# Understanding Vowel Perception Biases – It's Time to Take a Meta-analytic Approach

Linda Polka & Yufang Ruan
McGill University

Matthew Masapollo
Boston University

**Abstract**
This chapter reviews four recent studies designed to examine several theoretical accounts of directional asymmetries in vowel perception. The studies provide cross-language data on adults' discrimination of vowels that fall within a given phonetic category. The results show that asymmetries emerge using unimodal acoustic and visual vowels, regardless of native language, and also using schematic non-speech visual analogs. We then integrate the data across these four studies in a mini-meta-analysis. Collectively, the findings provide strong support for the Natural Referent Vowel framework's central claims that (1) asymmetries reflect a "language-universal" sensitivity to formant convergence (focalization) and 2) that this sensitivity is a speech-specific bias reflecting human sensitivity to the way that articulatory movements shape the acoustic and optical structures of speech. We advocate for further research adopting a meta-analytic approach.

Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (Eds.), ***A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*** (pp. 561-582). Dept. of English, School of Communication & Culture, Aarhus University.

## 1. Introduction

In previous work we discovered that, in infants and adults, vowel discrimination is often asymmetric such that discriminating a vowel change in one direction is significantly easier compared to discriminating the same vowels in the reverse direction (Polka & Bohn, 2003; 2011). For example, infants were more accurate when discriminating a change from /ɛ/ to /ae/ compared to the reverse direction of change from /ae/ to /ɛ/. In infants, similar asymmetries are found across language groups showing that this pattern reveals a generic, universal bias rather than an effect of language-specific attunement or categorization. In adults, asymmetries have been observed for non-native and within-category vowel contrasts. Figure 1 (left panel) shows directional asymmetries that have been reported in the literature; the arrow connecting two vowels shows the direction in which discrimination of the vowel pair was significantly higher. Directional asymmetries follow a consistent pattern – the easier direction is the one in which the vowel to be detected (the B vowel in an AB sequence) is the more peripheral vowel within a standard articulatory/acoustic vowel space (F1/F2) vowel space. This suggests that perception favors vowels produced with more extreme vocal tract constrictions or configurations. The Natural Referent Vowel (NRV) framework was formulated to account for these findings and to guide research into the nature and significance of this perceptual bias (Polka & Bohn, 2011). According to NRV, directional asymmetries in vowel discrimination reveal a universal perceptual bias that is phonetically grounded in human capabilities for speech production and perception. This perceptual bias is posited to reflect our exquisite sensitivity to the way that articulatory movements shape the physical speech signal. Specifically, we propose that this bias is due to the increased salience of vowels produced with more extreme articulatory maneuvers, which give rise to well-defined spectral prominences in the acoustic speech signal due to formant frequency convergence, also known as *focalization*. The *focal vowel bias* is supported by cross-linguistic research on phonemic vowel contrasts (Polka & Bohn, 2011; Tsuji and Cristia, 2017).

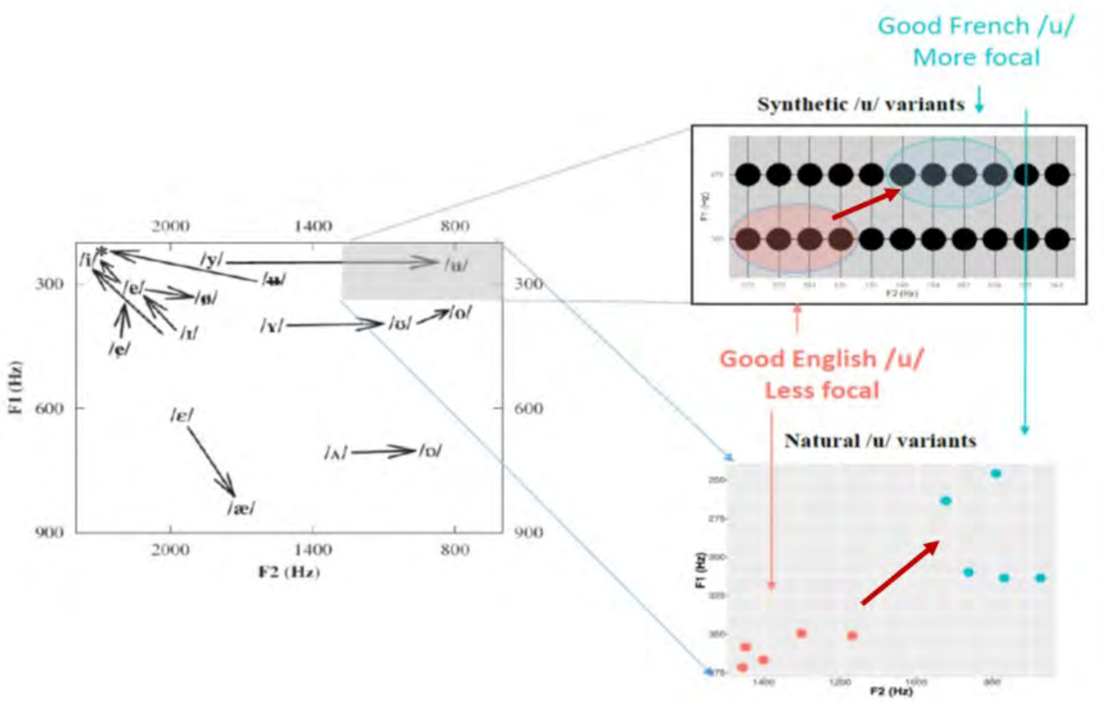



Figure 1.
**Left:** Directional asymmetries reported in the literature (from Polka & Bohn, 2011)
**Right top:** synthetic /u/ stimuli used in Masapollo et al JASA (2017): the red arrow shows the direction that was easier to discriminate for both French and English adults
**Right bottom:** natural /u/ stimuli used in Masapollo et al Cognition (2017); Masapollo et al JEP:HP& P (2018) and Masapollo et al JASA-EL (submitted); the red arrow shows the direction that was easier to discriminate for both French and English adults

The Native Language Magnet (NLM) model offers an alternative account of directional asymmetries in vowel discrimination (Kuhl et al, 2008). This model emerged from work investigating perception of within-category vowel variants. According to NLM, listening experience shapes perception to align with language-specific phonetic properties of native vowel categories. This leads to the formation of native language prototypes that act like perceptual magnets which attract less prototypic variants; NL magnets essentially warp the perceptual space around best or prototypic exemplars. One consequence of this magnet effect is asymmetric discrimination – detecting a change from a prototypic to a non-prototypic exemplar is harder compared to the reverse direction, i.e. a change from a non-prototypic to prototypic exemplar. Research focused on vowel contrasts has aligned with the NRV predictions pointing to language universal biases, while research focused on within-category differences has aligned with NLM predictions pointing to language-specific processes. Thus, a more direct and systematic comparison of these predictions is needed.

