

Vowel targets without a hyperspace effect.

(2004) *Language*, 80,377-378

D.H. Whalen
Haskins Laboratories
New Haven, CT

H.S. Magen
Haskins Laboratories
New Haven, CT

M. Pouplier
Haskins Laboratories
New Haven, CT

A.M. Kang
Haskins Laboratories
New Haven, CT

K. Iskarous
Haskins Laboratories
New Haven, CT

LETTERS TO LANGUAGE

Language accepts letters from readers that briefly and succinctly respond to or comment upon either material published previously in the journal or issues deemed of importance to the field. The editor reserves the right to edit letters as needed. Brief replies from relevant parties are included as warranted.

Vowel targets without a hyperspace effect March 17, 2004

To the Editor:

In a paper with broad implications published in this journal, Keith Johnson, Edward Fleming, and Richard Wright (hereafter, JFW; 'The hyperspace effect: Phonetic targets are hyperarticulated', *Language* 69.505–28, 1993) provided experimental evidence suggesting that listeners preferred vowels with more extreme values than those they produced. In their study, listeners were to pick a 'best exemplar' of a vowel from a systematic exploration of a vowel space created on a speech synthesizer. The chosen exemplars (averaged across male and female listeners) were more extreme in formant values than the average productions (by the males). JFW claimed this 'hyperspace' effect underlies all vowel production, even though the targets were more extreme than their speakers produced even when hyperarticulating. The analogy that they provide to motivate the hyperspace effect is that full vowels cannot be predicted from reduced vowels, so reduced vowels cannot be taken as the underlying form (p. 524). By assuming that careful speech, or hyperarticulation, is similarly unpredictable, JFW expected that vowel targets would be hyperarticulated.

In their perceptual study, this is what seemed to be happening: Preferred vowels were outside the articulatory space of the male talkers. We have elsewhere shown that the experimental data depends both on the acoustics of the synthetic stimuli used to measure the effect and on the dialect of the listeners (D. H. Whalen, Harriet S. Magen, Marianne Pouplier, A. Min Kang, and Khalil Iskarous, 'Vowel production and perception: Hyperarticulation without a hyperspace effect', *Language and Speech*, to appear, 2004). Here we summarize the results and further explore some of the theoretical reasons for finding that there is no need for a hyperspace effect.

First, for both studies, dialect affects production and thus shifts the vowel space. In JFW's data, the one non-California speaker shows that

this is true. The production values for their speaker from New York are given separately; if the perception results are compared to the New York speaker's production space, only two of eleven vowels exhibit a hyperspace effect. The fronting of back vowels in the California dialect is responsible for much of the apparent hyperspace effect, as confirmed by the Rhode Island speakers in our study.

Second, in our study, as in JFW's, even when speakers hyperarticulated, their productions were less extreme than their perceptions. This would imply that hyperarticulations should lie on a line between the full vowel productions and the perceptual (hyperspace) target. In both sets of production results, the directions in which the hyperarticulations went in formant space were quite sensible. For our data, the direction from full to hyperarticulated vowel was away from centralization for five vowels, decentralized for either F1 or F2 for five vowels, and aberrant for /u/ (where the full and hyperarticulations were nearly the same). The perceptual targets, by contrast, were fairly randomly distributed relative to either full or hyperarticulated vowels. If hyperarticulation is supposed to be asymptotically closer to a 'hyperspace' target, then there is no evidence of that trend in either our data or that of JFW.

Third, the preferred perceptual targets of the female listeners were essentially identical to those of the male listeners, even though those targets were 'reduced' from the standpoint of the female vowel spaces (see Whalen et al.). What this indicates is that both sets of listeners were normalizing the vowel space of this synthetic 'talker'. By collapsing the perceptual results but not the production results, JFW failed to address the issue.

