Interspeech 2024
1-5 September 2024, Kos, Greece

# In search of structure and correspondence in intra-speaker trial-to-trial variability

*Vivian G. Li*

Yale University, USA
`liguo.vivian@gmail.com`

## Abstract

Intra-speaker variability is present even when the talker is uttering the same words in the same social and linguistic context. Studies have revealed that such intra-speaker trial-to-trial variability is connected to speech perception and is actively regulated during speech production. However, the relevant parameters in the variability that are under active regulation remain largely unclear. This study contributes to the discussion by examining the distributional properties of intra-speaker variability. Following up on a study that showed formants on hundreds of repetitions of the same word, measured at different points along the trajectory, are all normally distributed, we ask if those normal distributions correspond, i.e. whether a particular repetition would hold a stable position in the distributions across measurement points. Our analysis of 300 repetitions of /i, oʊ/ showed that strong correspondence typically spans one to two measurement points, and the strength of correspondence is phoneme-dependent and position sensitive.

**Index Terms**: vowel, formant, target, trajectory, speech production

## 1. Introduction

Variability is a hallmark of natural speech. Even for the same speaker uttering the same word in the same social and linguistic context, it is unlikely, if ever, to find two repetitions with the exact same acoustic profile. This kind of *intra-speaker trial-to-trial* variability (*intra-speaker variability* or *variability* hereinafter) is the topic of this study.

*Intra-speaker variability* has been productively utilized for probing the mechanisms of and the relationship between the production and perception of vowels. Smaller intra-speaker variability for one vowel relative to another during production is found to correlate with closer category boundary position in categorical perception (i.e. fewer items along the two-vowel continuum are categorized as members of the former vowel than of the later one) [1]. Smaller intra-speaker variability for a vowel is also associated with stronger acuity in vowel discrimination (e.g. higher accuracy in ABX [2] discrimination tasks) [3, 4, 5, 6]. Recent sensorimotor studies suggest that varying the formant frequencies in the auditory feedback to talkers would induce compensatory adaptation, where talkers shift the vowel formants they produce in the opposite direction of feedback manipulation [7, 8]; intra-speaker variability in production (without perturbed feedback) correlates with the magnitude of compensation as well as the amount of variability in the compensation [9, cf. 10]. Further, reduction in the variability in the feedback elicits increased variability in production, indicating that the variability itself may be under active regulation by the speech motor system [11, cf. 12].

It remain unclear what the mechanisms are for implementing the active control of variability [13], or what the relevant parameters in the variability are, that are regulated by the speech motor system. One approach for identifying the relevant parameters is to study the distributional properties of the variability per se. [14] is the first study to provide empirical evidence that intra-speaker variability is normally distributed. Analyzing skewness and kurtosis in the F1 and F2 of 300 repetitions of /i/ and /oʊ/ produced by a single talker, [14] showed that the distributions of formant frequencies at different points along the trajectories are all by and large normal (see Fig. 1).

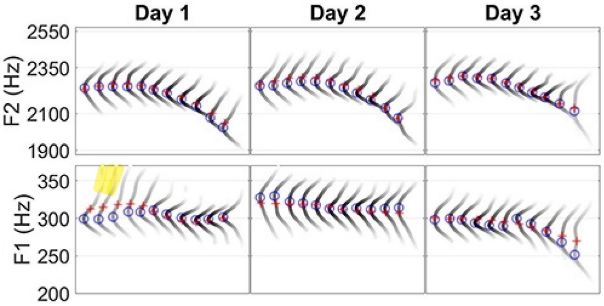

Figure 1: *Figure 4A from [14], showing the distribution of F1 and F2 values along 11 time-normalized measurement points (abscissas) of /i/ in* heed*, 100 repetitions per day. The grey contours are the probability density functions; on each contour, the blue circle represents the mode; the red dot represents the mean. The higher end of the distribution at measurement point1&2 of F1 in Day 1 are highlighted in yellow.*

