Is focalization the keystone connecting vowel perception and production?

Linda Polka[1], Matthew Masapollo[2], Lucie Ménard[3], Lucile Rapin[3] and Jean-Luc Schwartz[4]

[1]McGill University, Canada [2]University of Florida, USA [3]UQAM, Canada [4]GIPSA-lab Grenoble, France
linda.polka@mcgill.ca; mmasapollo@phhp.ufl.edu; menard.lucie@uqam.ca; lucilerapin@gmail.com;
jean-luc.schwartz@gipsa-lab.gernoble-inp.fr

ABSTRACT

Vowel perception is often asymmetric such that discriminating the same vowel contrast is easier in one direction compared to the reverse direction. According to one account, asymmetries reveal a perceptual bias favoring more extreme vowel gestures which give rise to salient acoustic signals due to increased formant convergence (*focalization*). Consistent with this articulatory account are findings that analogous asymmetries are elicited with silent lip-read videos of vowels, as well as other disembodied, animated displays that simulate the kinematic and configural properties of articulatory mouth gestures. However, our understanding of the relations between vowel constrictions, perceptual asymmetries, and formant patterns is incomplete. Here, focusing on the French /e/-/ø/ contrast, we examine whether formant proximities are related to perceptual asymmetries, and whether there are firm links between formant convergence and tongue constriction degree and lip aperture. We introduce the notion that focalization may serve as the *keystone* connecting vowel perception and production.

Keywords: vowels, perception, production, focalization, electromagnetic articulography

1. INTRODUCTION

Considerable research over the past 30 years has revealed that vowel perception by human adults and infants is often asymmetric: discriminating the same vowel contrast is easier in one direction compared to the same change presented in the reverse direction (for reviews and meta-analytic findings, see [1, 2]). This perceptual pattern is schematized in Figure 1, which shows a standard *F1-F2* plot of vowel contrasts that have been examined in infant vowel discrimination studies with arrows indicating the direction of change that was easier to discriminate.

According to the Natural Referent Vowel (NRV) framework [1], these asymmetries reveal a universal perceptual bias toward acoustic signals produced with more extreme vowel gestures, which act as *natural referent* vowels. Specifically, it is argued that vowels generated by more extreme gestures display heightened acoustic and perceptual salience due to well-defined spectral prominences formed by the convergence of adjacent formants (*focalization*; see also [3]). As Figure 1 illustrates, directional asymmetries can frequently be predicted from *F1-F2* vowel space location because, at least for oral vowels, focalization increases and reaches maximum levels at the corners of the vowel space. Critically, however, focalization and peripherality are distinct phonetic properties. Focalization considers the first three to four formants and their spectral position relative to each other, not just the vowel position within a simple *F1-F2* space.

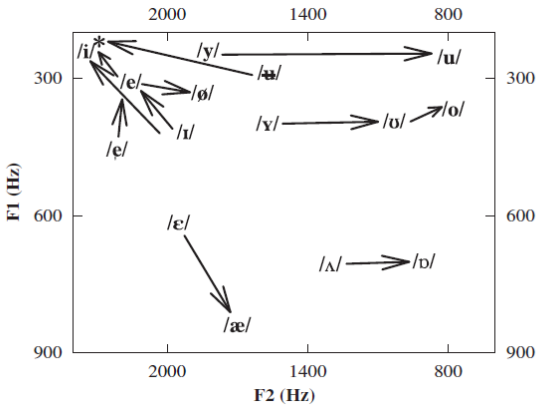

Figure 1: Schematic illustration of phonetic space based on the first two formant frequencies. Vowel contrasts showing asymmetries in infant vowel perception are plotted; arrows point in the direction that is easier to discriminate (adapted from Polka & Bohn, 2011).

The /e/-/ø/ vowel contrast is an example where perceptual asymmetries predicted by peripherality versus focality do not align, and thus, provides an ideal opportunity to test the NRV claim that articulatory gestures – acoustically specified by formant convergence – underlie perceptual asymmetries. Prior studies have examined asymmetries for the /e/-/ø/ vowel contrast in several languages including German [4], Dutch [5], and Danish [1]. In each study, the test stimuli were

