Editor's reply: Such historical perspectives are always welcome. More on Pedersen, elected an honorary member of the LSA in 1930, can be found in the 'Notes' of *Language* 30.194 (1954), where Bernard Bloch records his death on Sunday, October 25, 1953. Details on Pedersen's biography and bibliography are contained in Carol Henriksen's entry in *Lexicon grammaticorum* (ed. by H. Stammerjohann, Niemeyer, 1996, p. 710) and in Henning Andersen's in *The encyclopedia of language and linguistics* (ed. by R. E. Asher and J. M. Y. Simpson, Pergamon, 1994, vol. 6, pp. 2997–98). Pedersen's important contributions to Armenian studies are given an appreciation by Rüdiger Schmitt in his 'Einführung' leading off the reprint volume *Holger Pedersen: Kleine Schriften zum Armenischen* (Olms, 1982, pp. vii–xviii, containing 336 pages of articles dating from 1900 to 1924), and the luminous obituary in *Études celtiques* 7.244–45 (1955–56), written by no less a personage than J. Vendryes, offers additional, and well-deserved, praise.

Editor's note: The following letter is in response to the Letter to *Language* by D. H. Whalen, Harriet S. Magen, Marianne Pouplier, A. Min Kang, and Khalil Iskarous, 'Vowel targets without a hyperspace effect', *Language* 80.3.377–78, 2004 (hereafter, Whalen et al. 2004b); in that letter the authors discussed the results of their article, 'Vowel production and perception: Hyperarticulation without a hyperspace effect', *Language and Speech* 47.155–74 (hereafter, Whalen et al. 2004a); both are responses to Keith Johnson, Edward Flemming, and Richard Wright, 'The hyperspace effect: Phonetic targets are hyperarticulated', *Language* 69.505–28, 1993 (hereafter, JFW). Note that details of references are given only once, whether in this introductory paragraph, the letter, or the response.

Response to Whalen et al.

September 20, 2004

To the Editor:

There seems to be a basic misunderstanding of our point. Whalen et al. (2004b) attribute to us (JFW) the claim that vowel targets in speech production are 'more extreme than their speakers produced even when hyperarticulating' (377) so that 'no production, even the hyperarticulated ones, would ever live up to the hyperspace ideal' (378). Although this is an approximately accurate description of our perceptual 'method of adjustment' results (and also Whalen et al.'s 2004a results, as we show below), we were quite cautious when discussing the implications of our perceptual results. We described our results as 'consistent with hyperarticulated speech' (JFW, p. 505) and 'similar to hyperarticulated speech' (507), and noted that the outcome 'corresponds to hyperarticulated speech' (509) and 'reflects hyperarticulated versions' (519). We adopted this cautious stance because we assumed that 'listeners' perceptual expectations are based on experience' (516), and we felt that average vowel formants in a laboratory speaking task may not capture the full extent of the hyperarticulated vowel space. We noted that 'all of the formant values chosen in the perception task were represented in the productions of at least one speaker' (539). For us then, 'hyperspace' is shorthand for 'hyperarticulated vowel space' as observed in actual speech production, and our conclusion was, and remains, that 'phonetic targets are hyperarticulated' (505, 507, 523–25). Whalen and colleagues seem to think that we were trying to establish, on the basis of the perceptual hyperspace effect, that speech production targets are so extreme as to be unattainable by talkers, a view that is consistent with C. P. Browman & L. Goldstein's modeling of stop targets as points beyond the walls of the vocal tract ('Towards an articulatory phonology', *Phonology Yrbk.* 3.219–52, 1986), as if the speaker intends to poke the tongue through the alveolar ridge to produce [t], but we explicitly did not claim this.

The hyperspace effect in vowel perception has been replicated so many times[1] that it is widely assumed to be a feature of vowel perception (see also L. Polka & O.-S. Bohn, 2003 ('Asymmetries in vowel perception', *Speech Commun.* 41.221–31) for an insightful discussion of a psychoacoustic mechanism that may contribute to the effect). What is more, despite their claims to the contrary, Whalen et al. (2004a) also replicated the perceptual hyperspace effect. In most previous studies, the range of formant frequencies produced by speakers matched the range chosen by listeners. This seems to have not been the case in Whalen et al.'s study. So, while JFW were able to plot perceptual vowel spaces and production vowel spaces together on a single graph and directly observe the perceptual vowel space expansion that we dubbed the 'hyperspace effect', the vowel space expansion in Whalen et al.'s study was partially obscured by a perception/production mismatch in vowel formant range. So the direct statistical comparison of vowel formant frequencies that we and others have performed led to a confusing pattern of production-perception differences characterized by Whalen et al. as 'fairly randomly distributed'.