Recently, we have made significant progress towards disentangling these alternative views and confirming several central claims of the NRV framework. In Masapollo, Polka, Molnar & Ménard (2017), we systematically examined the role of universal and language-specific factors in vowel discrimination asymmetries. To do so, we synthesized an array of vowels that fall within the /u/ category. This /u/ vowel array is shown in Figure 1 (top right panel). The variants systematically varied in the proximity between their F1 and F2 values, in equal psychophysical steps along the mel scale. Critically, these variants were all clearly categorized as /u/ by both English and French adults, but also varied such that the best /u/ exemplars in French (circled in blue) were more focal than the best /u/ exemplars in English (circled in pink). The difference in focalization was due in part to the greater lip-rounding and protrusion that occurs in production of French /u/ compared to English /u/, which also increases F1 and F2 convergence for French /u/ productions compared to English /u/ productions**.**

Adults performed a categorial AX discrimination task designed to assess whether they show an asymmetric pattern in their discrimination of more-focal/French /u/ and less-focal/English /u/ tokens. Both monolingual English and monolingual French adults showed asymmetric discrimination as predicted by the NRV framework – showing better discrimination for a change from a less focal/English /u/ to a more-focal/French /u/ compared to the reverse direction. It is important to note that the NLM predicts that discrimination would be asymmetric but in opposite directions for French and English perceivers. Specifically, within each language group discriminating the change from a poor to good /u/ exemplar was expected to be better compared to the reverse (good to poor) direction. However, both French and English adults showed an asymmetry in the same direction and magnitude; thus there was no evidence that this pattern was affected by language experience as proposed by NLM. These findings confirm that the NRV bias reflects a sensitivity to formant convergence and also firmly establishes the presence of universal vowel processing biases that are distinct from the effects of language-specific attunement or prototype categorization.

A follow-up study provided evidence that the focal vowel bias can be observed when perceiving natural speech. This was shown by using auditory-visual recordings of English /u/ and French /u/ produced by a simultaneous bilingual female talker (Masapollo, Polka & Ménard, 2017). The F1 and F2 measures for these natural auditory /u/ variants are shown in Figure 1 – bottom right panel. A static screen shot showing one French /u/ token and one English /u/ token (taken at vowel midpoint) is also presented in Figure 2. As these images show, French /u/ and English /u/ are visually distinct. Video analyses confirmed that the lip-rounding and protrusion differences between these /u/ variants are conveyed in the dynamic visemes of these vowels. The focal vowel bias predicted by NRV was replicated when adults discriminated the natural French /u/ and English /u/ tokens presented in an auditory vowel discrimination task. As with the synthetic stimuli, both French and English adults showed the same directional asymmetry, which did not interact with language experience. The same finding emerged when we tested French and English adults' discrimination of the French /u/ and English /u/ tokens in a visual-only condition. As well, the focal bias was observed when English adults were tested in a bimodal (audio-visual) condition in which the auditory and visual channels were phonetically congruent, but not in a bimodal condition in which the audio and visual channels were phonetically-incongruent (French auditory /u/ dubbed onto English visual /u/; English auditory /u/ dubbed onto French visual /u/). These findings supply further evidence that the NRV bias reflects a universal sensitivity to formant convergence, independent of native-language categorization processes. Importantly, the finding that the same pattern emerges in visual vowel processing provides strong support for the NRV claim that this bias is phonetically grounded, reflecting a sensitivity to articulatory information available across different perceptual modalities.

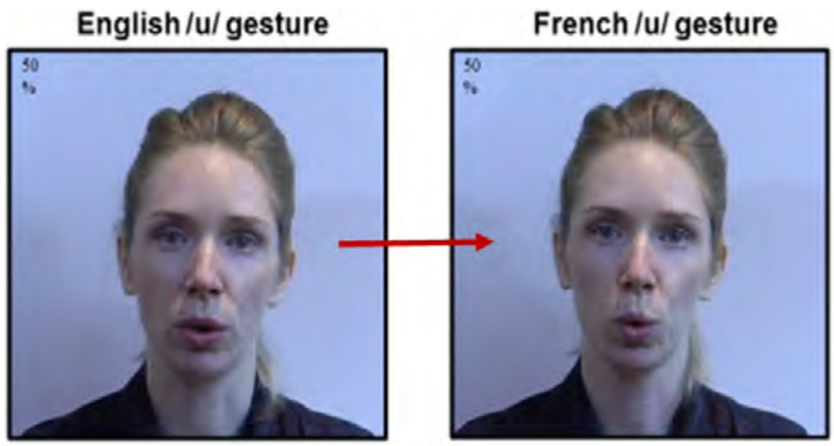



Figure 2. Model speaker's visual articulation at vowel midpoint. The red arrow shows the direction that was easier to discriminate for both French and English adults

In a subsequent study we probed the speech-specificity of the focal vowel bias in several ways (Masapollo, Polka, Ménard, Franklin, Tiede, & Morgan, 2018). First, we replicated the focal vowel bias in English adults using the same natural visual-only French /u/ and English /u/ stimuli while also tracking eye movements. Subjects attended selectively to the talker's mouth and also looked longer at the more focal /u/ tokens when discriminating these stimuli, confirming that articulatory features (increased lip rounding/ protrusion) specifying French /u/ drew more attention to the talking mouth. In a second study, no asymmetry was observed when English adults were tested with still images of the model speaker's face at vowel midpoint (as in Figure 2) where significant differences in lip rounding are observed across the two vowel types. This finding lends further support to idea that the focal vowel bias is tied to dynamic articulatory information, which is absent in a static image.

We gained further insights by testing adult discrimination using non-speech visual analogs of the lip movements for each vowel type. One visual analog condition was a point-light movie of the lip movements for each vowel token created from the video recordings by tracking four dots, two placed at the corners of the mouth and two placed on the top lip and bottom lip at the mouth mid-line as illustrated in Figure 3 (right top panel). The moving dots provide information on lip shape and movements. Although the moving dots are not recognized as a mouth, the French point light movies track a larger and more dynamic change in lip aperture compared to English point light movies. The same directional asymmetry that we observed for natural auditory and visual vowel tokens was observed when adults discriminated these point light movies; this was the case when subjects were told that the dots track lip movements and when they were not provided this information. However, the asymmetry was much weaker and failed to reach significance when the point light movies were rotated counter-clockwise by 45 degrees; in this orientation the configuration of the dots convey the same lip movement patterns but no longer depict a mouth-like shape. The point light analog findings suggest that adults require both the lip shape and movement patterns of these vowels to elicit the focal vowel bias, but recognition of a moving mouth is not required. In a second visual analog condition (Figure 3 – right bottom panel) we replaced the dots with a sideways figure-8 (∞) shape (aka a Lissajou curve) that changed in width and height over time to track the lip movements of each vowel token. This visual analog conveyed the distinct kinematic patterns present in lip movements over time for each vowel type but did not depict a mouth-like shape. Discrimination of the figure-8 analogs was not asymmetric providing further evidence that the focal vowel bias requires information specifying both lip shape and movement. Overall, the findings argue against an interpretation of the NRV bias as arising from simple, low level auditory or visual processes, and place the NRV bias squarely in the domain of speech perception.