This study is a follow-up inquiry of [14]; we ask: do the distributions along the formant trajectories correspond (correlate)? For instance, consider the tokens that are on the higher end of the distribution of F1 (from Day 1) at measurement point1 in Fig. 1, and the ones on the higher end of the distribution of F1 (from Day 1) at measurement point2, are they from the same set of repetitions? It may be tempting to assume that they do, but they do not have to, if all they need to satisfy is to form normal distributions. Fig. 2 illustrates two scenarios with seven schematic trajectories of F1 in each, focusing on two measurement points on the trajectories: the onset and the midpoint. Values in both panels are drawn from the same normal distributions: N (315, 15) for the onset and N (315, 10) for the midpoint. The left panel illustrates the scenario where the trajectories are parallel (non-crossing) to each other. In this scenario, the distributions at onset and midpoint correspond, as the position of a repetition relative to other repetitions in the distributions at both measurement points are the same. A repetition that is at

10.21437/Interspeech.2024-2456

the high end of the distribution at onset remains at the high end of the distribution at midpoint. The right panel illustrates an alternative scenario where the trajectories are not parallel to each other. The distributions at onset and midpoint in this scenario do not correspond, as each repetition's position in the distribution is variable at different measurement points. The repetition that is at the high end in the distribution at onset falls to the low end in the distribution at midpoint.

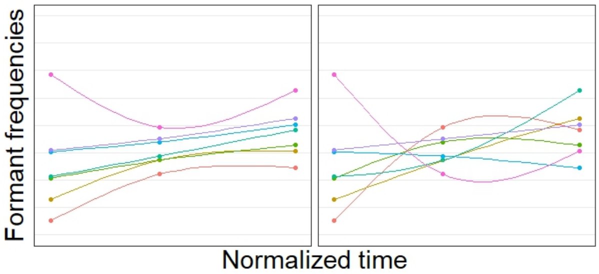

Figure 2: *Schematic trajectories of F1. Left: distributions at onset and midpoint correspond; Right: distributions at onset and midpoint do not correspond.*

Results from some studies indicate that the distributions may not correspond as perfectly as the left panel of Fig. 2 illustrates. Based on formant values measured at vowel onset, [15] distinguished *peripheral tokens*, whose formant values are farthest from the median (at onset), from *center tokens*, whose formant values are closest to the median (at onset). Subsequent analyses revealed that while peripheral tokens' formant values are closer to the median at vowel midpoint than at vowel onset (an inward centering effect), center tokens tended to diverge more from the median at vowel midpoint than at vowel onset (an outward effect). In the face of both inward and outward movement, the distributions of formant values at onset and midpoint may still correspond if the inward movement does not push non-center tokens to be as close to the median as the outward-pushed center tokens are, but not necessarily so. Thus, it raises the question that if the distributions do not correspond perfectly across all the measurement points on the trajectory, what is the range of strong correspondence (e.g. strong correspondence within 3 adjacent measurement points)?

A related question is, does the strength of correspondence vary by position (measurement point)? Scholarship in phonetics and phonology has long cherished the idea that vowel production involves targets [16, 17, 18]. In sustained production, target attainment may manifest acoustically as steady state in the vowel formants. For monophthongs, the midpoint, when the vowel is least likely to be under the influence of neighboring segments, is considered to be attaining or approximating the target. Studies have shown that on average, formant values at the vowel midpoint deviate less from the median than values at vowel onset do (an overall centering effect) [15]. Kinematic trajectories of tongue blade, tongue dorsum and jaw movement during vowel production exhibit similar patterns [19]. Systematic comparisons of variance at multiple points along formant trajectories in thirteen vowels show that for monophthongs, variance is smaller in the middle than at the onset or offset, but for diphthongs, variance is larger at the ends than in the middle of the vowel [20]. It is plausible that vowel targets affect the strength of correspondence as well, such that correspondence is stronger near monophthongs' midpoint or edges of diphthongs than elsewhere.

To summarize, this study seeks to explore how (strongly) the *distributions* of formant values measured at different points along the formant trajectories relate to each other. Specifically, we examine the scope (e.g. spanning x number of measurement points) of strong correspondence (correlation) between the distributions, and whether the strength of correspondence (correlation) is position-sensitive (near target vs. elsewhere).