Therefore, we present here a method for measuring vowel space area that permits comparisons when vowel formant ranges are mismatched. Using this method we then offer a reanalysis of results from JFW and the Whalen et al. (2004a) study showing that both data sets exhibit the hyperspace effect.

The area of the acoustic vowel space can be calculated using the formula for the area of a polygon in the F1/F2 plane in 1. We limit 1 to the point vowels /i/, /æ/, /ɑ/, and /u/, but the full set of vowels can be included so long as they are entered in an order that traces a polygon in the F1/F2 space.

(1)
$$\begin{aligned} A = {} & \frac{1}{2}(F1_i F2_u - F1_u F2_i) \\ & + \frac{1}{2}(F1_u F2_ɑ - F1_ɑ F2_u) \\ & + \frac{1}{2}(F1_ɑ F2_æ - F1_æ F2_ɑ) \\ & + \frac{1}{2}(F1_æ F2_i - F1_i F2_æ) \end{aligned}$$

In the analyses presented here formant frequencies were expressed on three different frequency scales—acoustic frequency, measured in cycles per second (Hz); auditory frequency, measured in auditory critical bands (Bark); and normalized frequency, measured on a log frequency scale using T. M. Nearey's (1977; *Phonetic feature systems for vowels*, U. Conn. dissertation) constant log interval point normalization method (CLIH-2).

To illustrate the differences between these measures of vowel space area (and confirm the utility of this approach), we measured the vowel space areas of the men, women, and children in G. Peterson and H. L. Barney's study of American English vowels ('Control methods used in the study of vowels', *J. Acoust. Soc. of Am.* 24.175–84, 1952).[2] These vowel space area estimates are shown in Table 1, presented at the end of the letter section (p. 649).

The acoustic vowel space area shows very large differences between men, women, and children that reflect differences in formant range (men have the lowest formant range, and children have the highest), while on the auditory and normalized frequency scales the difference between women's and children's vowel spaces are eliminated and both remain larger than the vowel space area found for men. The difference between men and women (larger vowel space for women) is consistent with prior research on other phonetic differences between men and women in the United States (e.g. D. Byrd, 'Relations of sex and dialect to reduction', *Speech Commun.* 15.39–54, 1994), which suggests that the gender differences that remain in the auditory and normalized vowel space areas reflect real gender differences in vowel production.

Similar calculations of vowel space area using the formant frequency data reported by Whalen et al. (2004a) and JFW are shown in Table 2 (note that all of these vowel spaces are for male voices). The area of the perceptual vowel space in both studies was comparable to the area encompassed in hyperarticulated speech and quite noticeably different from the vowel space area found in citation speech. This is true of the auditory and normalized vowel spaces as well as of the acoustic vowel space. The areas of the auditory and normalized vowel spaces support one observation of Whalen et al.—because these frequency scales place greater weight on low frequencies and the Rhode Island speakers produced backer/rounder [u] (with lower F2) than the JFW California speakers, the Whalen et al. hyperarticulated vowel space is larger in auditory and normalized space than is the JFW hyperarticulated vowel space.

From this reanalysis of the Whalen et al. and JFW vowel formant frequency data, we conclude that Whalen et al.'s listeners exhibited the perceptual hyperspace effect just as have listeners in a variety of other studies. This supports the notion that listeners expect vowels to sound more peripheral than they are in normal speech, and perhaps this means that when these listeners speak their phonetic targets are hyperarticulated.

Keith Johnson
[kjohnson@ling.ohio-state.edu]
Edward Flemming
[flemming@stanford.edu]
Richard Wright
[rawright@u.washington.edu]

Notes

[1] S. Lively, 1993 ('An examination of the perceptual magnet effect', *J. Acoust. Soc. of Am.* 93.2423); P. Iverson & P. K. Kuhl, 1996 ('Influences of phonetic identification and category goodness . . .', *J. Acoust. Soc. of Am.* 99.1130–40); A. Lotto, K. R. Kluender & L. L. Holt, 1995 ('Animal and computational models of development of graded vowel categories', *J. Acoust. Soc. of Am.* 98.2965); A. R. Bradlow, 1996 ('A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and in Spanish', *Phonetica* 53.55–85); E. Diesch, P. Iverson, A. Ketterman & C. Siebert, 1999 ('Measuring the perceptual magnet effect. . .', *Psychological Research* 62.1–19); K. Johnson, 2000 ('Adaptive dispersion in vowel perception', *Phonetica* 57.181–88); M. Barkat-Defradas, J.-E. Al-Tamimi & T. Benkirane, 2003 ('Phonetic variation in production and perception of speech', *ICPhS Barcelona*, 857–60); P. Iverson & B. G. Evans, 2003 ('A goodness optimization method for investigating phonetic categorization', *ICPhS Barcelona*, 2217–20); D. Mücke, 2003 ('Toward an auditory reference system for primary vowel types', *ICPhS Barcelona*, 997–1000); M. A. Kiliç & F. Öğüt, 2004 ('A high unrounded vowel in Turkish', *Speech Commun.* 43.143–54).

