THE NATURE OF ACOUSTIC GOALS REFERENCED DURING HYPERARTICULATED SPEECH

Carissa A. Diantoro & Melissa A. Redford

Department of Linguistics, University of Oregon, Eugene, OR 97403-1290 USA
carissad@uoregon.edu & redford@uoregon.edu

ABSTRACT

Hyperarticulation references goals defined in acoustic-perceptual space. The question is whether hyperarticulation enhances contrast between goals in this space or exaggerates salient aspects of the goal without reference to its position in space. To address this question, CVC minimal triplets with English front lax vowels, /ɪ, ɛ, æ/, were elicited from 30 participants during the first and last phase of a 3-phase experiment to map the front lax vowel space. In the middle phase, participants produced only words with /ɛ/ vowels. These were "misheard" by the experimenter either as /ɪ/ or /æ/ vowels. Measurement on corrected repetitions indicated that participants fronted corrected /ɛ/ relative to baseline but did not systematically lower or raise it in response to the misheard vowel quality; corrected /ɛ/ was louder relative to baseline only in the /ɪ/ condition. The mixed results are discussed with reference to the nature of acoustic goals.

Keywords: speech motor goals, speech production, speaking style.

1. INTRODUCTION

Lindblom and colleagues [1] demonstrated that speakers achieve the same formant frequencies for different vowels, /i, u, ɑ, o/, whether their jaw varies naturally with articulation or is fixed at a particular distance relative to the maxilla. Moreover, compensation for the articulatory perturbation is evident as soon as the first glottal pulse of vowel articulation. Lindblom and colleagues interpreted these results to suggest predictive control over speech articulation and so the existence of acoustic-perceptual motor goals. Many other perturbation studies since then have expanded on these findings [2]. Results from auditory feedback perturbation studies provide especially compelling support for the hypothesis that speech articulation is guided by acoustic perceptual motor goals [e.g., 3, 4, 5].

Auditory feedback perturbation studies use an elegant design to test the effect of self-monitoring via perceptual channels on motor goal attainment. We have borrowed and adapted the design here to investigate the effect of feedback from others (i.e., misperception) on production, using the effect to investigate whether acoustic-perceptual motor goals reference discrete phonological representations of sounds, as is typically supposed [e.g., 1, 2, 3, 4], or whether they are dispersed in an extra-linguistic multi-dimensional perceptual space that is used to control articulation, as proposed in [5]. The distinction is important because it has implications for a whole-word production hypothesis [6], which has been offered as a developmentally sensitive alternative to the psycholinguistic hypothesis of phonological encoding during speech production [e.g., 7]. In what follows, we present the logic behind our approach for addressing the study aim.

It is well established that when a listener mishears a speaker, the speaker adjusts their articulation to better communicate that which was intended. In the clear speech literature, this adjustment has come to be known as hyperarticulation [8]. Hyperarticulation is very often supposed to represent a kind of goal maximization [8, 9] and so it has been studied to infer something about the representations that guide speech articulation [e.g., 9, 10, 11]. In clear speech studies, hyperarticulation is interpreted to suggest not only acoustic-perceptual speech motor goals but also that such goals reference discrete phonological representations of sound. For instance, Johnson and colleagues [9] describe a perceptual "hyperspace" effect, linking listener expectations for maximally distinct vowel productions to the speaker's speech motor goals in production. They then explicitly link such goals to discrete phonological representations of sounds, noting, for example, that "the search for acoustic and articulatory correlates of phonological units, in order to be successful, should focus on carefully produced speech (pp. 525-6)."

The linking of phonological units and speech motor control predicts that hyperarticulated speech results in contrast enhancement [e.g., 10, 11], but so far this prediction has not been tested against the prediction from the alternative extra-linguistic hypothesis. Studies specifically designed to investigate contrast enhancement (versus global enhancement) have focused on 2-way contrasts [10, 11]. Yet, the robust evidence for targeted hyperarticulation presented in these studies need not imply linguistic contrast enhancement per se; it could simply reflect the speaker's ability to increase the

acoustic-perceptual distance between discrete speech motor goals in an extra-linguistic acoustic-perceptual space. To distinguish between these alternatives, a more rigorous test of the linguistic contrast enhancement hypothesis is required. The current study provides this test. To do this, we borrow and adapt a study design from the auditory feedback perturbation paradigm to investigate the effect of hyperarticulation on production given a 3-way contrast: Houde and Jordan [3] asked speakers to whisper CVC words with an /ɛ/ vowel. Formant frequencies for this vowel were shifted (or not) and presented over earphones. The shift (= perturbation) moved the /ɛ/ formants either towards the speaker's own /i/ vowel or towards the speaker's own /a/ vowel. Speakers compensated in real-time for the perturbation by shifting their production of /ɛ/ away from either /i/ or /a/, depending on the perturbation. More specifically, if the auditory feedback perturbation suggested that the speaker's /ɛ/ was moved towards the speaker's /i/ and so was confusable with /ɪ/, speakers adjusted their production so that /ɛ/ became more [æ]-like; if the /ɛ/ production was moved in the direction of /a/ and so became confusable with /æ/, it became more [ɪ]-like.