multiple tokens produced by one talker; formant values (averaged across tokens) are shown below in Figure 2. In all three studies, *F2* frequencies were higher for /e/ than /ø/. *F1* frequencies were similar for both vowels in the Danish and German productions and were lower for /e/ compared to /ø/ in Dutch. As shown in Figure 2, /e/ is more peripheral than /ø/ in *F1-F2* space in the German and Danish productions, but not in the Dutch productions. When these formant data were recently examined from the perspective of focalization [6], a consistent pattern emerged. As shown in Table 1 below, for each talker *F1* and *F2* frequencies were much closer in /ø/ compared to /e/; *F1* and *F3* distances showed the same pattern. These differences were consistent and robust; in each language there was no overlap in the magnitude of these formant distances between /e/ and /ø/. Some differences in *F2* and *F3* convergence were also noted, but in each language, they were much smaller, overlapped in magnitude across vowels, and were inconsistent in direction (e.g., more *F2-F3* convergence for /ø/ in Danish; for /e/ in Dutch).

These acoustic patterns are intriguing considering perceptual findings (see Table 2). In the Danish and the Dutch study, an asymmetry was observed showing better performance in the /e/ to /ø/ direction; this pattern was not predicted based on peripherality in either Danish or Dutch but aligns well with predictions based on focality in both languages. No asymmetry was observed in the German study; this may be because the acoustic differences between vowels were larger in this study. Overall, these findings bolster NRV by suggesting that asymmetries are associated with differences in focalization, whereas a peripherality-based prediction fails.

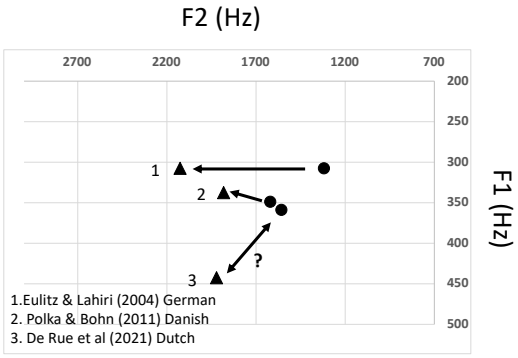

Figure 2: *F1-F2* plot of vowel stimuli from three studies that assessed perceptual asymmetries for /e/-/ø/. Mean values (across multiple tokens) are plotted for /e/ (triangles) and /ø/ (circles). An arrow connects the vowel pair to show the peripherality prediction asymmetry.

The NRV framework further posits that the processes underlying vowel perception asymmetries are sensitive to articulatory information, not focal acoustic patterns *per se* [7]. Especially consistent with this account are recent findings that asymmetries, comparable to those observed with acoustic vowel stimuli, are elicited with silent lip-read videos of vowels [7,8], as well as other disembodied, animated displays that simulate the kinematic and configural properties of orofacial speech movements [9]. Thus, in this view, focalization serves as the *keystone* (or lynchpin) connecting perceptual and articulatory patterns: perceivers directly extract information about distal vowel constrictions from proximal formant patterns. However, this account does not provide firm evidence that vowel constrictions are or are not recoverable from formant parameters, as a comprehensive understanding of the complex relations between vowel constrictions, formant patterns, and perceptual asymmetries is still being built [3,7].

		Danish (n=1) Polka & Bohn (2011)	German (n=1) Eulitz & Lahiri (2004)	Dutch (n=1) de Rue et al (2021)
F1-F2 (Hz)	e	1541	1817	1474
	Ø	**1271**	**1009**	**1198**
	difference	270	808	276
F1-F3 (Hz)	e	2239	2442	1980
	Ø	**1935**	**1709**	**1813**
	difference	304	733	167
F2-F3 (Hz)	e	698	**625**	**506**
	Ø	**664**	700	615
	difference	34*	75	109*

* values overlap

Table 1: Formant convergence analysis of vowel stimuli from three studies that assessed perceptual asymmetries for /e/-/ø/. Shaded values indicate the more focal vowel within the pair.

	Polka & Bohn (2011)	Eulitz & Lahiri (2004)	de Rue et al (2021)
Language	Danish	German	Dutch
Peripherality prediction (Figure 2)	/Ø/ to /e/ easier	/Ø/ to /e/ easier	?
Focality prediction (Table 1)	/e/ to /Ø/ easier	/e/ to /Ø/ easier	/e/ to /Ø/ easier
Subjects	Danish infants	German adults	Dutch adults
Perceptual task	conditioned headturn	MMN oddball paradigm	MMN oddball paradigm
Finding	/e/ to /Ø/ is easier	no asymmetry	/e/ to /Ø/ is easier

Table 2: Summary of acoustic and perceptual findings across three studies that assessed perceptual asymmetries for /e/-/ø/.

