NUMBER OF POSSIBLE BASIC VOWEL QUALITIES AND THEIR PSYCHOACOUSTICAL DISTANCE MEASURE

Antti Iivonen
University of Helsinki, Department of Phonetics

ABSTRACT

Spatial representation for the psychoacoustical vowel space and vowel resolution are discussed. A simple, but proper spatial approximation for the vowel space of the basic (major) vowel types is an F2/F1 vowel chart with Bark scales in which the F1 dimension is enlarged 60% in relation to F2. For other vowel types additional parameters are needed. If we want to display the spatial vowel resolution, a good approximation is achieved by a Critical Band Window (CBW-F1=1 Bark sized circle according to the F1 scale).

PSYCHOACOUSTICAL VOWEL SPACE (PVS)

Several suggestions have been presented in the literature for a simulation of listener's *psychoacoustical vowel space (PVS).* Psychoacoustical scales used include musical, Koenig, full logarithmic, mel, and Bark scale. The most usual parameter combination applied is F2/F1, but additionally several modifications of the F2/F1 space have been suggested: F2'/F1, F2–F1/F1, F2–F1/F1–F0. Some suggestions involve a three-dimensional vowel space (F1/F2/F3).

A successful spatial simulation of a PVS implies that every equal spatial distance corresponds to an equidistant psychoacoustical distance. This issue concerns the *scale problem.* The perceptual role of F3 in some vowel types should be discussed. Some parameter combinations (e.g. F2–F1/F1–F0) involve the question, whether computations are really carried out in the perceptual processing. One problem concerns the *vowel resolution.*

THE SCALE PROBLEM

In order to elucidate the scale problem, I made some experiments with the set of vowels proposed by Lindblom [5]. He calculated the first four formant values of 19 "quasi-cardinal vowels" representing psychoacoustically equal quantization steps. Lindblom used among other things the whole spectrum approach and Plomp's auditory distance metrics for calculations. For [i] and [y], Lindblom has the same F2 value probably due to the natural Swedish vowel formants. Many other languages like Finnish, French, and German have separate F2 values for those vowels. It seems to be plausible to add an hypothetical [ɨ] to the set (cf. Fig. 3 below).

The 19 vowels are presented in Fig. 1 according to full Bark scales. They are connected with lines. Equal psychoacoustical distances would imply that the nearest three vowels should in all cases form spatially equilateral triangles. Fig. 1 shows that this is not the case. The major observation is that the distances are on the average longer concerning F2. There are additionally some minor irregularities.

The conclusion is that F1 must be enlarged in relation to F2 in order to get a better spatial representation for (averaged) equal distances. The F1 dimension of an F2–F1/F1 space has been enlarged 100% in relation to F2 in [4] in order to achieve a better correspondence between the display and the phonetic experience. The calculations of the average distance needed for the enlargement in F2/F2 space showed that the proper enlargement should be approx. 60%. The corresponding modification of the vowel space has been carried out in Fig. 2.

Additional criteria for a PVS are the number of qualitative degrees in IPA vowel set. The maximal number of "horizontal" vowels is three which corresponds acoustically to six because of the effect of rounding on F2. The "vertical" dimension is problematic. There are four main degrees in IPA, but the intermediate degrees confuse the interpretation. Fig. 3 (below) seems to suggest that there exists room for 6 horizontal and 5 vertical (psychoacoustical) degrees.

It must be noticed, however, that the influence of F3 has been neglected so far. Its perceptual contribution is that the front vowels become brighter by means of the combined effect of F2 and F3 (cf. the discussion of perceptual integration in [7]).