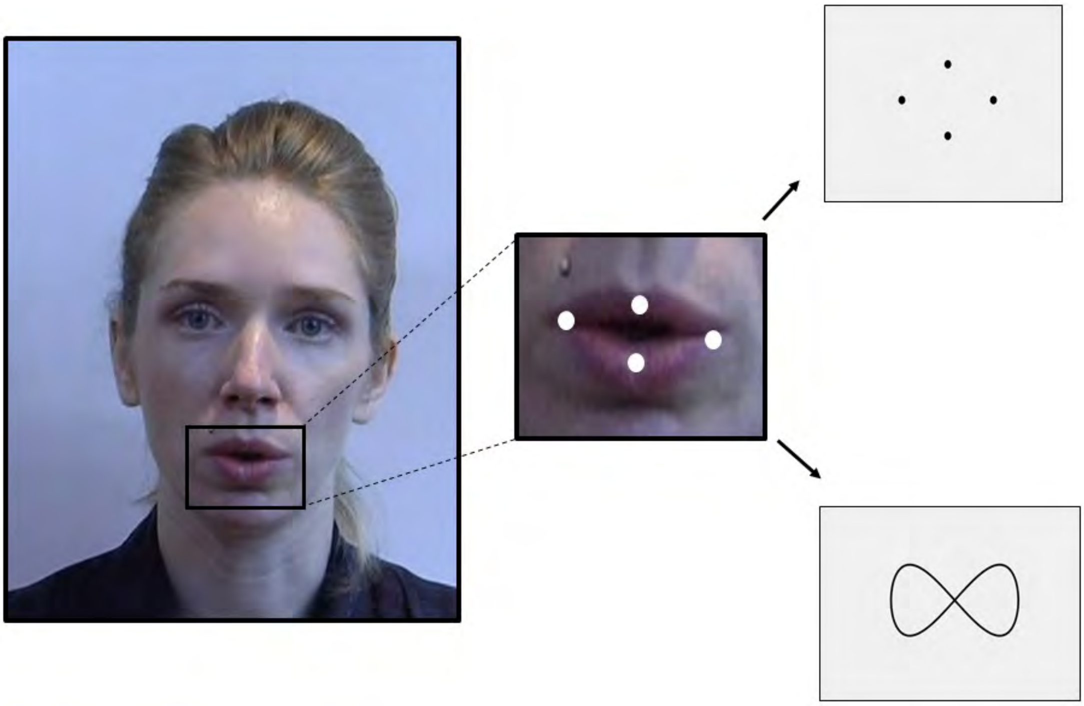


Figure 3. Dynamic non-speech visual analogs were created by tracking dots located at top/bottom and corners of the mouth. Point light movies (right top) conveyed lip shape and movement; Lissajou curves (right bottom) convey lip kinematics but not lip shape (Masapollo et al, 2018)

As a further test of the phonetic grounding of the NRV bias we examined task demands that impact phonetic processing (Masapollo, Franklin, Morgan, & Polka, submitted). In the work outlined above we used a categorial AX task with a 1500 ms inter-stimulus interval (ISI). This choice was based on prior work showing that phonetic processing is invoked by the memory demands imposed by a relatively long ISI. At a shorter ISI (e.g. 250 or 500 ms) perceivers can hold and compare acoustic details in auditory memory without engaging in phonetic encoding (e.g. Werker & Tees, 1983; Werker & Logan, 1985; Cowan & Morse, 1986). Auditory short memory fades quickly and thus when ISI is increased perceivers must rely on an encoded form of the stimulus to complete the task. Thus, prior work suggests that auditory processing is engaged when the ISI is short and phonetic processing is invoked when the ISI is longer (e.g. 1000 ms or 1500 ms). Thus, if the NRV bias is a phonetic bias, it should be reduced or absent when the ISI is shortened creating memory demands that favor auditory processing. English adults tested with natural productions of English /u/ and French /u/ showed reliable directional asymmetries when the ISI was 1500 ms but not when the ISI was shortened to 1000 ms or 500 ms. This finding, which emerged for both visual-only and auditory-only stimuli, contributes further evidence that the directional asymmetries expose a bias that is phonetically grounded.

There is no doubt that speech perception is strongly influenced by experience with a specific language. Collectively, the work summarized above confirms that universal perceptual processes also play a role in shaping adult vowel perception. Two published meta-analyses support these same conclusions with respect to infant vowel perception. The first meta-analysis, which included 19 articles containing 119 experimental records obtained using different behavioral and physiological methods, established that attunement to the native language begins to emerge in the first year of life (Tsuji & Cristia, 2013). The second meta-analysis was conducted on an updated dataset that also includes acoustic measures of the stimuli used (Tsuji & Cristia, 2017). This meta-analysis showed that spectral acoustic distinctiveness and order effects predicted by the NRV framework are reliable predictors of effect size in infant vowel discrimination tasks.

The work of Tsuji and Cristia inspired us to take a meta-analytic approach to assess predictions from the NRV framework with respect to adult vowel perception. As a first step we conducted a mini-meta-analysis integrating data across the four adult studies summarized above. Our mini-meta-analysis addressed several questions. First, what is the effect size due to the focalization bias when data are combined across the four studies summarized above? Second, as predicted by NRV, is the focalization bias effect size similar across language groups and across stimulus modalities? Third, as predicted by NRV, is the focalization bias effect size reduced when memory demands are decreased (by ISI manipulations) to promote acoustic processing and disfavor phonetic processing? To address the latter two questions we analyzed the effect of several moderator variables on the focalization bias effect size.

## 2. Method

### Database

Data extracted from four studies were included in this mini meta-analysis (see Table 1). We included data from 16 test conditions that utilized dynamic speech or non-speech analogs. We excluded one record that utilized static visual images as our hypotheses pertain to perception of dynamic speech or speech-like events. The resulting data base included findings obtained with diverse stimulus types including synthetic speech, natural speech (in auditory, visual and AV modalities), as well as point-light and Lissajou (∞) analogs of vowel lip movements. Despite the variability in stimulus types, the data from these different studies are suited for a meta-analytic approach, since all conditions utilize the same AX discrimination task to test adults (for restrictions, see next sections). For this mini-meta-analysis we used A prime scores, which was the dependent variable reported in each study.

