converging results. In an ANOVA using RT as the dependent variable, we observed only a significant main effect of order of stimulus presentation [$F(1, 28) = 6.102$, $p = .020$, $\eta 2p = .179$] showing faster response for discriminating less to more focal vowel changes. There was no main effect of native language [$F(1, 28) = 1.646$, $p = .210$], or native language X order interaction [$F(1, 28) = .191$, $p = .665$].

14. In addition, each high vowel in Canadian French has a lax allophonic variant, [ɪ], [ʊ] and [ʏ]. In Canadian English, the tense/lax pairs /i-ɪ/, and /u-ʊ/ occur, but not /y-ʏ/. Further, cross-language vowel production data indicates that the tense/lax high vowel allophones that occur in Canadian French are spectrally closer than the comparable tense/lax vowels that occur Canadian English (Escudero & Polka, 2003).

References

Boë, L.-J. (1999). Modelling the growth of the vocal tract vowel spaces of newly-born infants and adults: consequences for ontogenesis and phylogenesis, *Proc. of the Int. Congress of Phonetic Sciences*, 2501-2504.

Boersma, Paul & Weenink, David (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.19, http://www.praat.org/

Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P. (1986). Cinéradiographie des Voyelles et Consonnes du Français [Cineradiographic study of French vowels and consonants], Institut de Phonétique de Strasbourg, Strasbourg, France.

Cowan, N., & Morse, P. (1986). The use of auditory and phonetic memory in vowel discrimination. *J. Acoust. Soc. Am.*, 79, 500-507.

Cutler, A. (2012). Native listening: Language experiences and the recognition of spoken words. MIT Press, Cambridge, MA.

Diesch, E., Iverson, P., Kettermann, A., & Siebert, C. (1999). Measuring the perceptual magnet effect in the perception of /i/ by German listeners. *Psych. Res.*, 62, 1–19.

Dufour, S., Brunelliere, 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. *J. of Psycholing. Res.*, 42, 161-173.

Escudero, P., & Polka, L. (2003). A cross-language study of vowel categorization and vowel acoustics. In Sole, M.J., D. Recansens and J. Romero (Eds.), *Proc. of the Int. Congress of Phonetic Sciences*, Barcelona, 861-864.

Feldman, N., Griffiths, T.L., & Morgan, J. (2009). The influences of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference. *Psych. Rev.,* 116(4), 752-782.

Frieda, E. M., Walley, A. C., Flege, E. J., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Percept. Psychophys.*, 61(3), 561-577.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psych. Bull.*, 75, 424-429.

Iverson, P., & Kuhl, P.K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoust. Soc. Am.*, 97, 553–562.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Percept. Psychophys.*, 62, 874-886.

Johnson, K., E. Flemming, and R. Wright (1993). The hyperspace effect: Phonetic targets are hyperarticulated, *Language* 69, 505-528.

Karypidis, C. (2007). Order effects and vowel decay in short-term memory: The neutralization hypothesis. *Proc. of the Int. Congress of Phonetic Sciences, Saarbrücken,* 657–660.

Kent, R.D., & Read, C. (2002). The acoustic analysis of speech. Singular/Thomson Learning.

Kuhl, P. K. (1991). Human adults and human infants show a “Perceptual Magnet Effect” for the prototypes of speech categories: Monkeys do not. *Percept. Psychophys.*, 50, 93–107.

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. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. deBoysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259-274). Dordrecht, Netherlands: Kluwer Academic Publishers.

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). *Phil. Trans. of the Royal Society B,* 363, 979-1000.

Lively, S.E. & Pisoni, D.B. (1997). On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *J. Exp. Psych. and Human Percept. Perf.*, 23(6), 1665-1679.

Lotto, A.J., Kluender, K.R., & Holt, L.L. (1998). Depolarizing the perceptual magnet effect. *J. Acoust. Soc. Am.*, 103(6), 3648-3655.

MacLeod, A., Stoel-Gammon, C., & Wassink, A.B. (2009). Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers. *J. Phonetics*, 37, 4, 374-387.

MacMillan, N.A., Goldberg, R.F., & Braida, L.D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *J. Acoust. Soc. Am.*, 84, 4, 1262-1280.

Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A users guide (2nd ed.). New York: Cambridge University Press.