In the current study, we use overt feedback from a listener about the category with which their /ɛ/ production is being confused. If speakers respond as they do in auditory feedback perturbation studies by adjusting intended /ɛ/ away from the misheard vowel, then we will have obtained strong evidence for the connection between speech motor goals and contrastive (i.e., phonological) speech sound representations. If not, then the possibility that such goals are extra-linguistic remains a viable alternative hypothesis.

2. METHODOLOGY

2.1. PARTICIPANTS

Participants were college-aged (18 to 22 years old) adults recruited via word-of-mouth through friend networks. Thirty participants completed the study, all reported normal hearing and typical speech-language development (none had a history of speech-language therapy); a pure-tone hearing screen test confirmed normal hearing.

2.2. MATERIALS

Materials were CVC minimal triplet words with the English front lax vowels, /ɪ, ɛ, æ/. There were 10 triplets. Table 1 lists the minimal triplet words.

/ɪ/	/ɛ/	/æ/
bid	bed	bad
bit	bet	bat
sid	said	sad
sit	set	sat
kid	ked	cad
kit	ket	cat
lid	lead	lad
lit	let	lat
rid	red	rad
writ	rhett	rat

Table 1: List of minimal triplets used in the study.

Each word was paired with a picture for elicitation purposes. Pictures were cartoon-like representations of the objects/events denoted by the CVC word. These were printed in color and laminated to create picture cards. The backside of each card was labelled with the target word. Nonce words were treated as proper names (i.e., Sid, Ket, Rhett) and paired with a cartoon character. Verbs (e.g., sit, said, sat) were paired with events. Adjectives (e.g., red, rad) were paired with objects, where the attribute of the object was highlighted during word-picture training. All picture–word correspondences were easily learned by all participants during a brief training that preceded the elicitation task. If a picture–word correspondence was forgotten during the task, the card was flipped over and shown to the speaker as a written reminder of the correspondence.

2.3. PROCEDURE

The elicitation task took place in adjacent sound-attenuated experimental rooms. The conjoining wall had a window. The experimenter and participant sat facing one another on either side of the window. This set up was used to enhance the plausibility of the mishearing manipulation: Although the participant was able to hear the experimenter over a baby monitor, they could not if the baby monitor was turned off. The participant could therefore easily imagine then how it might be difficult for the experimenter to hear them. The set up also allowed us to collect in person speech data while obeying pandemic-related restrictions on said data collection.

Elicitation occurred in three phases: a mapping phase, a misperception phase, and a remapping phase. The experimenter used the pictures to elicit all 30 words twice in different fixed random orders during the mapping and remapping phase. During the misperception phase, only /ɛ/ words were elicited and the experimenter "misheard" the participant as having produced either the matched [ɪ]-word ("ih" condition) or the matched [æ]-word ("ae" condition). The

participant was obliged to correct the experimenter by repeating the [ɛ]-word. The correction was done twice in a row on each trial (i.e., for each [ɛ]-word); each [ɛ]-word was also elicited twice during the misperception phase, again in random order. As in auditory feedback perturbation studies, the specific misperception manipulation was between subjects: half of the participants were assigned to the "ih" condition and half to the "ae" condition.

Participant speech was digitally recorded with a sampling rate of 44,100 Hz. A lavaliere microphone was attached to the participant's shirt to maintain a fixed mic-to-mouth distance.

2.4. ACOUSTIC MEASURES

Acoustic segmentation and measurement were completed on speech elicited from the 15 participants in the "ih" condition and the 15 in the "ae" condition. Vowels were segmented and labelled for each CVC word by a trained research assistant using Praat [12]. Vowel identity was determined with reference to the intended CVC word, based on the fixed random order used during the picture-based elicitation task. Overall vowel duration was automatically extracted from the segmentations. F1, F2, and vowel amplitude were automatically extracted at 3 equal intervals around vowel midpoint; specifically, at points equal to 40%, 50%, and 60% of the vowel duration. Formant frequency values were based on the automatic formant tracking algorithm used in Praat. The tracks were inspected by hand vowel by vowel against the spectrogram. Tracking parameters were adjusted where necessary (e.g., if F3 was tracked in lieu of F2).