1. Masapollo et al (2018) JEP:HPP (experiment 5)
2. Masapollo et al (submitted) JASA - EL
3. Masapollo et al (submitted) JASA - EL
4. Masapollo et al (submitted) JASA - EL
5. Masapollo et al (submitted) JASA - EL
6. Masapollo et al (2018) JEP:HPP (experiment 4)
7. Masapollo et al (2017) JASA (experiment 2)
8. Masapollo et al (2017) Cognition (experiment 2)
9. Masapollo et al (2017) Cognition (experiment 2)
10. Masapollo et al (2018) JEP:HPP (experiment 1)
11. Masapollo et al (2018) JEP:HPP (experiment 3.2)
12. Masapollo et al (2017) Cognition (experiment 1)
13. Masapollo et al (2017) Cognition (experiment 1)
14. Masapollo et al (2017) Cognition (experiment 3)
15. Masapollo et al (2018) JEP:HPP (experiment 3.1)
16. Masapollo et al (2017) JASA (experiment 2)

Table 1. References for each condition entered in the meta-analysis

### Moderator Variables – speech only conditions.

The effect of three moderator variables – language, modality, and ISI – was examined for the 12 conditions conducted with speech stimuli. The four conditions using non-speech visual analogs were removed because there is no data on ISI or language with these stimulus types and our hypotheses concerning these moderators pertain specifically to speech processing. For each condition, participants' native language, stimulus modality, and ISI (inter-stimulus interval) were coded as moderators. Only English- and French-speaking participants have been tested in included studies (EN=9, FR=3). The data base included three stimulus modalities: audio-only (*n*=6), visual-only (*n*=5), and audio-visual (*n*=1) stimuli. Due to limited audio-visual data, the conditions were collapsed to form two modality types: audio-only and AV or visual-only. By collapsing AV with Visual-only conditions we can examine whether the focalization bias effect is affected by the presence vs absence of visual speech information. The data set included three levels of ISI: 1500ms (*n*=8), 1000ms (*n*=2), and 500ms (*n*=2). Prior studies of vowel discrimination reveal a gradient decay in auditory memory (and decline in discrimination performance) as ISI is increased up to 2000 ms, especially for within category stimuli. This decay (and associated decline) is quite steep between 500 and 1000 ms and very gradual between 1000 ms and 1500 ms (Cowan & Morse, 1986; Experiment 2). For this reason (and given our limited data on ISI) the ISI conditions were collapsed to form to two ISI types: short ISI (500ms) and long ISI (1000ms and1500ms).

## Meta-Analytic Procedures

The analyses were conducted with the open-source package "metafor" (Viechtbauer, 2010) in R (R Core Team, 2018). Our effect size of interest represented the difference in discrimination by direction of vowel contrast. We calculated effect sizes based on the unbiased accuracy score (A' prime score, see Masapollo, Polka, & Menard, 2017 footnote two for more details) for each direction. Since each participant was tested in both directions, there were two dependent outcome values per sample.

Based on these values, we calculated Hedges' *g* effect size to represent the difference between directions within a sample. Like Cohen's d, Hedges' g reports the effect size in standard deviation units of the dependent variable while including a correction factor for small sample sizes. We also used Pearson correlation coefficient r for a within-subject experimental design correction. The calculations are found in Borenstein, Hedges, Higgins, & Rothstein (2009a).

Effect sizes were weighted by their inverse variance and entered into a random-effects model (Borenstein, Hedges, Higgins, & Rothstein, 2009b). A random-effects model without any moderators was applied to estimate the overall effect of focalization: *model = effect size, effect size* *variance, method = maximum likelihood estimator, weighted = TRUE.* was calculated to further estimate the proportion of heterogeneity over the total variability (Higgins, Thompson, Deeks, & Altman, 2003). Then, the *Q*-test of heterogeneity was performed to test whether the heterogeneity among the true effects was significant. We also conducted a sensitivity test on the overall effect of focalization (all conditions; no moderators)**,** leaving one effect size out at one time, this was done to detect influential cases and check the stability of the overall focalization bias effect size.[1]

Next, we ran three analyses to examine the effects of each of our three moderators. Because only speech conditions were included in the moderator analysis, we initially fitted a random-effects model including only the 12 speech conditions as a base model, *model = effect size, effect size variance, method = maximum likelihood estimator, weighted = TRUE*. For each moderator analysis, a mixed-effects model with a moderator, *model = effect size, effect size variance, mods = ~ moderator, method = maximum likelihood estimator, weighted = TRUE*, was used. The effect of each moderator was estimated and a *z*-test was conducted to examine whether the coefficient was significantly different from zero. Then a likelihood ratio test was conducted to compare each full model (all speech conditions; including the moderator) with the base model (all speech conditions; no moderator).

## 3. Results

In total, four studies with 16 experimental conditions (with a total of 242 adults tested) were included in this mini meta-analysis. Each condition is assigned a number and Table 1 provides the specific reference for each condition. The forest plot shown in Table 2 includes all conditions listed in order of effect size (ES) magnitude, with smallest ES at the top running to the largest ES at the bottom. Each line in this Figure 2 provides details on one condition ("tree") in the overall forest plot. The details provided include (from left to right) the assigned number, the language group tested, stimulus type, stimulus modality, ISI used in the AX task, the sample size, the mean and standard deviation for A prime scores for AX trials with a central (English /u/) to peripheral (French /u/) direction of change, mean and standard deviation for A prime scores for the reverse direction of

[1] We do not report a funnel plot analysis, which is typically conducted to assess publication bias, because this meta-analysis featured appropriate data from studies in our lab, all of which have been published or submitted for publication. Thus, there were no "file drawer" conditions excluded from this meta-analysis.

change, a plot showing mean difference in A prime scores across the two directions plotted as a square with 95% confidence intervals indicated. The dotted line is at zero (indicating no difference between directions); a mean value greater than zero corresponds to a directional asymmetry indicating a focalization bias. The squares that are plotted for each condition vary in size to show the weight of that condition in the overall meta-analysis, with larger squares denoting a greater weighting. The weighting is jointly determined by effect size, sample size, and margin of error (confidence interval). The last column on the far right, shows hedge's g for each condition and the confidence intervals (5%, 95%) around this effect size estimate. At the bottom of the forest plot the observed outcome is plotted as a diamond. The horizontal mid-point of the diamond corresponds to the overall effect size computed across all 16 conditions. The left and right points of the diamond correspond to the confidence limits (left = 5%, right = 95%) around the combined effect size. In this meta-analysis, the confidence interval is quite narrower for the combined effect size making it difficult to visualize. A smaller confidence interval is expected given that combining data across studies often yields a more precise estimate of the effect size. To the right of the diamond, Hedges' *g* and confidence limits (5%, 95%) around it are also reported. The statistics for the heterogeneity test (Q test of heterogeneity, residual heterogeneity proportion ) are indicated at the bottom left.