## 2. Method

To examine the correspondence between the distributions of formant values, we need first a large set of repetitions of the same vowels. A larger number of repetitions is desirable because, for instance, in the extreme case of two repetitions, at all measurement points, there will always be one repetition that is above the mean and one below. If the two repetitions are sufficiently apart, it would be very likely that the distributions correspond, i.e. the measurements that are above the mean are from the same repetition. However, if there are twenty repetitions, it would be more likely that some repetitions may overlap or cross-over with each other, resulting in changes in relative position in the distribution.

Here we use the same dataset as [14], which contains four target syllables, /hid/ (*heed*), /gik/ (*geek*), /oʊd/ (*owed/ode*), and /doʊt/ (*dote*), each recorded for 300 repetitions from a native speaker of American English. Details on data collection are described in [14].

Raw formant values of the first three formants (F1, F2, F3) extracted (in Hertz) at eleven time-normalized points (inclusive of both ends) along the trajectories are transformed to the *mel* scale [21], so that each unit increment would correspond to the same amount of perceived increase in frequency. The values, grouped by measurement point, formant, and syllable, are then normalized (centered relative to the mean and scaled by standard deviation) to derive their relative positions in the distributions. A value of 2, for instance, indicates that the particular repetition at the measurement point of interest is two standard deviations above the mean, putting the particular repetition higher than 97.5% of all repetitions at that measurement point. After the normalization, a formant in a given repetition is indexed by eleven numbers that denote the repetition's relative position in the distribution at each measurement point.

Correlation tests between the positional indices are carried out to examine the strength of correspondence between distributions at different measurement points, as well as the scope of strong correspondence. Regression analyses are run to test if the strength of correspondence is positional sensitive.

## 3. Results

Pairwise Pearson's correlation coefficients were calculated for pairs of positional indices at different measurement points on the same formant. Following [22], a correlation coefficient with an absolute value equal to or larger than 0.7 is considered strong; a value between 0.7 and 0.3 is considered moderate; a value lower than 0.3 is considered weak.

### 3.1. Results for /i/

For indices on the F1, F2 and F3 of /i/ (/gik, hid/ together), all pairs except three exhibited moderate-to-strong correlation ($r \geq 0.3$). The three pairs of exception (at point 11 & 1, point 11 & 2, and point 11 & 3) were all on F2, with correlation coefficients approaching moderate strength (0.21, 0.24 and 0.29 respectively). All the tests were statistically significant ($p <$

0.0001).

Correlation is weaker on pairs of larger distance than on pairs that are closer. For F1 and F3, strong correlation is observed for pairs less than three points apart ($r$ = 0.9 for adjacent pairs; $r$ = 0.8 for pairs with a distance of two points). For F2, strong correlation is observed for pairs less than four points apart ($r$ = 0.95, 0.84, 0.76 for pairs of distance 1, 2, 3 respectively).

### 3.2. Results for /oʊ/

For indices on the F1, F2 and F3 of /oʊ/ (/oʊd, doʊt/ together), on average, only pairs less than four points apart on F1, pairs within a distance of five points on F2 and pairs within a distance of six points on F3 showed moderate-to-strong correlation ($r \geq$ 0.3); only adjacent pairs on F1, and pairs with a distance of two or less on F2 or F3 showed strong correlation ($r \geq 0.7$). All pairs except ten (6% in 165 pairs) were statistically significant ($p < 0.05$).

Similar to /i/, correlation was stronger between closer points than between more distant points. However, the span of moderate-to-strong correlation was much narrower on /oʊ/ than on /i/. Note that on the F1 of /oʊ/, the average correlation coefficients fell below 0.3 (weak) when the indices were four points apart, whereas indices on the F1 of /i/ were of moderate correlation ($r$ = 0.44) even ten points apart. Another notable contrast is that for pairs of all distances on /oʊ/, the average correlation coefficient tended to be higher on F3 than on F2 or F1. On /i/ however, pairs within seven points apart had higher average coefficients on F2 than on F1 or F3.