Maeda, S. (1979). An articulatory model of the tongue based on a statistical analysis. *J. Acoust. Soc. Am.*, 65, S22. DOI: 10.1177/0956797614533571.

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model. In W.L. Hardcastle & A. Marshal (Eds.), *Speech Production and Speech Modeling* (131-149). Dodrecht, The Netherlands: Kluwer Academic.

Marian, V., Blumfeld, H.K., & Kaushanskaya, M. (2007). Language Experience and Proficiency Questionnaire (LEAP-Q). *J. Speech, Lang. and Hearing Res.*, 50(4), 940-967.

Martin, P. (2002). Vowel system of Quebec French. From acoustics to phonology. *La Linguistique*, 38, 71–88.

Masapollo, M., Polka, L., & Ménard, L. (2016). The role of visual-phonetic information from lip movements on the natural referent vowel bias. Poster presented at the *15th Meeting of the Assoc. for Laboratory Phonology*, Cornell University, Ithaca, NY.

Ménard, L., Schwartz, J-L. & Boë, 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. *J. Speech, Lang, and Hearing Research*, 47, 1059- 1080.

Ménard, L., Davis, B., Boë, L.-J, & Roy, J.-P. (2009). Producing American-English During Vocal Tract Growth: A Perceptual Categorization Study of Sunthesized Vowels. *Journal of Speech, Language, and Hearing Research*, 52, 1268-1285.

Nam, Y.-J. (2015). The role of acoustic-phonetic bias in consonant manner perception. Ph.D. Thesis, McGill University.

Nam, Y.-J., & Polka, L. (2016). The phonetic landscape in infant consonant perception is an uneven terrain. *Cognition*, 155, 57-66.

Noiray, A. Cathiard, M.-A., Ménard, L., & Abry, C. (2011). Test of the Movement Expansion Model: Anticipatory vowel lip protrusion and constriction in French and English speakers. *J. Acoust. Soc. Am.*, 129 (1), 340-349.

Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels. *J. Acoust. Soc. Am.*, 32, 693-703.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Percept. Psychophys.*, 52(1), 37-52.

Polka, L., & Werker, J.F. (1994). Developmental changes in perception of non-native vowel contrasts. *J. Exp. Psych. Human Percept. Perf.*, 20(2), 421-35.

Polka, L. & Bohn, O. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants, *J. Acoust. Soc. Am.*, 110, 577-592.

Polka, L. & Bohn, O. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, 221-231.

Polka, L., & Bohn, O. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *J. Phonetics*, 39, 467-478.

Pons, F, Albareda-Castellot, B., & Sebastian-Gallés. (2012). The interplay between input and initial biases: asymmetries in vowel perception during the first year of life. *Child Development*, 83(3), 965-976.

Repp, B.H., Healy, A.F., & Crowder, R.G., 1979. Categories and context in the perception of isolated steady-state vowels. *J. Exp. Psychol.: Human Percept. Perform.* 5, 129–145.

Repp, B.H., & Crowder, R.G. (1990). Stimulus order effects in vowel discrimination. *J. Acoust. Soc. Am.*, 88, 2080–2090.

Rosch, E. (1975). Cognitive reference points. *Cognition,* 7, 532– 547.

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.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997). The Dispersion-Focalization Theory of vowel systems. *J. Phonetics*, 25, 255–286.

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.

Stevens, K. N. (1989). On the quantal nature of speech. *J. Phonetics*, 17, 3–46. Tyler, M.D., Best, C.T., Faber, A. & Levitt, A.G. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica*, 71, 4-21.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Dev*, 7, 49-63.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percept. Psychophys.,* 37, 35–44.