## The Effect of Focalization

The estimated overall effect size of the focalization bias was .34, with 95% CI [.24, .44], $z$=6.76, $p$<.0001 (See Table 2). This corresponds to a small to medium effect size using the classification offered by Cohen (1988). The amount of total heterogeneity (i.e. between studies variation), , was .0069 and 18.11% of total variability was explained by the heterogeneity instead of sampling error, Q (15)=21.76, $p$ =.114. This indicates that inconsistency among studies was relatively low. The result of the sensitivity test showed that there was no influential case and the estimated overall effect size varied from .32 to .36. Thus the effect size is stable within the small to medium range.

Table 2. The forest plot of the overall model showing 16 effect sizes.

| Study | Language | Type | Modality | ISI | n | Cen->Pher Mean | Cen->Pher SD | Pher->Cen Mean | Pher->Cen SD | Hedge's g [95%CI] |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | EN | Lissajou curves | V | 1500 | 16 | 0.61 | 0.08 | 0.62 | 0.09 | −0.14 [−0.68, 0.39] |
| 2 | EN | Natural | V | 500 | 14 | 0.93 | 0.03 | 0.93 | 0.03 | −0.02 [−0.37, 0.33] |
| 3 | EN | Natural | V | 1000 | 16 | 0.87 | 0.08 | 0.86 | 0.08 | 0.15 [−0.07, 0.37] |
| 4 | EN | Natural | A | 1000 | 16 | 0.79 | 0.08 | 0.77 | 0.06 | 0.20 [−0.21, 0.62] |
| 5 | EN | Natural | A | 500 | 15 | 0.81 | 0.09 | 0.79 | 0.09 | 0.22 [−0.02, 0.46] |
| 6 | EN | Rotated point−light | V | 1500 | 16 | 0.69 | 0.06 | 0.66 | 0.11 | 0.25 [−0.39, 0.88] |
| 7 | FR | Synthetic | A | 1500 | 15 | 0.98 | 0.02 | 0.97 | 0.04 | 0.32 [ 0.06, 0.59] |
| 8 | EN | Natural | V | 1500 | 15 | 0.86 | 0.08 | 0.83 | 0.08 | 0.35 [ 0.09, 0.60] |
| 9 | FR | Natural | V | 1500 | 13 | 0.83 | 0.12 | 0.78 | 0.1 | 0.44 [ 0.03, 0.85] |
| 10 | EN | Natural | V | 1500 | 16 | 0.8 | 0.09 | 0.76 | 0.1 | 0.44 [ 0.19, 0.70] |
| 11 | EN | Point−light | V | 1500 | 16 | 0.68 | 0.08 | 0.63 | 0.09 | 0.51 [ 0.16, 0.86] |
| 12 | FR | Natural | A | 1500 | 13 | 0.91 | 0.04 | 0.85 | 0.1 | 0.59 [ 0.12, 1.06] |
| 13 | EN | Natural | A | 1500 | 15 | 0.85 | 0.05 | 0.8 | 0.09 | 0.62 [ 0.17, 1.07] |
| 14 | EN | Natural | AV | 1500 | 15 | 0.89 | 0.06 | 0.85 | 0.06 | 0.65 [ 0.21, 1.08] |
| 15 | EN | Point−light | V | 1500 | 16 | 0.68 | 0.08 | 0.62 | 0.06 | 0.72 [ 0.02, 1.42] |
| 16 | EN | Synthetic | A | 1500 | 15 | 0.96 | 0.03 | 0.93 | 0.04 | 0.82 [ 0.26, 1.37] |

RE Model for All Studies ($Q = 21.76$, $df = 15$, $p = 0.114$; $I^2 = 18.11\%$) — 0.34 [ 0.24, 0.44]

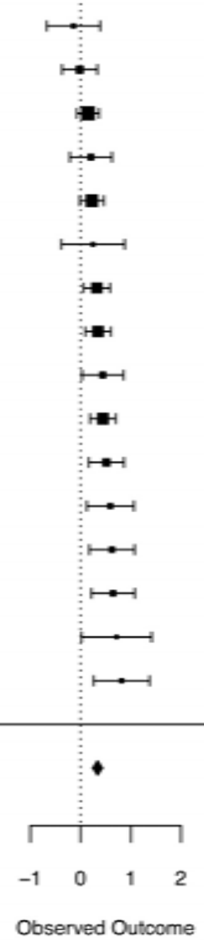



Table 2

## Moderator Analyses

## Effect of Focalization - Speech conditions only

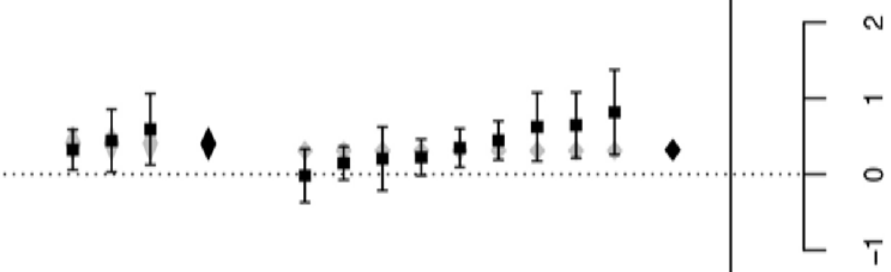
We initially fit a base model without any moderators by using 12 speech conditions (2, 3, 4, 5, 7, 8, 9, 10, 12, 13, 14, and 16), then each full model (including a moderator ) was compared. The estimated effect size of the focalization bias for speech conditions was .33 with 95% CI [.23, .44], $z$=6.36, $p$ <.0001. This corresponds to a small to medium effect size. The inconsistency among conditions was relatively low ( = .0057,  = 17.65%, Q (11)=16.46, $p$=.125). The result is consistent with the overall effect of focalization reported above.

### Language

A moderator analysis was conducted with the 12 speech conditions to determine if the focalization bias effect size was modulated by language experience. The resulting forest plot is shown in Table 3. The upper portion of the Table 3 shows the conditions conducted with French adults listed in order of effect size magnitude; the lower portion show the conditions conducted with English adults listed in order of effect size magnitude. For each language group, the estimated effect size and CI for that language group is plotted right below the corresponding section and is superimposed on each condition as a light grey diamond. The magnitude of the focalization bias did not differ across the two languages groups in our data set (beta=-0.10, 95% CI [-0.34, .15], $z$=-0.77, $p$=.439). As expected the full model with language as a moderator is not better than the base model (LRT=.57, $df$=1, $p$=.449). Both groups displayed a small to medium effect size for focalization bias. The weighted average effect size for French-speaking samples ($n$=3) was .40 with 95% CI [.20, .60], $z$=3.90, $p$<.0001. The weighted average effect size for the English-speaking samples ($n$=9) was .32 with 95% CI [.19, .45], $z$=4.93, $p$<.0001. French and English listeners completed identical perceptual tasks in the following conditions: 7 and 16, 8 and 9, 12 and 13. Within these matched conditions, the largest difference in focalization bias across language groups was observed for synthetic speech and within each direction discrimination performance was consistently higher for French adults than English adults.