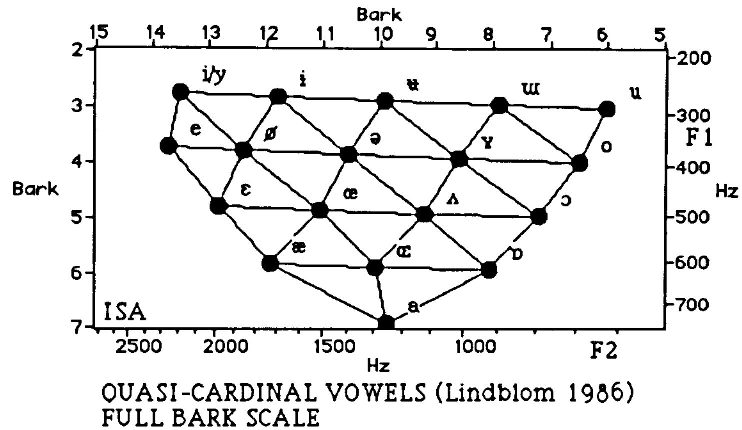

Figure 1. Lindblom's (1986) 19 quasi-cardinal vowels presented in a F2/F1 space according to full Bark scales. (The figures are produced by means of Intelligent Speech Analyser (ISA) developed by Raimo Toivonen.)

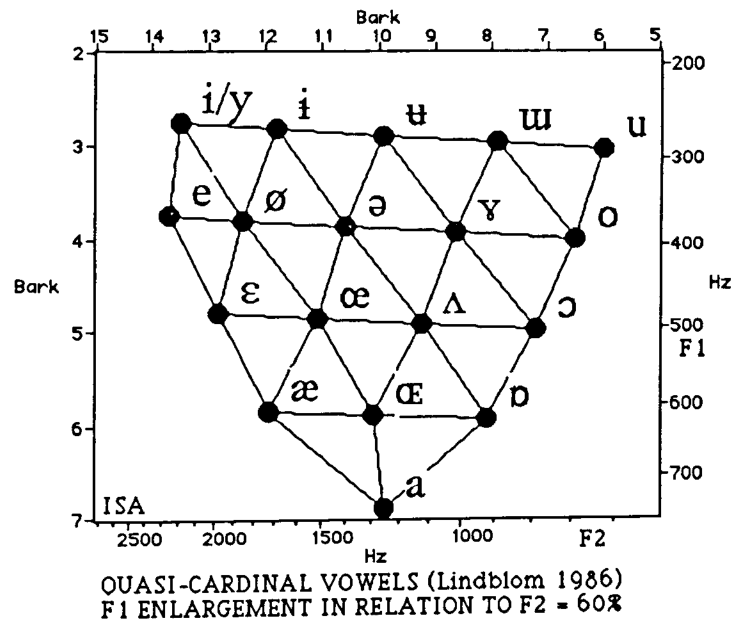

Figure 2. The same vowels as in Fig. 1 presented in an F2/F1 space according to full Bark scales with the enlarged F1 dimension After the enlargement (60%). the spatial distances correspond on the average better the psychoacoustical equi-distancies between the vowel points.

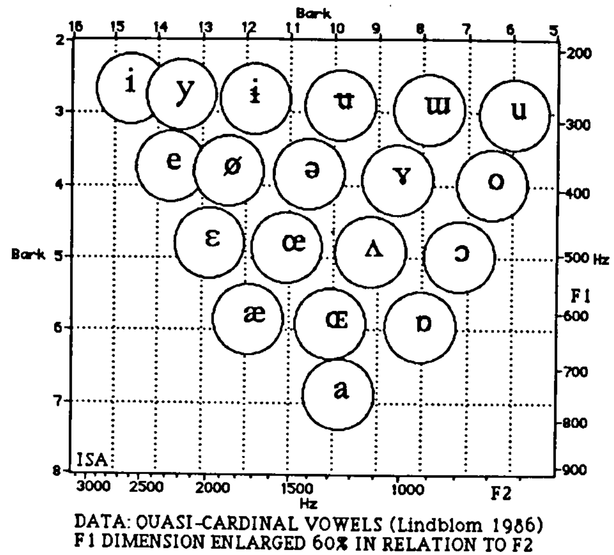

Figure 3. The same vowel set as in Fig. 1–2 presented as 1 Bark sized circles (Critical Band Windows according to F1 dimension). A hypothetical [ɨ] added. Note the overlapping of [i] and [e] with [y] and [ø].

