Asymmetries in Vowel Perception: Effects of Formant Convergence and Category "Goodness"

Matthew Masapollo[1,2], Linda Polka[1,2], and Lucie Ménard[2,3]

[1] School of Communication Sciences and Disorders, McGill University
[2] Centre for Research on Brain, Language, and Music, McGill University
[3] Département de Linguistique, Université du Québec à Montréal

ABSTRACT

The mechanisms underlying directional asymmetries in vowel perception are a focus of debate. One account – the Natural Referent Vowel (NRV) framework – suggests that asymmetries reflect a *language-universal* phonetic bias, such that listeners are predisposed to attend to vowels with extreme articulatory postures, which display high formant convergence. A second account – the Native Language Magnet (NLM) theory – suggests that asymmetries reflect a *language-specific* bias favoring "good" exemplars of native vowel categories. We examined whether listeners display asymmetries influenced by formant proximity and/or language experience. Specifically, we tested English adults in a same-different discrimination task, using /u/ vowels that systematically differed in their degree of formant proximity (between F1 and F2) and stimulus goodness. Results revealed asymmetries as predicted by NRV when vowel pairs exhibited a relatively larger difference in their F1-F2 convergence patterns, and as predicted by NLM when vowel pairs exhibited a relatively smaller difference in their F1-F2 convergence patterns.

Keywords: Natural Referent Vowel framework; Native Language Magnet theory; vowel perception

1. INTRODUCTION

Considerable research indicates that directional asymmetries often occur in infant vowel perception, such that discrimination of a vowel change presented in one direction is consistently easier compared to the same change presented in the reverse direction (for reviews, see [1, 2]). In particular, vowel discrimination is easier for infants when they are presented with a change from a less peripheral to a more peripheral vowel within articulatory-acoustic vowel space (defined by F1-F2). For example, young infants perform better at detecting the change from /e/ to /i/ than the reverse change from /i/ to /e/. Infants from across cultures exhibit these perceptual asymmetries in their discrimination of both native and non-native vowel contrasts, indicating that they are *not* the result of experience listening to a specific language. Furthermore, infants also prefer listening to more peripheral vowels (e.g., /i/) over less peripheral vowels (e.g., /ɪ/), regardless of their phonemic status in the ambient language [2], suggesting that discrimination asymmetries reflect a language-universal perceptual bias for vowels with extreme articulatory-acoustic properties. By the end of the first year of life, however, these asymmetries begin to fade for native vowel contrasts, but remain in place for non-native vowel contrasts [2, 3]. In adulthood, directional asymmetries are reduced for native contrasts and maintained or enhanced for non-native vowel contrasts ([2], cf. [4]).

On the basis of these findings, Polka and Bohn [2] advance the Natural Referent Vowel (NRV) framework, which argues that infants initially display a *language-universal* bias favoring peripheral vowels, which serve as natural reference points (or perceptual anchors) within the vowel space, but that this bias gradually shifts as needed to optimize vowel processing in a specific language. Concurring with Schwartz and colleagues [5], Polka and Bohn further argue that vowels found at the periphery of articulatory-acoustic vowel space, such as /i/, /a/ and /u/, exhibit spectral properties that make them acoustically and perceptually salient, and that listeners thus treat these vowels as perceptual reference points during vowel discrimination tasks. Specifically, peripheral vowels are conjectured to be more salient due to local formant convergence, or *focalization*. As two neighbouring formants move close together in frequency, there is a mutual reinforcement of their acoustic energy, such that the amplitude of each formant increases. Vowel spectra with marked peaks are hypothesized to be easier for listeners to perceive and memorize [5]. Maximal focalization is observed for vowels found at the periphery of the vowel space, which have the most extreme vocal tract postures. For example, F2 and F3 are close in frequency for /i/ (which is the highest front vowel), and F1 and F2 are close in frequency for /a/ (which is the lowest back vowel) and /u/ (which is the highest back vowel). Focalization, then, could be a central phonetic cue underlying asymmetries.