**Table 3.** The forest plot for speech conditions: language as a moderator of focalization effect size.

| Study | n | Cen->Pher Mean | Cen->Pher SD | Pher->Cen Mean | Pher->Cen SD | Hedge's g [95%CI] |
|---|---|---|---|---|---|---|
| *French* | | | | | | |
| 7 | 15 | 0.98 | 0.02 | 0.97 | 0.04 | 0.32 [ 0.06, 0.59] |
| 9 | 13 | 0.83 | 0.12 | 0.78 | 0.1 | 0.44 [ 0.03, 0.85] |
| 12 | 13 | 0.91 | 0.04 | 0.85 | 0.1 | 0.59 [ 0.12, 1.06] |
| RE Model | | | | | | 0.40 [0.20, 0.60] |
| *English* | | | | | | |
| 2 | 14 | 0.93 | 0.03 | 0.93 | 0.03 | −0.02 [−0.37, 0.33] |
| 3 | 16 | 0.87 | 0.08 | 0.86 | 0.08 | 0.15 [−0.07, 0.37] |
| 4 | 16 | 0.79 | 0.08 | 0.77 | 0.06 | 0.20 [−0.21, 0.62] |
| 5 | 15 | 0.81 | 0.09 | 0.79 | 0.09 | 0.22 [−0.02, 0.46] |
| 8 | 15 | 0.86 | 0.08 | 0.83 | 0.08 | 0.35 [ 0.09, 0.60] |
| 10 | 16 | 0.8 | 0.09 | 0.76 | 0.1 | 0.44 [ 0.19, 0.70] |
| 13 | 15 | 0.85 | 0.05 | 0.8 | 0.09 | 0.62 [ 0.17, 1.07] |
| 14 | 15 | 0.89 | 0.06 | 0.85 | 0.06 | 0.65 [ 0.21, 1.08] |
| 16 | 15 | 0.96 | 0.03 | 0.93 | 0.04 | 0.82 [ 0.26, 1.37] |
| RE Model | | | | | | 0.32 [0.19, 0.45] |




Moderator Analysis: beta = −0.10, 95%CI [−0.34, 0.15], p = 0.439.

Table 3

**Stimulus Modality**

A moderator analysis was conducted with the 12 speech conditions to determine if the effect size for the focalization bias was modulated by stimulus modality. The forest plot is shown in Table 4. The upper portion shows the auditory-only conditions ordered by effect size magnitude. The lower portion shows the visual and AV conditions also ordered by effect size magnitude. For each modality, the estimated effect size and CI for that language group is plotted right below the corresponding section and is superimposed on each condition as a light grey diamond. Recall that two stimulus modality types were included in the analysis: audio-only and AV or visual-only. The effect size related to focalization did not differ across modalities types (beta = -0.07, 95% CI [-0.28, .14], $z$=-0.66, $p$=.511). As expected, the full model is not better than the base model (LRT=.43, $df$=1, $p$=.511). A small to medium effect size was observed in each modality. The weighted average effect size for audio-only conditions ($n$=6) was .36 with 95% CI [.22, .50], $z$=4.99, $p$<.0001, while the estimated effect size for AV /visual-only conditions ($n$=6) was .31 with 95% CI [.16, .45], $z$=4.09, $p$<.0001.

**ISI**

A moderator analysis was conducted with the 12 speech conditions to determine if the effect size for the focalization bias was modulated by ISI. The forest plot is shown in Table 5. Recall that two ISI types were included in the analysis: short ISI (500ms) and long ISI (1000 or 1500 ms). The upper portion shows the short ISI conditions ordered by effect size magnitude. The lower portion shows the long ISI conditions also ordered by effect size magnitude. For each ISI type, the estimated effect size and CI for that language group is plotted right below the corresponding section and is superimposed on each condition as a light grey diamond. The effect size related to focalization differ across ISI types (beta=.23, 95% CI [< .01, .47], $z$=1.96, $p$=.050). The full model is slightly better than the base model (LRT=3.77, $df$=1, $p$=.052). The estimated effect size for short ISI conditions ($n$=2) was .15 with 95% CI [-0.05, .34], $z$=1.44, $p$=.149. The estimated effect size for long ISI conditions ($n$=10) was .38 with 95% CI [.27, .50], $z$=6.62, $p$<.0001. The weighted averaged effect size was significant within the long ISI condition but not within the short ISI condition, which suggests that ISI is a potential moderator of the focalization effect size that may gain support if the data related to ISI was augmented.