While the findings reported in Polka *et al.* [6] are compatible with NRV, their interpretation is post hoc and based on a small sample of vowels produced by one talker in each language (3-4 tokens per talker). The current research provides a stronger empirical test of the core NRV assumptions regarding asymmetries by examining the articulatory and

acoustic properties of the /e/-/ø/ contrast in Canadian French with a broader sample of talkers, using state-of-the-art electro-magnetic articulography (or EMA; [10,11]). We addressed two specific research questions. First, does focalization provide a more direct, principled account for perceptual asymmetries than vowel space peripherality? Second, is there a relationship between degree of articulatory constriction and formant proximity for the /e/-/ø/ contrast? In this paper, we report findings that address the first research question and outline how they guide ongoing analyses to address the second question.

2. METHODS

2.1. Participants

We acquired data from twenty native speakers of Canadian French (mean age: 28 years, *SD* ± 4.9 years; range, 19 to 39 years; all male). None reported a history of a speech, language, or hearing disorder.

2.2. Stimuli

Talkers produced randomized blocks of all ten Canadian French vowels (/i y u e ø o ɛ œ ɔ a). To control the vocal tract configuration at the onset of stimulus production, each vowel was embedded in the carrier phrase, "V as WORD," where V was one of ten vowels and WORD was a CVCV word with /p/ as C and the target vowel V in both vowel positions. For example: «a comme papa» («a as papa»).

2.3. Recordings

Speech movements were recorded using an AG500 EMA system (Carstens Medizinelektronik GmbH, Bovenden, Germany). Dynamic, articulatory sensors were affixed on the midsagittal surfaces of the upper lip, lower lip, jaw, tongue-tip, tongue-blade, and tongue-dorsum. Static, head reference sensors were affixed on the gingiva above the upper left and right incisors and behind each ear on the left and right mastoid processes. Participants were recorded while producing ten repetitions of each vowel, resulting in 100 tokens collected per participant. Stimuli were orthographically cued randomly. Simultaneous audio (with a sampling rate of 44.1 kHz) was recorded using an omnidirectional condenser head-worn microphone (Audio-Technica BP892) and the EMA sensor signals (with a sampling rate of 250 Hz) were recorded using Carsten's `CS5RECORDER` and `CS5VIEWER`.

2.4. Acoustic Analyses

This work is based on the analysis of 10 productions of /e/ and /ø/ by each talker (200 tokens per vowel and 400 vowels in total). To avoid coarticulation effects, only the first (isolated) vowel in the sentence «V as CVCV» was analysed. Speech signals were down sampled to 22.05 kHz (half the sampling frequency). Vowel onset and vowel offset were manually identified based on the appearance and disappearance of the first two formant frequencies on the spectrogram, respectively, using Praat [12]. The first three formants were extracted at vowel midpoint using Linear Predictive coding.

2.5. Articulatory Kinematic Analyses

The raw kinematic data first underwent a series of standardized pre-processing steps to rotate and translate each sensor position signal to the occlusal plane and then, using information from the head reference sensors, were corrected for head motion artifacts. The acoustic and kinematic movement signals were then computed, synchronized, and visualized together using the MATLAB-based toolbox *Mview* [13,14], developed by Mark Tiede at Haskins Laboratories. Ongoing analyses are based on the sensors located on the lips and tongue-dorsum because they provide the most direct information about vowel constrictions. Vertical lip separation (the Euclidean distance between the sensors affixed on the upper- and lower-lip sensors), lip protrusion (relative to the upper incisors), and the vertical displacement of the tongue-dorsum (relative to the upper incisors) at vowel midpoint are being extracted from the movement time series for each vowel token.

3. RESULTS

F1 and *F2* formant frequencies (averaged across tokens) are plotted in Figure 3 below for each of the 20 talkers. For every talker, *F2* frequencies were higher for /e/ than /ø/. As well, for every talker, vowel height was higher (*F1* frequencies lower) for /e/ compared to /ø/. These data confirm that, in Canadian French, /e/ is a more peripheral vowel than /ø/ in the standard *F1-F2* vowel space. The vowels were also analyzed to assess differences in formant distance or focalization. Findings are summarized in Table 3. These data clearly show that *F1* and *F2* frequencies are closer in productions of /ø/ compared to /e/; *F1* and *F3* formant distances show the same pattern. Both patterns are robust and evident in all 20 talkers (χ = 20; p <.0001). Some differences in *F2* and *F3* convergence were noted. Group means show that /e/ is more focal than /ø/ with respect to *F2-F3;* however, the *F2-F3* difference (on average) across vowels was less than half of the differences observed for *F1-F2* and *F1-F3*. Overall, this pattern is weak and inconsistent; it was evident for only 13 out of 20 talkers (χ = 1.8; p =.179). The other 7 talkers show the reverse patten (more focal *F2-F3* for /ø/) or no difference across the vowels. The formant proximity (focalization) patterns are shown schematically in