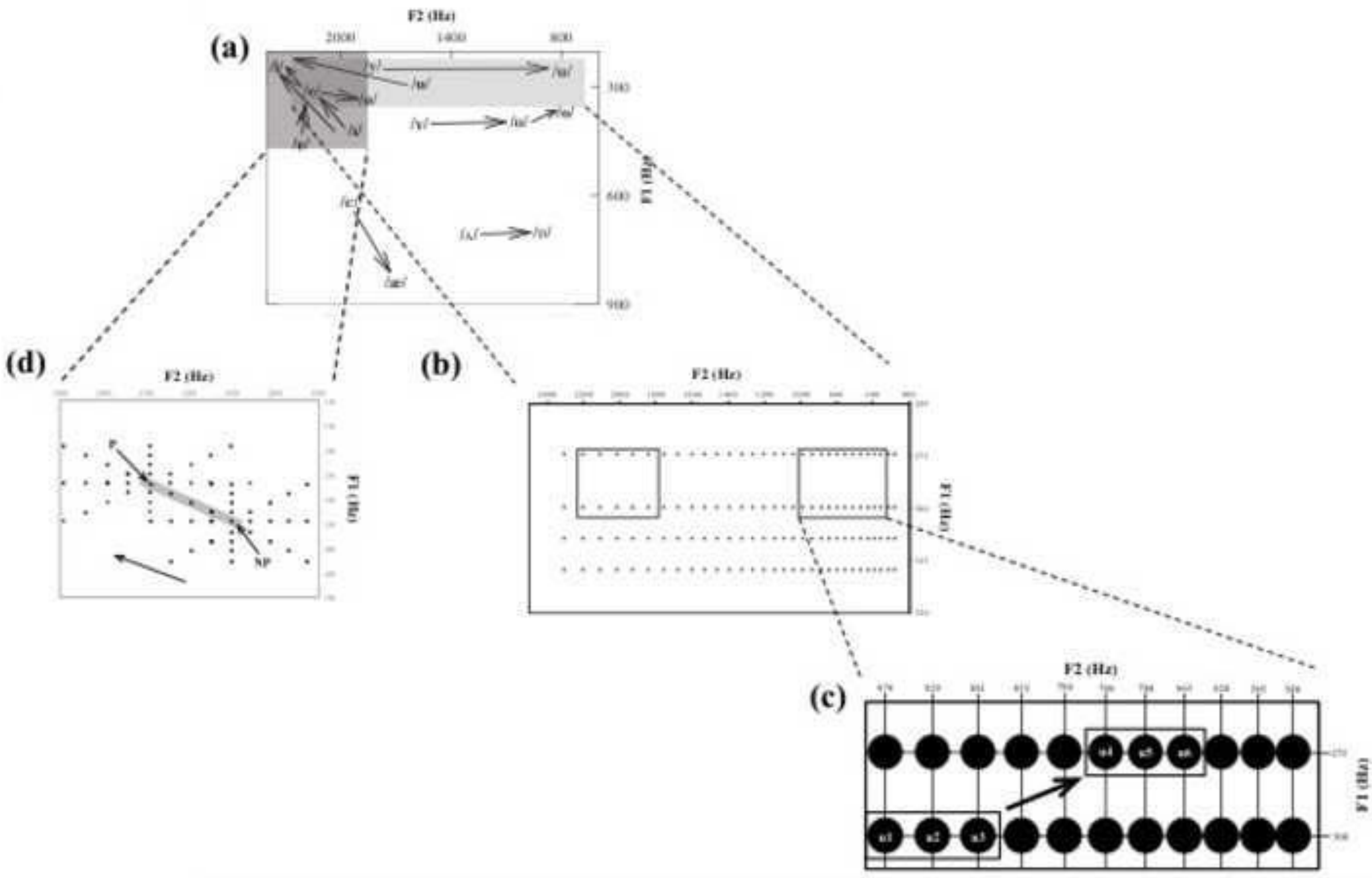
Whalen, D. H., Magen, H. S., Pouplier, M., Kang, A. M., & Iskarous, K. (2004). Vowel production and perception: Hyperarticulation without a hyperspace effect. *Language and Speech 47*, 155-174.

Zwicker, E., & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *J. Acoust. Soc. Am.*, 68, 1523-1525.