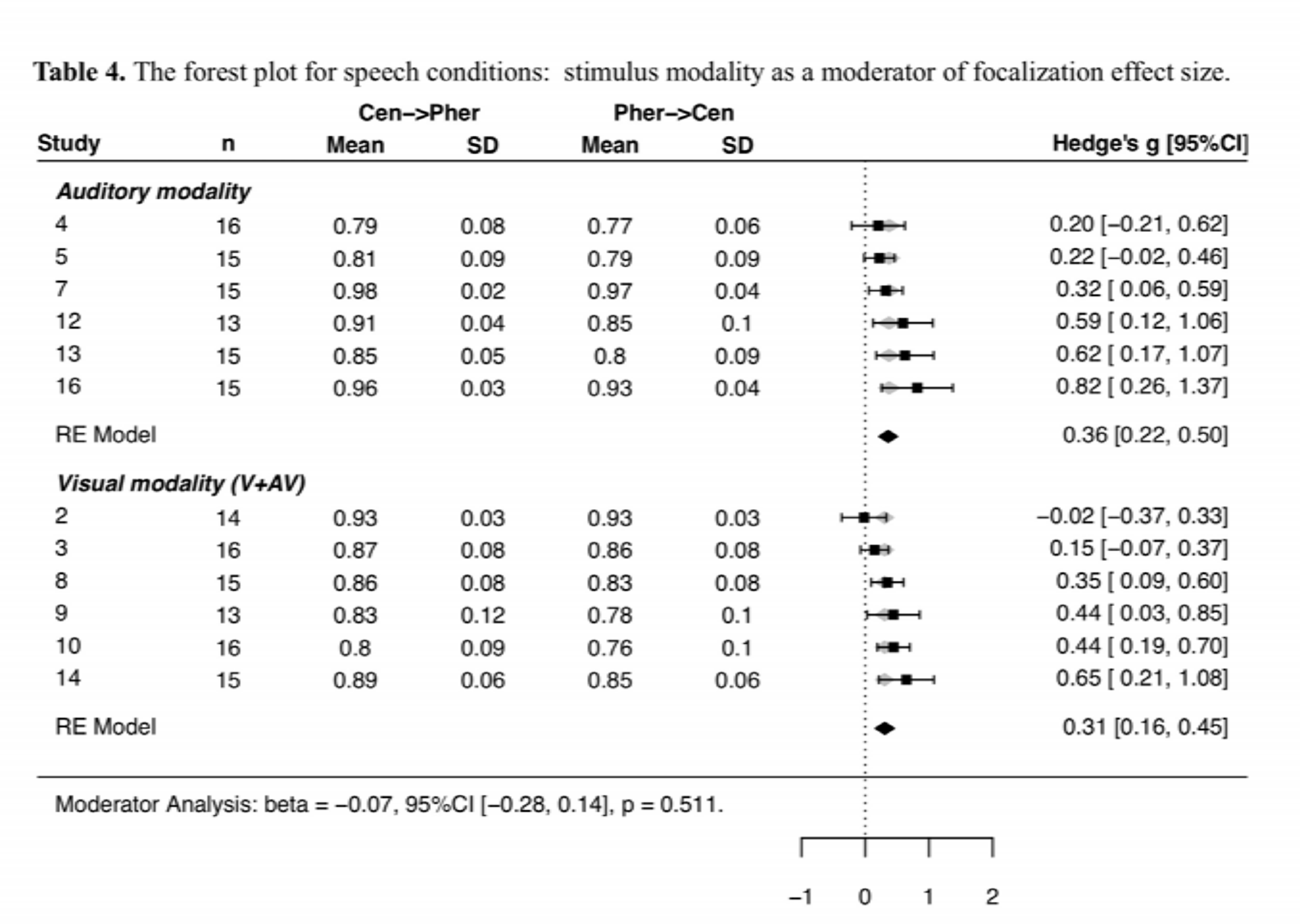


**Table 4.** The forest plot for speech conditions: stimulus modality as a moderator of focalization effect size.

| Study | n | Cen–>Pher Mean | Cen–>Pher SD | Pher–>Cen Mean | Pher–>Cen SD | Hedge's g [95%CI] |
|---|---|---|---|---|---|---|
| ***Auditory modality*** | | | | | | |
| 4 | 16 | 0.79 | 0.08 | 0.77 | 0.06 | 0.20 [–0.21, 0.62] |
| 5 | 15 | 0.81 | 0.09 | 0.79 | 0.09 | 0.22 [–0.02, 0.46] |
| 7 | 15 | 0.98 | 0.02 | 0.97 | 0.04 | 0.32 [ 0.06, 0.59] |
| 12 | 13 | 0.91 | 0.04 | 0.85 | 0.1 | 0.59 [ 0.12, 1.06] |
| 13 | 15 | 0.85 | 0.05 | 0.8 | 0.09 | 0.62 [ 0.17, 1.07] |
| 16 | 15 | 0.96 | 0.03 | 0.93 | 0.04 | 0.82 [ 0.26, 1.37] |
| RE Model | | | | | | 0.36 [0.22, 0.50] |
| ***Visual modality (V+AV)*** | | | | | | |
| 2 | 14 | 0.93 | 0.03 | 0.93 | 0.03 | –0.02 [–0.37, 0.33] |
| 3 | 16 | 0.87 | 0.08 | 0.86 | 0.08 | 0.15 [–0.07, 0.37] |
| 8 | 15 | 0.86 | 0.08 | 0.83 | 0.08 | 0.35 [ 0.09, 0.60] |
| 9 | 13 | 0.83 | 0.12 | 0.78 | 0.1 | 0.44 [ 0.03, 0.85] |
| 10 | 16 | 0.8 | 0.09 | 0.76 | 0.1 | 0.44 [ 0.19, 0.70] |
| 14 | 15 | 0.89 | 0.06 | 0.85 | 0.06 | 0.65 [ 0.21, 1.08] |
| RE Model | | | | | | 0.31 [0.16, 0.45] |

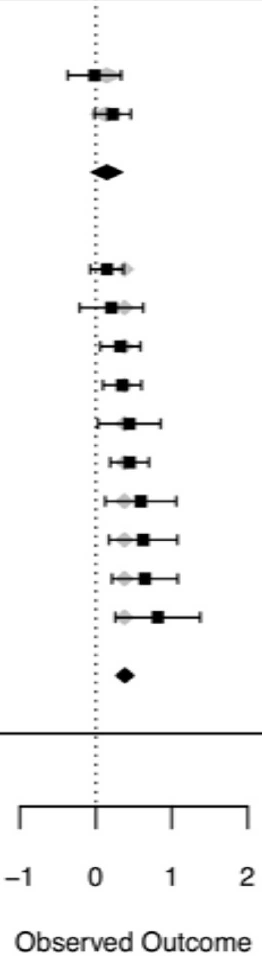
Moderator Analysis: beta = –0.07, 95%CI [–0.28, 0.14], p = 0.511.



Table 4

**Table 5.** The forest plot for speech conditions: ISI as a moderator of focalization effect size .

| Study | n | Cen–>Pher Mean | Cen–>Pher SD | Pher–>Cen Mean | Pher–>Cen SD | Hedge's g [95%CI] |
|---|---|---|---|---|---|---|
| ***Short ISI*** | | | | | | |
| 2 | 14 | 0.93 | 0.03 | 0.93 | 0.03 | −0.02 [−0.37, 0.33] |
| 5 | 15 | 0.81 | 0.09 | 0.79 | 0.09 | 0.22 [−0.02, 0.46] |
| RE Model | | | | | | 0.15 [−0.05, 0.34] |
| ***Long ISI*** | | | | | | |
| 3 | 16 | 0.87 | 0.08 | 0.86 | 0.08 | 0.15 [−0.07, 0.37] |
| 4 | 16 | 0.79 | 0.08 | 0.77 | 0.06 | 0.20 [−0.21, 0.62] |
| 7 | 15 | 0.98 | 0.02 | 0.97 | 0.04 | 0.32 [ 0.06, 0.59] |
| 8 | 15 | 0.86 | 0.08 | 0.83 | 0.08 | 0.35 [ 0.09, 0.60] |
| 9 | 13 | 0.83 | 0.12 | 0.78 | 0.1 | 0.44 [ 0.03, 0.85] |
| 10 | 16 | 0.8 | 0.09 | 0.76 | 0.1 | 0.44 [ 0.19, 0.70] |
| 12 | 13 | 0.91 | 0.04 | 0.85 | 0.1 | 0.59 [ 0.12, 1.06] |
| 13 | 15 | 0.85 | 0.05 | 0.8 | 0.09 | 0.62 [ 0.17, 1.07] |
| 14 | 15 | 0.89 | 0.06 | 0.85 | 0.06 | 0.65 [ 0.21, 1.08] |
| 16 | 15 | 0.96 | 0.03 | 0.93 | 0.04 | 0.82 [ 0.26, 1.37] |
| RE Model | | | | | | 0.38 [0.27, 0.50] |




Moderator Analysis: beta = 0.23, 95%CI [< 0.01, 0.47], p = 0.050.