Figure 4 below. The dotted arrows highlight the more focal formant pattern for /ø/ compared to /e/, especially with respect to *F1-F2* and *F1-F3* formant proximity.

4. DISCUSSION

4.1. Acoustic-Perceptual Relations

Here we examined the Canadian French /e/-/ø/ contrast, which is articulatorily distinguished by lip-rounding (compression and protrusion). To confirm the focalization patterns observed for this vowel pair in prior studies, we recorded and analyzed a multi-talker sample of Canadian French /e/ and /ø/ vowel productions. Our acoustic findings align with the earlier (single talker) analyses reported in Danish, Dutch, and German. In our French corpus, /e/ was consistently more peripheral than /ø/ in *F1*/*F2* space. French /ø/ was also more focal than French /e/ with respect to *F1-F2* and *F1- F3* proximity. These formant patterns were quite robust and uniform across talkers. Like prior studies, weak and inconsistent differences in *F2-F3* proximity were noted. These data provide further evidence that when perceptual asymmetry predictions based on peripherality and focalization conflict, focalization is the winner. These findings support the core NRV claims that perceptual asymmetries reflect a sensitivity to articulatory patterns that are specified by focalization patterns.

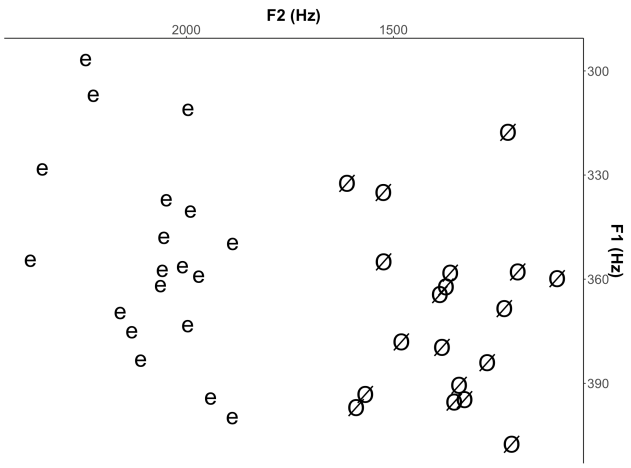

Figure 3: F1/F2 plot of French /e/ and /ø/ productions. Each symbol corresponds to the mean values (across 10 tokens) for 1 talker.

Location in the vowel space (peripherality) often predicts asymmetries for contrasts involving vowel height or front/back dimensions. However, this approach does not work for contrasts that involve lip-rounding, whereas a prediction based on formant convergence is more successful (see also [15]). These findings support the NRV claim that focalization provides a more principled account for perceptual asymmetries. Further research with other lip-rounding contrasts is needed to firmly establish this position.

		French (n=20)
F1-F2 (Hz)	e	1742
	Ø	**1019**
	difference	713
F1-F3 (Hz)	e	2335
	Ø	**1839**
	difference	509
F2-F3 (Hz)	e	588
	Ø	821
	difference	200

Table 3: Formant convergence analysis of French /e/-/ ø / vowels produced by 20 talkers. Shaded values indicate the more focal vowel within the pair.

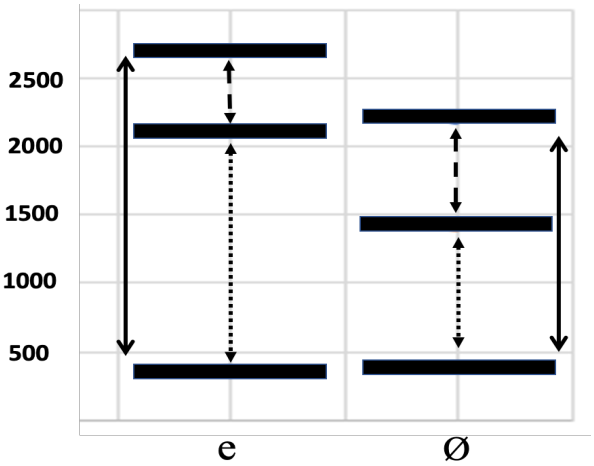

Figure 4: Schematic highlighting formant convergence patterns observed for French /e/ and /Ø/. The formant values shown correspond to mean values across 20 talkers. Lines connected with arrows indicate the close formant proximity for /ø/ compared to /e/, especially for *F1-F2* (dotted lines) and *F1-F3* (solid line).