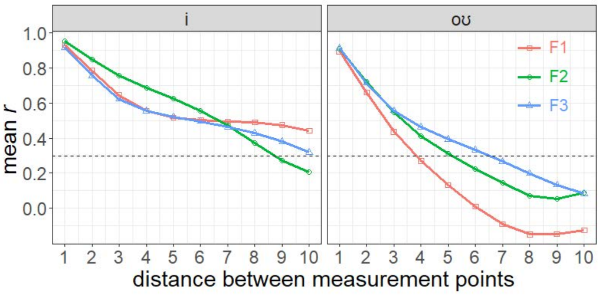

Figure 3: *Mean Pearson's correlation coefficient (r) by distance between points. Adjacent points, e.g. point 1 & 2, are of a distance of 1. A pair that is 10 points apart, e.g. point 1 & 11, are of a distance of 10.*

### 3.3. Effect of target

Since correlation is generally strong between adjacent measurement points, and weak between distant measurement points throughout the formant trajectory (see Fig. 4), positional differences (if it is present) may be hard to observe when the distance between pairs is very small (= 1) or very large. We therefore chose pairs of distance 3 and 4 for /i/ and pairs of distance 2 and 3 for /oʊ/ (whose average correlation coefficients are near 0.7), to avoid ceiling/floor effect. If correspondence (correlation) strength is stronger near targets, then among the chosen pairs, the ones that span over targets are expected to demonstrate stronger correlation than the ones that don't.

Following the generalizations in [20], the target for /i/ is taken to be at the vowel midpoint, the targets for /oʊ/ are taken to be at the vowel onset and offset. Accordingly, pairs at point 3 & 6, 4 & 7, 5 & 8, 6 & 9, 3 & 7, 4 & 8, 5 & 9 for /i/, and pairs at point 1 & 3, 2 & 4, 8 & 10, 9 & 11, 1 & 4, 2 & 5, 7 & 10, 8 & 11 for /oʊ/ are labelled as target-covering pairs.

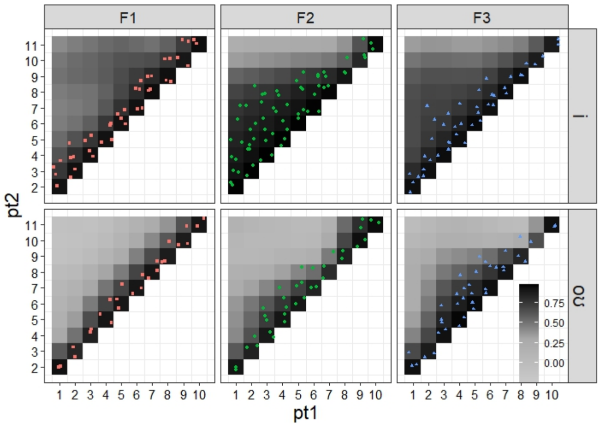

Figure 4: *Heatmap showing correlation strength between pairs of indices. Darker tiles represent stronger correlation. Each dot overlaying on the heatmap represents one pair of indices whose correlation coefficient is ≥ 0.7 (correlations coefficients are calculated for each syllable separately). The earlier point (pt1) in the pair is represented on the x axis, the later point (pt2) in the pair is represented on the y axis.*

Separate liner mixed effects models were fitted for /i/ and /oʊ/ using the lme4 package [23] in R [24]. The dependent variable was correlation coefficients between the chosen pairs (for /i/ there were 90 correlation coefficients = 15 pairs × 2 words × 3 fomants; for /oʊ/ there were 102 correlation coefficients = 17 pairs × 2 words × 3 formants). The base model included by-word random intercepts, and three fixed factors: *ptDist* (the distance between pairs, adjusted to 0 for the base level to facilitate interpretation of results, e.g. for /i/, *ptDist* for pairs of distance 3 is 0), *formant* (F1, F2, F3), and their interaction. Adding *target* (target-covering or not) to the base models improved model fit for both /i/ ($\chi^2(1) = 8.95$, $p = 0.003$), and /oʊ/ ($\chi^2(1) = 29.14$, $p < 0.0001$). Adding the interaction between *target* and *formant* further improved model fit (for /i/: $\chi^2(2) = 8.32$, $p = 0.016$; for /oʊ/: $\chi^2(2) = 15.92$, $p = 0.0004$). The addition of *target:ptDist* interaction, or *target:formant:ptDist* three-way interaction did not improve the models.