Table 5

## 4. Discussion

In this chapter we present a qualitative review of recent adult cross-language research designed to examine the perceptual and cognitive processes underlying directional asymmetries in vowel perception. We also report the results of a mini-meta-analysis of this body of research which was undertaken to gain a more rigorous and comprehensive assessment of this work. This meta-analysis was limited to data from the series of studies reviewed above in which we assessed the focalization effect for a sub-phonemic contrast (more focal French /u/ vs less focal English /u/) across diverse stimulus types, language groups, and task demands.

Overall, the meta-analytic findings support the central tenets of the NRV framework. The collective evidence confirms that adults display a stable and reliable directional asymmetry in vowel discrimination that is tied to focalization differences. This bias has a small to medium effect size when measured in A prime units, which is a conservative (unbiased) index of discrimination.

Analyses of several variables that potentially moderate the focalization effect were also in line with NRV predictions. As predicted, the perceivers' native language did not modulate the focalization effect size. This further strengthens our claim that the NRV bias is distinct from the NLM effect. However, given the limited language diversity in this initial met-analysis, this should be re-assessed with an augmented data set. Importantly, although we claim that NLM and NRV describe distinct factors that shape vowel perception, they are not mutually exclusive. Interactions between these biases may emerge in other contexts or language groups.

Also as predicted, differences in stimulus modality did not modulate the focalization effect size. Thus, the focal vowel bias appears to be multi-modal and comparable in magnitude when assessed via vision or audition. This provides strong support for our claim that NRV is phonetically grounded and cannot be explained by general auditory processing biases alone.

The moderator analysis indicates a marginal trend for the focalization effect size to be modulated by the inter-stimulus interval used in the AX task. Thus, the NRV-based prediction regarding ISI was not firmly supported in this meta-analysis. However, the observed trends within the long and short ISI subgroups suggest that this factor may emerge in data set that is augmented with additional studies that include short ISI conditions. Thus, further research addressing this issue is needed to draw a firm conclusion regarding this task variable.

Overall most of the NRV frame-work predictions were supported in individual studies and also backed up in our integrative meta-analysis. As a next step, it will be informative to augment this meta-analysis to include data from other sub-phonemic and phonemic contrasts, other language groups, and other discrimination tasks. We invite researchers to contribute appropriate data to us as we begin to build a more comprehensive data set. Specifically vowel discrimination data (published or unpublished) that can be analyzed to assess effects of directional asymmetries in adult listeners, with native or non-native contrasts, will be informative.

Most of us are familiar with meta-analysis as a big undertaking that involves a thorough and comprehensive collection and integration of work within a specific field of research. However meta-analysis has much to offer and can be implemented on many different scales - with just a few experiments or with a large and multi-faceted data set. The main benefit of this approach is that it provides a way to look beyond an individual study and ground our interpretation in a more precise estimate of effect size gathered from a body of data rather than the dichotomous outcome of a single study. The focus on effect size (instead of null hypothesis tests) also pushes us to ask a deeper question – how big is an effect and is the magnitude of this effect modulated (or not) by specific factors as predicted by our hypothesis or conceptual framework. Thus integrating data in a meta-analytic framework provides both a more comprehensive and a more rigorous test of our hypotheses. By uncovering the strengths as well as the limitations of a body of research from a data analytic perspective, meta-analysis can also guide and motivate future research in productive directions. Following the example of Cummings (2012), we encourage our fellow speech scientists to add a meta-analytic perspective and tools to their research program.

## Acknowledgements

This chapter is supported by NSERC funding to L. Polka and CSC funding to Y. Ruan. Special thanks to Sho Tsuji and Christina Bergmann for their helpful input on the analyses and manuscript.

**References**

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009a). Effect Sizes Based on Means *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009b). Fixed-Effect Versus Random-Effects Models *Introduction to Meta-Analysis*: John Wiley & Sons, Ltd.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, (2nd Edition). Hillsdale: NJ: Erlbaum.

Cowan, N., & Morse, P. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, *79*, 500-507.

Cummings, G. (2012). *Understanding the New Statistics – Effect sizes, Confidence Intervals and Meta-Analysis*, Routledge, Taylor and Francis Group , LLC.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e), *Philos. Trans. Royal Soc. B* 363, 979-1000.

*Masapollo, M., Franklin, L., Morgan, J., & Polka, L. (submitted). Asymmetries in vowel perception arise from phonetic encoding strategies. *Journal of the Acoustical Society of America – Express Letters*

*Masapollo, M., Polka, L., & Menard, L. (2017). A universal bias in adult vowel perception – By ear or by eye. *Cognition*, *166*, 358-370. doi:10.1016/j.cognition.2017.06.001

*Masapollo, M., Polka, L., Menard, L., Franklin, L., Tiede, M., & Morgan, J. (2018). Asymmetries in unimodal visual vowel perception: The roles of oral-facial kinematics, orientation, and configuration. *J Exp Psychol Hum Percept Perform*. doi:10.1037/xhp0000518

*Masapollo, M., Polka, L., Molnar, M., & Menard, L. (2017). Directional asymmetries reveal a universal bias in adult vowel perception. *J Acoust Soc Am, 141*(4), 2857. doi:10.1121/1.4981006

Polka, L., & Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Communication*, *41*(1), 221-231. doi:10.1016/s0167-6393(02)00105-x

Polka, L., & Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467-478. doi:10.1016/j.wocn.2010.08.007

R Core Team. (2018). R: A language and environment for statistical computing Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Tsuji, S., & Cristia, A. (2014). *Percetual attunement in vowels: A meta-analysis. Developmental Psychobiology*, *56*(2), 179-191.

Tsuji, S., & Cristia, A. (2017). *Which Acoustic and Phonological Factors Shape Infants' Vowel Discrimination? Exploiting Natural Variation in InPhonDB*. Pa-

per presented at the Interspeech 2017.
Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *J Stat Softw*, *36*(3), 1-48.
Werker, J. F., & Tees, R.C. (1983). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, *75*(6), 1866-1878.
Werker, J. F., & Logan, J.S. (1985), Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*, 35-44.
Higgins, J. P. T., Tompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medicine Journal*, *327*(7414), 557-560.

*studies providing data for the meta-analysis.