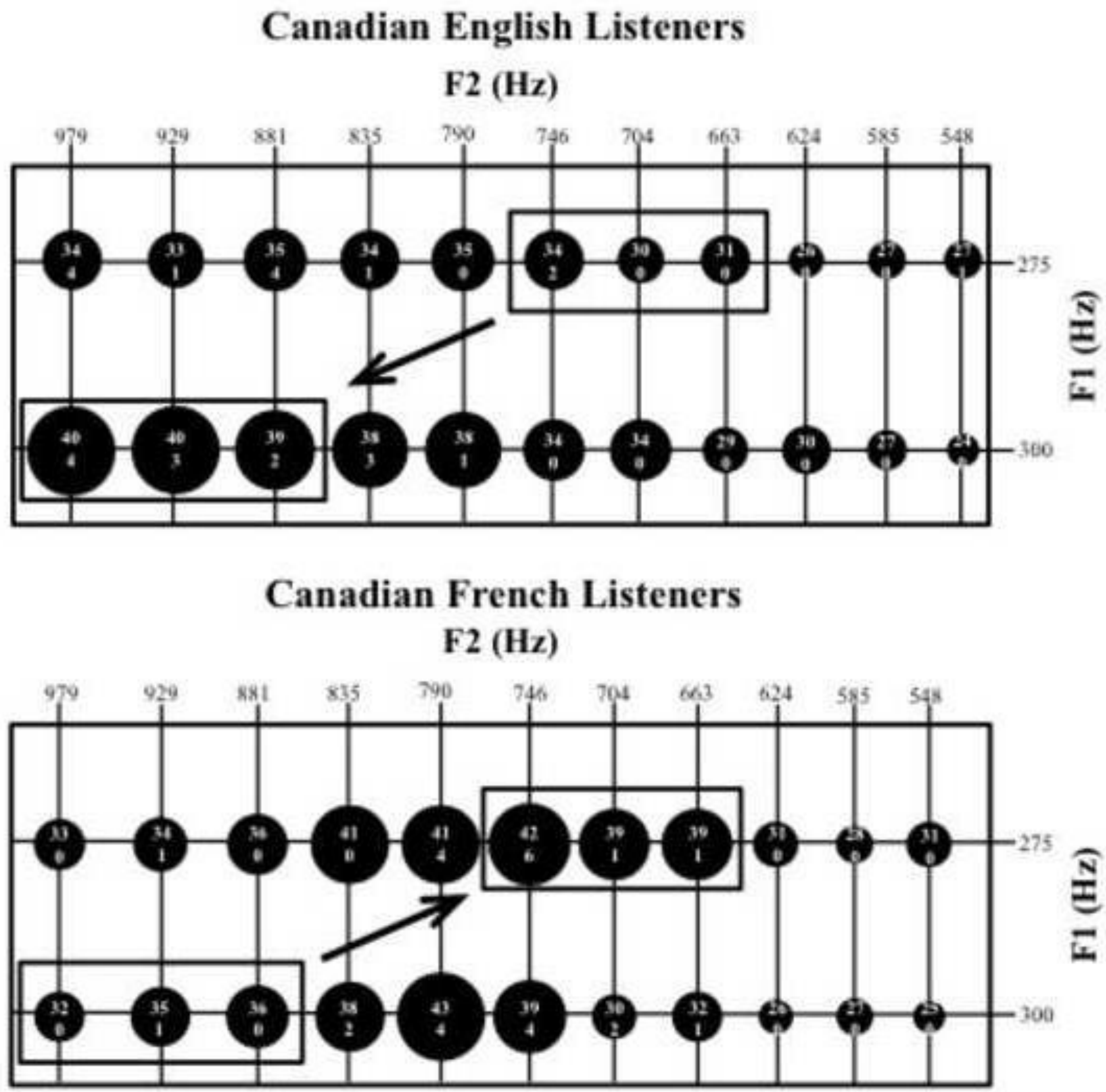
Stimulus Pair	Order of Vowel Change		Statistical p Values (α = .05)
Stimulus Pair	Less to more focal	More to less focal	*p*
u_1/u_4	0.97	0.95	0.151
u_1/u_5	0.97	0.96	0.098
u_1/u_6	0.97	0.96	0.095
u_2/u_4	0.97	0.95	0.051
u_2/u_5	0.97	0.95	0.187
u_2/u_6	0.97	0.96	0.114
u_3/u_4	0.97	0.94	0.004
u_3/u_5	0.96	0.94	0.053
u_3/u_6	0.97	0.91	0.000

TABLE I. Mean *A'* values by stimulus pair and presentation order for the cross-set discrimination analyses (statistical values are listed on the far right).

Figures (TIFF, EPS, PS, or JPEG only) Click here to download Figures (TIFF, EPS, PS, or JPEG only) Figure_1.jpg



Figures (TIFF, EPS, PS, or JPEG only) Click here to download Figures (TIFF, EPS, PS, or JPEG only) Figure_2.jpg



Figures (TIFF, EPS, PS, or JPEG only) Click here to download Figures (TIFF, EPS, PS, or JPEG only) Figure_3.jpg