Directional asymmetries in infant and adult

vowel perception are not limited to *across-category* vowel pairs. Studies by Kuhl and colleagues report that English listeners' discrimination from a prototypical to a non-prototypical exemplar of the *same native* vowel category (English /i/) is harder compared to the same change presented in the reverse direction [6, 7, 8, see also 9], while no asymmetry was observed around a non-native vowel prototype (Swedish /y/); these effects were reversed for Swedish listeners [7]. These findings led Kuhl [10] to formulate the Native Language Magnet (NLM) theory, which argues that perceptual learning, using the statistical properties of the input speech, biases listeners toward native language phonetic category prototypes. Here, the best exemplars of a native vowel category are said to "pull" similar auditory representations towards itself much as a magnet attracts iron.

However, a serious challenge to NLM comes from other studies showing that some listeners perceive the non-prototypical vowels to be outside of the category of the prototypical vowels. For example, listeners frequently identified Kuhl's [6] non-prototypical /i/ exemplars as /e/ [8, 11, 12]. This raises the possibility that the so-called "perceptual magnet effect" may be explained instead as an NRV bias – a change from /e/ to /i/ is easier to detect than a change from /i/ to /e/ [2]. Moreover, the prototypical /i/ vowel employed by Kuhl [6] was more focal (between F2 and F3) compared to the non-prototypical /i/. Thus, it is not clear whether within-category vowel discrimination asymmetries are best accounted for by focalization, language experience, or both factors.

In this investigation, the effects of formant convergence and category "goodness" on vowel discrimination were examined. Specifically, we tested Canadian English listeners' discrimination of vowel pairs contrasting a less focal with a more focal exemplar from within the same native vowel category, using variants of /u/. The /u/ variants systematically differed in both their degree of formant proximity (between F1 and F2) and perceptual goodness judgments. Crucial to this study's design is that the prototypical English /u/ stimuli are *less* focal compared to the non-prototypical /u/ stimuli. This being the case, NRV and NLM make opposite predictions with respect to directional asymmetries. NRV predicts that English listeners will be more accurate at discriminating a vowel change in the less to more focal direction, whereas NLM makes the *opposite* prediction that English listeners will be more accurate in discriminating a vowel change in the more to less focal direction.

2. METHODS

2.1. Participants

Fifteen normal-hearing, monolingual Canadian English speaker-listeners from Ontario participated in this experiment (mean age = 25, SD = 11, 6 males and 9 females). Their language background was assessed using the Canadian English version of the Language Experience and Proficiency Questionnaire [13]. All participants were raised in monolingual English homes, did not begin learning a second language earlier than 10 years of age, and did not use a second language on a regular basis.

2.2. Stimuli

The stimuli were 22 /u/ tokens from a previous study by Molnar [14]. The vowels were synthesized, using the Variable Linear Articulatory Model [15], to simulate productions by an adult male speaker. The /u/ tokens varied in F1 (275 or 300 Hz) and F2 (548-979 Hz, in equal mel steps). The stimuli were identical in all other respects: f0, F3, F4, and F5 were 120 Hz, 2522 Hz, 3410 Hz, and 4159 Hz, respectively. The stimuli were 400-ms in duration, and had the same intonation contour. The stimulus array is shown in Figure 1.

Molnar [14] presented these stimuli to native Canadian English speakers-listeners (n=15) from the Ontario region. Note that these were *not* the same subjects who participated in the present experiment. Each stimulus was identified and rated for "goodness" on a 5-point scale (1=very poor, 5=very good). Listeners identified all of the vowels as /u/ (the mean percent of /u/ identification was above 85% for all of the tokens). In Figure 1, relative differences in /u/ category goodness are displayed as differences in the size of each token; the average "goodness" rating (from Molnar [14]) is also indicated in the center of each stimulus.