2.4. ANALYSES

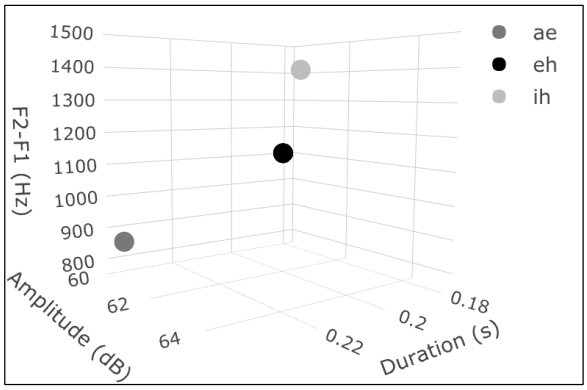
Mixed effects models were built using the lme4 package [13] in R [14] to test for the fixed effects of vowel, elicitation phase, and/or misperception condition while controlling for random effects of word and speaker, where appropriate. P-values were calculated using Satterthwaite's Method [15] with the lmerTest package [16] in R studio. Dependent measures were either raw measurement values or normalized values. Normalized values were used in the critical analyses that tested for the fixed effect of condition ("ih" versus "ae") and manipulation ("correction" versus "control") on /ɛ/ production. To compute these values, F1 and F2 frequencies, vowel amplitude, and vowel duration were each averaged for each word within each phase within each speaker. Word-to-word difference values were then computed by subtracting mapping phase values from misperception phase values (= correction) or from re-mapping phase values (= control).

3. RESULTS

3.1. FRONT LAX VOWEL SPACE

Figure 1 shows the position of the front lax vowels in acoustic space based on elicitations from the mapping phase of the elicitation task. Results from a mixed effects model on raw measurement values confirmed that the vowels differed from one another along all dimensions shown: [F2-F1 model intercept is "ae" at 853.55, SE = 16.95, $p < .001$; "eh" $\beta = 264.03$, SE = 14.60, $p < .001$; "ih" $\beta = 558.40$, SE = 21.48, $p < .001$. Amplitude model intercept is "ae" at 60.01, SE = 1.73, $p < .001$; "eh" $\beta = 0.73$, SE = 0.21, $p < .01$; "ih" $\beta = 0.42$, SE = 0.35, $p = > .05$. Duration model intercept is "ae" at 0.24, SE = 0.008, $p < .001$; "eh" $\beta = -0.05$, SE = 0.003, $p < .001$; "ih" $\beta = -0.06$, SE = 0.004, $p < .001$].

Figure 1: The 3 lax vowel targets in acoustic space. Error



bars indicate 95% confidence interval.

As expected, the vowel in /ɛ/ words occupied a position in acoustic space that was intermediate to the vowels in /ɪ/ and /æ/ words. Also, intended /ɛ/ was closer in acoustic space to intended /ɪ/ compared to /æ/. This relative distance between intended vowels in space was also as expected.

3.2. EFFECT OF MISHEARING ON /Ɛ/ PRODUCTION

Normalized values were used to test for the critical fixed effects of elicitation phase (correction vs. control) and misperception condition ("ih" vs. "ae"). The models thus included only a random slope and intercept for word. Analyses indicated a significant interaction between phase and condition on F1 [$\beta = 22.19$, SE = 6.32, $p < .001$]: F1 was raised in response to misperception, but only in the "ih" condition [$\beta = 25.06$, SE = 4.54, $p < .001$]. The interaction was not significant on F2. Instead, there was only a main effect of phase in both the "ih" condition [$\beta = 62.38$, SE = 10.23, $p < .001$] and "ae" condition [$\beta = 42.75$, SE = 10.99, $p < .001$]. There was no simple effect of misperception condition on /ɛ/. Importantly, whether

/ɛ/ was misheard as /ɪ/ or as /æ/, participants adjusted by raising F2. Misperception also had an effect on /ɛ/ amplitude and duration, but only in the "ih" condition [amplitude: $\beta = 5.63$, SE = 0.53, $p < .001$; duration: $\beta = 0.03$, SE = 0.007, $p < .001$]. The effect of misperception on /ɛ/ can be seen in Figure 2, which shows the position of the participants /ɛ/ vowel in acoustic space as a function of misperception condition; its default position, based on elicitations obtained during the mapping phase of the experiment, is also shown.