Fourth, the method used by JFW to test the perception seems to lead to more extreme judgments than listeners actually prefer. This 'method of adjustment' allows the listener to pick vowels from a grid until the best match for an English vowel (as specified by the keyword at the top of the computer screen) is found. Subjects are inclined to explore the whole space and to prefer the edges. We found this to be the case when the same stimuli were judged via a rating task rather than the method of adjustment. The rating task presented each stimulus separately for a judgment as to vowel identity and goodness of representing that vowel. These results were much closer to the

productions than the method of adjustment perceptions were. The appearance of a hyperspace effect, then, seems to have been caused by several factors.

These results show us that listeners are remarkably capable of adjusting to a speaker's vocal characteristics, even when that 'speaker' does not really exist. The formants used to create the stimuli for both JFW and Whalen et al. specified a vocal tract of a certain length, roughly that of an average male talker, with F0 values typical for a male talker. The perceptual result was a simulated talker whose formant values were somewhat more extreme than those of the average American talker, and listeners were quite consistent in identifying those vowels as such. To fully test the hyperspace proposal, it would be necessary to use a range of synthetic voices differing in the dimensions of the underlying vocal tract in the way that the speakers' vocal tracts vary. Even so, such a demonstration could be interpreted as showing only that vowel spaces are different, both in extent and location, and that listeners are sensitive to the information in the vowel signal that tells them what that speaker is like.

One implication of JFW's original proposal is that the true study of a language's vowel space depends not on acoustic measurements but on perceptual testing of preferred targets. However, JFW recommend only that hyperarticulated speech be elicited (pp. 525–26), despite asserting that it requires hyperspace targets to predict production. Their recommendation seems closer to the underlying state of affairs, and even the need for hyperarticulations is not well established. The notion that hyperarticulations cannot be predicted from full articulations just because full articulations cannot be predicted from reduced ones does not necessarily hold. At least for the point vowels, the physiological mechanisms that would make a narrower, and thus more extreme, constriction are fairly easy to predict given a full vowel. Languages tend to maximize both the clarity of a particular vowel and its distance from other vowels in the system.

JFW's claim that vowel spaces need a hyperspace description has far-reaching implications, including the elicitation issue just discussed. No production, even the hyperarticulated ones, would ever live up to the hyperspace ideal. It appears that the hyperspace effect is an artifact of the subjects and methodology used. Thus it seems that full vowels are still sufficient to describe hyperarticulated, full, and reduced vowels.

D. H. WHALEN
[whalen@haskins.yale.edu]

HARRIET S. MAGEN
[hmagen@ric.edu]

MARIANNE POUPLIER
[marianne.pouplier@yale.edu]

A. MIN KANG
[min.kang@yale.edu]

KHALIL ISKAROUS
[iskarous@haskins.yale.edu]

Bichakjian revisited

May 4, 2004

To the Editor:

While I am reluctant to get involved in a debate whose fine points I have not studied at length, I think it important, in light of recent discussion on evolutionary forces in language (e.g. Frederick Newmeyer in *Language* 79.583–99 (2003) and Bernard Bichakjian in *Language* 80.2–3 (2004)), to make a few general points about an 'evolutionary approach' of language development and our understanding of language lineages:

1. If the 'evolutionary approach' means that linguistic changes 'on the whole have proceeded along a unidirectional course' and 'were probably driven by a Darwinian-like process whereby linguistic features . . . were continuously replaced with ever-more advantageous alternatives' (Bichakjian, p. 2), then we need to be told why the Indo-European languages exhibit widely diverse phonetological, morphological, and syntactic systems that stand in conflict with 'a unidirectional course'. While changes are advantageous in some neurological, auditory, or communicational sense (cf. my *Essays on time-based linguistic analysis*, Oxford University Press, 1996, p. 294), being advantageous does not necessarily trigger a change. How does one get around the problem that simplifying (however conceptualized) a sound segment, a morphological construct, or a syntactic pattern may simultaneously make an item in the context more complex and perhaps less advantageous? A change may violate the connatural implicational relationships between the item changed and some item in the context. Is being advantageous in sound systems based on 'ease of articulation' or on making an item fit more easily into a given environment? Complexity, which can be measured, can be more or less advantageous than something else. Epenthesis is evidently advantageous in one respect but more complex in a purely quantitative sense. Conversely, a deletion may leave incompatible items juxtaposed.