Six of the vowel stimuli were selected from the full array to be used in the present experiment. To assess effects of focalization and category goodness on vowel discrimination, two vowel sets were selected: one set included less focal/high /u/ category goodness tokens, and the other included more focal/low /u/ category goodness tokens. The selected stimuli are marked and labelled in Figure 1. Stimuli 1, 2, and 3 form the less focal set with an average F1-F2-distance = 629 Hz (723 mels); this set received the highest perceptual goodness ratings (M=4.0). Stimuli 4, 5, and 6 form the more focal set with an average F1-F2-distance = 402 (511 mels); this set received lower goodness ratings (M=3.2).

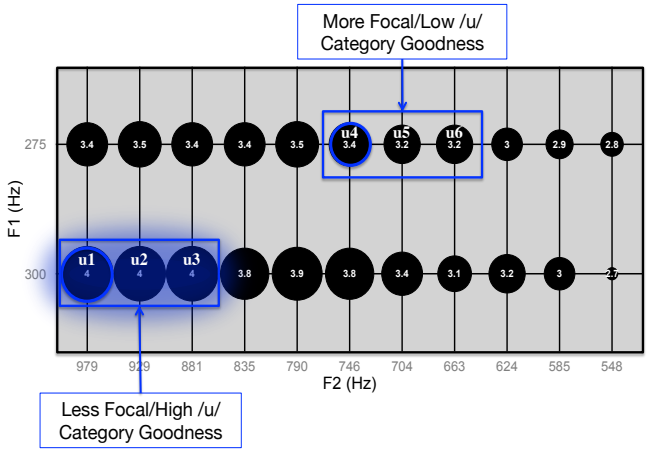

Figure 1: The formant frequencies for the entire array of /u/ vowel stimuli. The six stimuli used in the present study are marked by the blue squares, and labelled 1 thru 6. The relative differences in /u/ category goodness are displayed as differences in the size of each token (larger for higher goodness). The numbers within the tokens list the average goodness ratings. The blue highlighted region indicates the location of the prototypical English /u/ vowels in articulatory/acoustic space.

2.3. Apparatus

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

2.4. Procedure

We employed an AX same/different discrimination task. On each trial, participants heard two stimuli, separated by an inter-stimulus interval of 1500-ms, and then judged whether they were the "same" or "different." Participants initiated a trial by pressing a response key, and then pressed one of two labelled buttons to indicate whether the second stimulus was the same as the first [A] or different from the first [X]. Before the experimental session, 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. Participants completed 180 trials in a session. They heard every possible pair of the 6 stimuli (including each stimulus being paired with itself), 5 times, in both presentation orders. No feedback was provided.

3. RESULTS

The critical research question was whether focalization, native language experience, or both factors, influence directional asymmetries in adult vowel perception. If formant convergence influences vowel perception, then discrimination will be heightened when stimulus pairs change in the direction going from the less to more focal /u/ tokens. Conversely, if category goodness influences perception, then discrimination will be heightened when pairs change in the direction going from the more to less focal /u/ tokens. It is also possible that both effects will be observed with asymmetries consistent with NRV evident for larger acoustic differences and asymmetries consistent with NLM evident for smaller acoustic differences.

The first analysis focused on discrimination of cross-set vowel pairs. The mean percent correct responses were calculated for each participant for all different vowel pairs contrasting a vowel from the less focal set (stimuli 1, 2, 3) with a vowel from the more focal set (stimuli 4, 5, and 6); separate scores were computed for each order of presentation (less to more focal; more to less focal). Mean percent correct responses were submitted to a two-way analysis of variance (ANOVA) with stimulus pair (1/4, 1/5, 1/6, 2/4, 2/5, 2/6, 3/4, 3/5, 3/6) and order of stimulus presentation (less to more focal vs. more to less focal) as within-subjects factors. There was a main effect of order of presentation, $F(1,112) = 11.704$, $p=.004$, $\eta_p^2 = .455$, such that listeners were better at detecting less to more focal vowel changes ($M=95.2$, $SE=1.2$), compared to more to less focal vowel changes ($M=84.0$, $SE=2.7$) (shown in Figure 2). There was also a main effect of stimulus pair, $F(1,112)=2.795$, $p=.007$, $\eta_p^2 = .166$, but no two-way interaction, $F(8,112)=1.067$, $p=.391$. This analysis was repeated using A' scores (using different and same pair scores) to ensure that the observed effects were not due to response-bias. A' is a response bias-free measure of perceptual sensitivity that represents each participant's hit rate as a function of their false-alarm rate [17, 18]. Here, only a main effect of order of stimulus presentation was found.