Results of the final models are summarized in Table 1. For F1 of /i/, the pairs covering the target did not differ significantly from the non-target-covering pairs (*Coef* = -0.011, *SE* = 0.042, $p = 0.8$); F3 showed a similar pattern (*Coef* = 0.089, *SE* = 0.060, $p = 0.139$). The effect of target on F2 however was significantly different from that on F1 (*Coef* = 0.169, *SE* = 0.060, $p$ = 0.005), indicating that the pairs that span the midpoint of F2 on /i/ showed a stronger correlation than the pairs at the edges. For /oʊ/, there was a significant lowering effect of target on pairs on F1 (*Coef* = -0.083, *SE* = 0.041, $p < 0.001$), and the pattern on F2 was similar (*Coef* = 0.012, *SE* = 0.058, $p = 0.834$), indicating that on F1 and F2 of /oʊ/, pairs that span over targets (i.e. at vowel edges) had a weaker correlation than pairs that span the middle of the vowel. The magnitude of the weakening effect of target was enhanced on F3 (*Coef* = -0.195, *SE* = 0.058, $p < 0.001$).

Table 1: _Summary of the final models for /i/ (top) and /oʊ/ (bottom): correlation $\sim$ (ptDist+target)*formant+(1| word)._

| Predictor | Coef | SE | t value | p |
|---|---|---|---|---|
| (Intercept) | 0.651 | 0.040 | 16.376 | <0.001 |
| ptDist | -0.09 | 0.042 | -2.128 | 0.033 |
| target=yes | -0.011 | 0.042 | -0.253 | 0.800 |
| formant=F2 | 0.027 | 0.051 | 0.526 | 0.599 |
| formant=F3 | -0.067 | 0.051 | -1.317 | 0.188 |
| ptDist:formant=F2 | 0.032 | 0.060 | 0.537 | 0.591 |
| ptDist:formant=F3 | 0.027 | 0.060 | 0.452 | 0.651 |
| target=yes:formant=F2 | 0.169 | 0.060 | 2.816 | 0.005 |
| target=yes:formant=F3 | 0.089 | 0.060 | 1.479 | 0.139 |
| (Intercept) | 0.697 | 0.064 | 10.956 | <0.001 |
| ptDist | -0.219 | 0.041 | -5.315 | <0.001 |
| target=yes | -0.083 | 0.041 | -2.012 | 0.044 |
| formant=F2 | 0.054 | 0.048 | 1.125 | 0.261 |
| formant=F3 | 0.14 | 0.048 | 2.944 | 0.003 |
| ptDist:formant=F2 | 0.049 | 0.058 | 0.832 | 0.405 |
| ptDist:formant=F3 | 0.077 | 0.058 | 1.315 | 0.188 |
| target=yes:formant=F2 | 0.012 | 0.058 | 0.21 | 0.834 |
| target=yes:formant=F3 | -0.195 | 0.058 | -3.341 | 0.001 |

## 4. Summary and discussion

In this study we set out to examine the strength of correspondence between distributions (represented by positional indices) of formant values at measurement points along the formant trajectories, the scope of strong correspondence, and its positional sensitivity. Based on 600 tokens of /i/ and /oʊ/, we found that for /i/ the strength of correspondence is generally moderate, as nearly all pairs of positional indices had a correlation higher than 0.3. The scope of strong correspondence on /i/ is limited to within two or three measurement points. Correspondence is weaker on /oʊ/ than on /i/, as pairs of positional indices that are four or more points apart on /oʊ/ showed weak correlation ($r \leq 0.3$); the scope of strong correspondence is also narrower on /oʊ/ than on /i/: on /oʊ/, strong correspondence is only found for adjacent pairs or pairs that are one measurement point apart.

### 4.1. Phonemic identity

These differences between /oʊ/ and /i/ may be rooted in their phonemic identities: /oʊ/ involves two vocalic qualities whereas /i/ involves only one. Maintaining a strong correspondence over a large span may be more challenging when one needs to transition out of the first element in the diphthong quickly in order to reach the second element, than when one only needs to produce a single vocalic quality. Additionally, we observed that for /oʊ/, correspondence tended to be stronger on F3 than on F2 or F1. This may be linked to the fact that /oʊ/ is a rounded vowel and F3 is relevant to lip rounding.