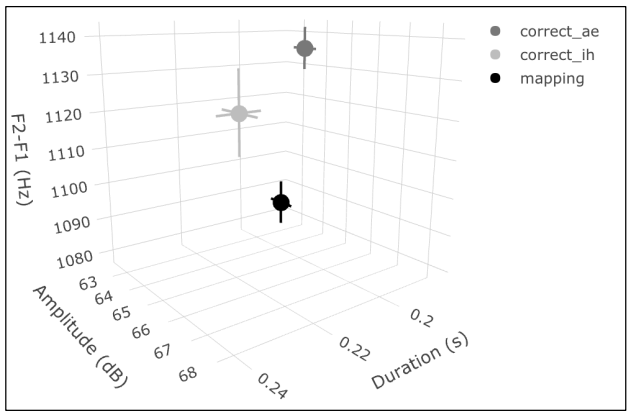


Figure 2: Mid-front lax vowel target shown as a function of misperception condition ("ih" vs. "ae"). The default position of the same vowel is also shown. Error bars indicate 95% confidence intervals.

4. CONCLUSION

Hyperarticulation references goals defined in acoustic-perceptual space. The question addressed in this study was whether this goal incorporates paradigmatic information relevant to linguistic contrast. The results were that, whether the experimenter misheard intended /ɛ/ as /ɪ/ or as /æ/, participants corrected the misperception in the same way – that is, they adjusted their articulation so that the vowels they produced varied in the same direction away from the control vowel along the measured dimensions (vowel formants, amplitude, duration). This result is inconsistent with the prediction of linguistic contrast enhancement. When differences by misperception condition were found, these indicated that /ɪ/ misperception elicited a greater degree of articulatory adjustment to /ɛ/ than /æ/ misperception. This result suggests an implicit awareness that [ɛ] is more apt to be confused with [ɪ] than with [æ] due to the relative positions of these sounds in acoustic space. Note that such a suggestion does not require speech motor goals to reference phonemic contrast; it requires only that the participant note the acoustic similarity between the experimenter's CVC word production due to misperception (e.g., "bid") and the intended CVC word that the participant had produced (e.g., "bed"), and then seek to further differentiate their own production from that of the experimenter's mistaken production.

5. ACKNOWLEDGMENTS

This research was wholly supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD) under grant R01HD087452 (PI: Redford). The content is solely the authors' responsibility and does not necessarily reflect the views of NICHD.

6. REFERENCES

[1] Lindblom, B., Lubker, J., Gay, T. 1979. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *J. Phonetics*, vol. 7, no. 2, pp. 147-161.

[2] Perrier, P., Fuchs, S. 2015. "Motor equivalence in speech production," In: Redford, M. A. (eds), *Handbook of Speech Production*. Wiley, 225–247.

[3] Houde, J. F., Jordan, M. I. 2002. Sensorimotor Adaptation of Speech I: Compensation and Adaptation. *J. Speech Lang. Hear. Res.*, vol. 45, pp. 295-310.

[4] Purcell, D. W., & Munhall, K. G. 2006. Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *J. Acoust. Soc. Am.*, vol. 120, pp. 966-977.

[5] Davis, M., Redford, M. A. 2019. The emergence of discrete perceptual-motor units in a production model that assumes holistic phonological representations. *Front. Psychol.*, vol. 10, pp. 2121.

[6] Redford, M. A. (2019). Speech production from a developmental perspective. *J. Speech Lang. Hear. Res.*, vol 62, pp. 2946-2962.

[7] Levelt, W. J. 1989. Speaking: From intention to articulation. MIT press.

[8] Lindblom, B. 1990. "Explaining phonetic variation: A sketch of the H&H theory," In: Hardcastle, W.J., Marchal, A. (eds) *Speech Production and Speech Modelling.* Springer, Dordrecht. pp. 403-439.

[9] Johnson, K., Flemming, E., Wright, R. 1993. The hyperspace effect: Phonetic targets are hyperarticulated. *Language*, pp. 505-528.

[10] Schertz, J. 2013. Exaggeration of featural contrasts in clarifications of misheard speech in English. *J. Phonetics,* vol. 41, no. 3–4, pp. 249–263.

[11] Wedel, A., Nelson, N., & Sharp, R. 2018. The phonetic specificity of contrastive hyperarticulation in natural speech. *J. Mem. Lang.,* vol 100, 61-88.

[12] Boersma, P., Weenink, D. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.37, retrieved 16 December 2020.

[13] Bates D, M.chler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.,* vol. 67, no. 1, pp. 1–48.

[14] RStudio Team. 2021. RStudio: Integrated Development Environment for R. RStudio.

[15] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bull.*, vol 2, pp. 110-114.
[16] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. J. *Stat. Softw*., vol. 82, pp. 1-26.