Taken together, the results of this analysis strongly support the NRV hypothesis [2, 5] that focalization affects adult listeners' perception of vowel differences. However, these data alone cannot be used to argue against the NLM hypothesis [10]. It is possible that asymmetries consistent with NLM will occur among vowel pairs that are closer to the prototype stimulus in acoustic space. Kuhl [6] reported large NLM effects for vowels close to the prototype stimulus, and very small effects for vowels further from the prototype.

To assess this possibility, a second analysis was conducted to examine discrimination of within-set vowel pairs. To do so, the mean percent correct

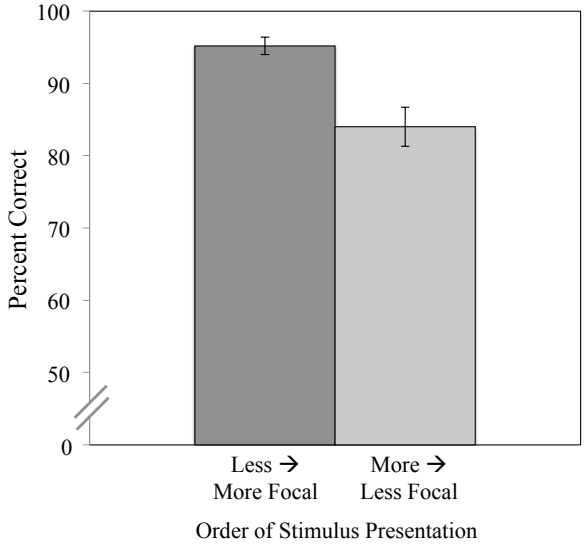

Figure 2: Directional asymmetries in cross-set vowel discrimination consistent with NRV. Mean percent correct responses for all different vowel pairs contrasting a vowel from the less focal set (stimuli 1, 2, 3) with a vowel from the more focal set (stimuli 4, 5, 6) Error bars represent standard errors.

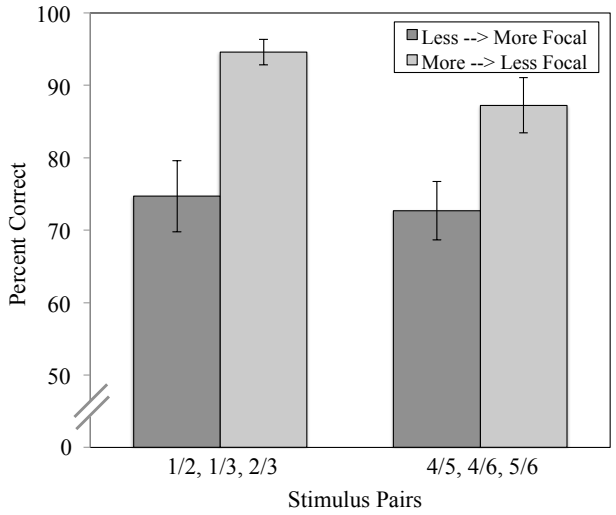

Figure 3: Directional asymmetries in within-set vowel discrimination consistent with NLM. Mean percent correct responses for all different vowel pairs contrasting stimuli within the less focal set (stimuli 1, 2, 3) and within the more focal set (stimuli 4, 5, 6). Error bars represent standard errors.