### 4.2. Positional sensitivity

Investigations on positional sensitivity showed promising results. Regression analyses revealed significant differences in the strength of correlation between indices pairs that span over the targets and the pairs that do not, on F2 of /i/, and F1, F2 and F3 of /oʊ/. The interpretation of the findings however is not straight forward. First, although F2 of /i/ showed stronger correspondence (correlation) near the target (vowel midpoint) than elsewhere, it may not capture the full pattern. Correspondence appears moderate or strong throughout the first half or nearly three quarters of F2 and F3 of /i/ (Fig5). The significant difference between the target region near vowel midpoint and regions at vowel edges is likely due to the weak correlation at vowel offset. Second, the direction of the effect of target on /oʊ/ counters our expectation. Indices pairs that are in the region of targets (edges of /oʊ/) showed weaker correlation than pairs in the middle of /oʊ/. One possible interpretation is that targets show a weakening effect on correspondence. Given that variance in formant frequencies is smaller near targets [20] and that tokens that move closer to the median and tokens that move away from the median are both present [15], weakened correspondence may arise as the compounded effect of the two. The reason that such a weakening effect is not observed near the midpoint of /i/ may be attributed to the oddity of /i/ itself: unlike many other monophthongs that show a significant decrease in the variance of formant frequencies in the middle of the vowel, /i/ tends to have a stable amount of variance [20]. Thus /i/ may not be an ideal candidate for observing target related effects. Another possible interpretation of the results is that perhaps correspondence tends to be strong when the speech motor system is focused on approaching a vowel target, and weakens when it prepares to transition away from the vowel target. This would provide a unified explanation for the weaker correlation observed at the offset of /i/, and at the onset and offset of /oʊ/.

### 4.3. Working with limited data

Analyses in this study are based on a small published dataset. Utilizing an existing dataset, as opposed to collecting new data, could be a merit rather than a drawback. Firstly, since research data collection is often funded by taxpayers, uncovering insights without collecting new data helps avoid unnecessary expenditure of public funds. The resources saved can then be allocated to fund other valuable research endeavors. Secondly, despite the growing number of publicly available corpora of understudied languages, research based on such corpora, especially non-descriptive ones, is still rare [25]. Part of the reason might be the limitations (in size and variation) of the data from such corpora. Some language archives, e.g. the UCLA archive [26], contains on average only a few short recordings per language. The dataset used in the current study is limited in that it features repetitions of four target words from a single speaker. As such this study may serve as inspiration for researchers exploring ways to extract insights from published (and often small) data.

## 5. Conclusion

Taken together, our findings indicate that the strength of correspondence is phoneme-dependent and position sensitive. The precise characterization of its patterns and mechanisms would require further research. For instance, the current study is based on formant frequencies measured at time-normalized points; future studies may examine the effect of duration. In this study we only analyzed two vowels; studies that include a larger set of vowels would be desirable.

## 6. References

[1] S.-C. Chao, D. Ochoa, and A. Daliri, “Production variability and categorical perception of vowels are strongly linked,” *Frontiers in Human Neuroscience*, vol. 13, p. 96, 2019.

[2] A. M. Liberman, K. S. Harris, J. A. Kinney, and H. Lane, "The discrimination of relative onset-time of the components of certain speech and nonspeech patterns," *Journal of experimental psychology*, vol. 61, no. 5, pp. 379–388, 1961.

[3] J. S. Perkell, F. H. Guenther, H. Lane, M. L. Matthies, E. Stockmann, M. Tiede, and M. Zandipour, "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," *The Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2338–2344, 2004.

[4] V. M. Villacorta, J. S. Perkell, and F. H. Guenther, "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *The Journal of the Acoustical Society of America*, vol. 122, no. 4, pp. 2306–2319, 2007.

[5] J. S. Perkell, H. Lane, S. Ghosh, M. L. Matthies, M. Tiede, F. H. Guenther, and L. Ménard, "Mechanisms of vowel production: auditory goals and speaker acuity," in *Proceedings of the Eighth international Seminar on speech production, Strasbourg, France*. Citeseer, 2008, pp. 29–32.