[2] To be explicit, we took the average formant values for /i/, /æ/, /ɑ/, and /u/ from Peterson & Barney 1952 and for the acoustic vowel space area calculation entered them directly in 1. The auditory vowel space areas were calculated by converting the acoustic frequency measurements to the auditory Bark scale before entering them into 1. The normalized vowel formants were taken by subtracting the average log(F1) from the vowel's log(F1), and the average log(F2) from the vowel's log(F2). Note though that the vowel space area is the same whether or not the log(F) values are normalized by subtracting the average log(F). These normalized values were then entered into 1.

Whalen et al. reply: Johnson, Flemming, and Wright (JFW) continue to misinterpret the main feature of their own 1993 findings and our failure to replicate (2004a,b): Listeners choose more extreme values of synthetic vowels as best exemplars than might be expected because synthetic vowels are not optimal realizations. Thus this effect, to the extent that it replicates, is an experimental artifact. This will remain true even if the newly proposed measuring technique of JFW is used.[1] Identification of steady-state vowels is not terribly accurate. For example, when the values for English obtained in a large study (Peterson & Barney 1952) were resynthesized as steady-state vowels like these used by

JFW, correct identification was only 72.7% (J. Hillenbrand & R. T. Gayvert, 'Identification of steady-state vowels synthesized from the Peterson and Barney measurements', *J. Acoust. Soc. of Am.* 94.668–74, 1993), compared with 95.4% for natural tokens (J. Hillenbrand, L. A. Getty, M. J. Clark & K. Wheeler, 'Acoustic characteristics of American English vowels', *J. Acoust. Soc. of Am.* 97.3099–111, 1995). When we consider that quality ratings are even more variable than identification, we would expect listeners to pick more extreme values in synthesis than they would use in their production, allowing the least ambiguous pattern to be selected. This supports our original conclusion: Any appearance of a hyperspace effect is a methodological artifact.

The theoretical implication that 'perhaps this means that when these listeners speak their phonetic targets are hyperarticulated' (see above) is incoherent because the vowel spaces that are used to determine the perceptual space's extreme values is a generic one, not any one speaker's. The female speakers' space is certainly not adequately represented, because their space is larger than the perceptual one chosen.[2] With the new measure, even the location of the 'expanded' space within the F1/F2 plane is deemed to be irrelevant. How can such an abstract measure be the basis of anyone's production? There is currently no evidence to support the notion that hyperspace plays a role in production or in the perception of natural speech.

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]

Notes

[1] The low front vowel appears to account for most of the difference in space sizes with this new measure, which is a very weak indication of expansion.

[2] JFW currently seem to be suggesting that the differences that remain in vowel space size are due to choice rather than anatomy. Byrd (1994) does not show a larger vowel space for women, only that they choose less reduced forms of central vowels than men. No one has claimed that children 'choose' to use a larger vowel space, only that their smaller vocal tracts result in a larger vowel space.

	ACOUSTIC (kHz^2)	AUDITORY ($Bark^2$)	NORMALIZED ($logHz^2$)
Men	412	15.7	0.119
Women	682	21.0	0.139
Children	910	22.8	0.134

Table 1. Vowel space areas for men, women, and children calculated from average formant frequencies reported in Peterson & Barney 1952.

		ACOUSTIC (kHz^2)		AUDITORY ($Bark^2$)		NORMALIZED ($logHz^2$)	
		WMPKI	JFW	WMPKI	JFW	WMPKI	JFW
PRODUCTION	citation		310		9.99		0.069
	hyperarticulation	437	440	16.56	13.67	0.130	0.092
PERCEPTION	goodness rating	535		16.66		0.112	
	MOA	629	594	18.43	18.95	0.121	0.130

Table 2. Vowel space areas in acoustic, auditory, and normalized frequency scales from Whalen et al. (WMPKI) and JFW; MOA = method of adjustment.