4.2. Articulatory-Acoustic Relations

To further substantiate the NRV account, the next step is to examine how vowel constrictions are tied to focalization. In one prior study [7], image processing of 2-D video images of a model talker's face during vowel production showed more lip compression was associated with more focal *F1-F2* patterns. Ongoing analyses with the current EMA recordings are exploring how the lips and tongue-dorsum work in concert to alter the overall spectrum of /e/ and /ø/. We predict that the constriction degree of the tongue-dorsum and lips (horizontal and vertical aperture) will be greater during productions of /ø/ compared to /e/, and the extent of constriction will be correlated with the proximity of formants,with lip-rounding showing a larger effect on lowering *F3*.

5. REFERENCES

[1] L. Polka and O-S. Bohn, "Natural referent vowel (NRV) framework: An emerging view of early phonetic development," *Journal of Phonetics,* vol. 39, pp. 467-478, 2011.

[2] S. Tsuji and A. Cristia, "Which acoustic and phonological factors shape infants' vowel discrimination? Exploiting natural variation in InPhonDB," In Proc. of Interspeech, 2017*,* pp. 2108-2112.

[3] J.L. Schwartz, C. Abry, L.J. Boë, L. Ménard and N. Vallée, "Asymmetries in vowel perception, in the context of the Dispersion-Focalization Theory," *Speech Communication,* vol. 45, pp. 425–434. 2005.

[4] C. Eulitz and A. Lahiri, "Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition," *Journal of Cognitive Neuroscience*, vol.164, pp. 577-583, 2004.

[5] N. de Rue, T. Snijders and P. Fikkert, (2021). "Contrast and conflict in Dutch Vowels," *Frontiers in Human Neuroscience*, vol. 15, 2021. [Online serial] Available: doi: 103389/fnhum.2021.629648.

[6] L. Polka, M. Masapollo and O-S. Bohn, "Predicting asymmetries in vowel perception: formant convergence succeeds where peripherality fails," presented at the 181st Meeting of the Acoustical Society of America, Seattle, USA, 2021.

[7] M. Masapollo, L. Polka and L. Ménard,. "A universal bias in adult vowel perception - By ear or by eye," *Cognition*, vol. 166, pp. 358-370, 2017.

[8] M. Masapollo and F.H. Guenther, "Engaging the articulators enhances perception of concordant visible speech movements." *Journal of Speech, Language, and Hearing Sciences*, vol 62, No.10, pp.3679-3688, 2019.

[9] M. Masapollo, L. Polka, L. Ménard, L. Franklin, M. Tiede and J.L. Morgan, "Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation and configuration," *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 44, No.7, pp. 1103-1118, 2018.

[10] J.S. Perkell, M.H. Cohen, M.A. Svirksy, M.L. Matthies, I. Garabieta and M.T.T. Jackson, "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *Journal of the Acoustical Society of America,* Vol.92, No.6, pp. 3078–3096, 1992.

[11] T. Rebernik, J. Jacobi, R. Jonkers, A. Noiray and M. Wieling, "A review of data collection practices using electromagnetic articulography," *Laboratory Phonology*, Vol.12, No.6, 2021.

[12] P. Boersma and D. Weenink, (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved 17 December 2022 from http://www.praat.org/.

[13] M. Tiede, "MVIEW: Software for visualization and analysis of concurrently recorded movement data," Haskins Laboratory report, New Haven, CT. 2005.

[14] M. Tiede, "MVIEW: Multi-channel visualization application for displaying dynamic sensor movement," Haskins Laboratory report, New Haven, CT. 2010.

[15] C.T. Best and A. Faber, "Developmental increase in infants' discrimination of nonnative vowels that assimilate to a single native vowel," presented at the International Conference on Infant Studies, Brighton, UK, 2000.

6. ACKNOWLEDGEMENTS

The work reported here was supported by the Natural Sciences and Engineering Research Council of Canada under grant numbers 105397 (L.Polka) and 312395 (L. Ménard).