[6] M. K. Franken, D. J. Acheson, J. M. McQueen, F. Eisner, and P. Hagoort, "Individual variability as a window on production-perception interactions in speech motor control," *The Journal of the Acoustical Society of America*, vol. 142, no. 4, pp. 2007–2018, 2017.

[7] D. W. Purcell and K. G. Munhall, "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," *The Journal of the Acoustical Society of America*, vol. 120, no. 2, pp. 966–977, 2006.

[8] E. N. MacDonald, R. Goldberg, and K. G. Munhall, "Compensations in response to real-time formant perturbations of different magnitudes," *The Journal of the Acoustical Society of America*, vol. 127, no. 2, pp. 1059–1068, 2010.

[9] D. R. Nault and K. G. Munhall, "Individual variability in auditory feedback processing: Responses to real-time formant perturbations and their relation to perceptual acuity," *The Journal of the Acoustical Society of America*, vol. 148, no. 6, pp. 3709–3721, 2020.

[10] E. N. MacDonald, D. W. Purcell, and K. G. Munhall, "Probing the independence of formant control using altered auditory feedback," *The Journal of the Acoustical Society of America*, vol. 129, no. 2, pp. 955–965, 2011.

[11] B. Parrell and C. A. Niziolek, "Formant variability is actively regulated in vowel production," *Neuroscience*, vol. 7, no. 9, pp. 907–915, 2020.

[12] H. Wang and L. Max, "Inter-trial formant variability in speech production is actively controlled but does not affect subsequent adaptation to a predictable formant perturbation," *Frontiers in Human Neuroscience*, vol. 16, p. 890065, 2022.

[13] K. S. Kim, J. L. Gaines, B. Parrell, V. Ramanarayanan, S. S. Nagarajan, and J. F. Houde, "Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech," *PLOS Computational Biology*, vol. 19, no. 7, p. e1011244, 2023.

[14] D. H. Whalen and W.-R. Chen, "Variability and central tendencies in speech production," *Frontiers in Communication*, vol. 4, p. 49, 2019.

[15] C. A. Niziolek, S. S. Nagarajan, and J. F. Houde, "What does motor efference copy represent? evidence from speech production," *Journal of Neuroscience*, vol. 33, no. 41, pp. 16 110–16 116, 2013.

[16] C. P. Browman and L. Goldstein, "Articulatory gestures as phonological units," *Phonology*, vol. 6, no. 2, pp. 201–251, 1989.

[17] I. Lehiste and G. E. Peterson, "Transitions, glides, and diphthongs," *The journal of The acoustical society of America*, vol. 33, no. 3, pp. 268–277, 1961.

[18] G. E. Peterson and H. L. Barney, "Control methods used in a study of the vowels," *The Journal of the acoustical society of America*, vol. 24, no. 2, pp. 175–184, 1952.

[19] B. Parrell, M. Tiede, V. Gracco, and D. M. Shiller, "Kinematic evidence of centering during vowel production," in *Proc. 12th Intl. Seminar on Speech Production (ISSP2020)*, 2021, pp. 76–79.

[20] J. Jibson, "Variability of formant values at different time points of vowels," in *Proceedings of Meetings on Acoustics*, vol. 39, no. 1. AIP Publishing, 2019.

[21] G. Fant, "Analysis and synthesis of speech processes," in *Manual of phonetics*, B. Malmberg, Ed. Amsterdam: North-Holland, 1968, pp. 173–177.

[22] C. P. Dancey and J. Reidy, *Statistics without maths for psychology*. Pearson education, 2007.

[23] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[24] R Core Team, "R: A language and environment for statistical computing," 2021. [Online]. Available: https://www.R-project.org/

[25] D. H. Whalen and J. McDonough, "Under-researched languages: Phonetic results from language archives," in *The Routledge handbook of phonetics*, W. F. Katz and P. F. Assmann, Eds. Routledge, 2019, pp. 51–71.

[26] "UCLA Phonetics Lab Language Archive," accessed: 20240-06-12. [Online]. Available: http://archive.phonetics.ucla.edu