responses were calculated for each participant, for all different vowels within the less focal set (stimuli 1, 2, 3) and for all vowel pairs within the more focal set (stimuli 4, 5, 6); separate scores were computed for each order of presentation (less to more focal; more to less focal). Mean percent correct responses were submitted to a two-way ANOVA with stimulus pair (1/2, 1/3, 2/3, 4/5, 4/6, 5/6) and order of stimulus presentation (less to more focal vs. more to less focal) as within-subjects factors. Consistent with NLM predictions, there was a main effect of order of presentation,, $F(1,70)=10.081$, $p=.007$, $\eta_p^2 =.419$, showing that listeners were better at detecting vowel changes in the more to less focal direction (towards the English /u/ prototype) (for 1/2, 1/3, 2/3: $M=94.6$, $SE=1.8$; for 4/5, 4/6, 5/6: $M=87.2$, $SE=3.8$), compared to the reverse direction (i.e., less to more focal and towards the non-prototypic English /u/) (for 1/2, 1/3, 2/3: $M=74.7$, $SE=4.9$; for 4/5, 4/6, 5/6: $M=72.7$, $SD=4.0$) (shown in Figure 3). There was a marginal effect of stimulus pair, $F(5, 70)=2.339$, $p=.051$, $\eta_p^2 =.143$, but no two-way interaction, $F(5,70)=.979$, $p=.437$. This analysis was also repeated using *A'* scores; only a main effect of order of stimulus presentation was found. Although this analysis failed to show a significant interaction, there is a trend for the asymmetries to be stronger for vowel pairs closer to the prototype (i.e., stimulus 1).

4. DISCUSSION

The present findings demonstrate that asymmetries in vowel perception pattern as predicted by NRV when vowel pairs exhibit a relatively larger difference in their F1-F2 convergence patterns, and as predicted by NLM when vowel pairs exhibit a relatively smaller difference in their F1-F2 convergence patterns. This suggests that both *language-universal* and experience-dependent *language-specific* phonetic biases function to shape vowel perception. These findings suggest that the prevailing effect is modulated by degree of acoustic distance between the contrasting vowels such that NLM effects are evident for stimuli close to the prototype and NRV effects arise when focalization differences are larger but still fall within the same vowel.

To further investigate how language-independent and language-specific aspects of phonetic perception interact, we are examining Canadian French listeners' ability to discriminate the present vowel contrasts. Canadian French provides an interesting test bed for our hypotheses. Molnar [14] reported that native Canadian French listeners rate stimuli 4, 5 and 6 (the more focal variants) as the prototypical exemplars of their /u/ category. This being the case, we expect them to show parallel asymmetries (better in the less to more focal direction) for both cross-set and within-set vowel discrimination. Future cross-language studies testing English and French-learning infants' discrimination abilities for these vowel contrasts will also be informative for theory.

5. REFERENCES

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

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

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

[4] Tyler, M., Best, C., Faber, A., & Levitt, A. (2014). Perceptual assimilation and discrimination of non-native vowel contrasts, *Phonetica*, *71*, 4-21.

[5] 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.

[6] 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.

[7] 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.

[8] Iverson, P., & Kuhl, P.K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97, 553–562.

[9] Hawkins, S. and Barrett Jones, S. (2004) The perceptual magnet effect reflects phonetic context. *Journal of the Acoustical Society of America* 115(5, Part 2), 2630.

[10] 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.

[11] Lively, S.E. & Pisoni, D.B. On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology and Human Perceptual Performance*, 23(6), 1665-1679.

[12] Lotto, A.J., Kluender, K.R., & Holt, L.L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America*, 103(6), 3648-3655.

[13] Marian, V., Blumfeld, H.K., & Kaushanskaya, M. (2007). Language Experience and Proficiency Questionnaire (LEAP-Q). *Journal of Speech, Language and Hearing Research*, 50(4), 940-967.

[14] Molnar, M. (2011). Behavioral and electrophysiological correlates of vowel perception in monolingual and simultaneous bilingual users of Canadian English and Canadian French. Ph.D. Thesis, McGill University.

[15] 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. *Journal of Speech, Language, and Hearing Research*, 47, 1059- 1080.

[16] Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, 8, 185–190.

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

[18] Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424-429.