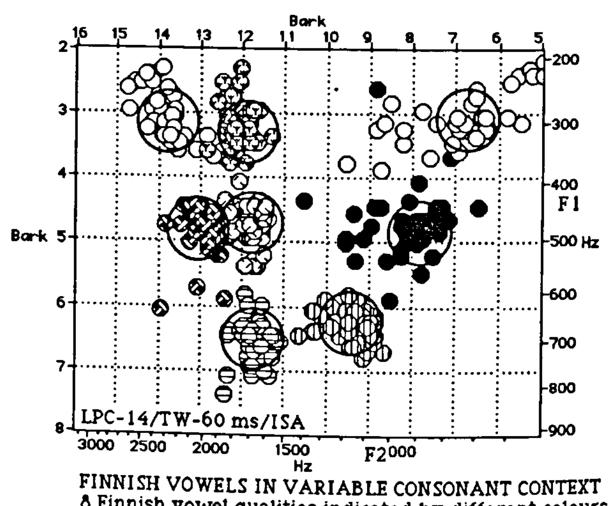

Figure 4. A scattergram of Finnish stressed vowel occurrences (N=352) representing 8 phonemic qualities produced by a male speaker AA in two and three syllable words. The vowels occurred in 11 different symmetrical consonant contexts. A Critical Band Window (CBW-F1) is drawn on the densest accumulation of the single occurrences of each phoneme.

If the F3 is not taken into consideration, the spatial distances between [e] and [ø] (cf. Fig. 1 and 2) as well as between [i] and [y] become too short (these effects can been seen also below in Fig. 3).

The conclusion is that F2/F1 representation (with enlarged F1) can be characterized as an *approximate framework* for basic or *major* vowel types (cf. the notion in [4]), but it cannot be a proper psychoacoustical framework for *all possible* vowel qualities. For [i], [e], [y], and [ø] F2' could be used for display. Besides, the F2/F1 charts must be understood as relative maps, because the absolute vowel positions depend on the vocal tract length. Other articulatory factors should be discussed, too. It is, however, another issue.

PHONETIC VOWEL RESOLUTION

According to Flanagan's experiments [1], the difference limen (DL) for formant perception of *synthetic* vowels is some 3-5%. That small DL would imply that there exist more than 400 perceptually different vowel qualities in the F2/F1 space [2]. The number of possible phonetic major vowel types must be much smaller. Flanagan's result [1] seems to reflect more the *discrimination* of vowels than their phonetic *perception*. Besides, the perception of *natural* vowels is another issue. Nagakawa *et al.* [6] observed 6–13% DLs in the perception of the F2 differences.

Instead of the 3–5% resolution, I suggested [2] that the Critical Band of the ear could be a proper measure for phonetic vowel resolution. If the whole F2/F1 vowel space (with *full Bark* scales) is filled with one-Bark-sized windows (= Critical Band Window, CBW; each comprising one critical band, i.e. 1 Bark), we get about 45 different vowel points [2]. If we fill the vowel space, which has an *enlarged F1* dimension, we get the illustration of Fig. 3. It shows the 19 vowel points discussed above, which comprise 1 Bark each (CBW-F1), according to the F1 scale. A hypothetical vowel [ɨ] has been added. There remains empty space for some additional CB windows. A fronted [a] and a back [ɑ] could be added. The number of CB windows is near that of the IPA vowels (the new IPA vowel chart in JIPA 23 (1) 1993 contains 28 basic vowel symbols). The suggested representation implies that F1 resolution is better than that of F2.

A Critical Band Window (specifically, CBW-F1) can be understood as an area that scans its surrounding space to check, if there exists a perceptually distinct psychoacoustical distance from other vowels. If two CBWs overlap, it can be assumed that the listener may have difficulty distinguishing the vowels considered.

The relationship of speech production and speech perception is a complicated issue. An example of the CBW application is shown in Fig. 4. The scattergram of Finnish vowels and a CBW-F1 on each vowel phoneme show that the major part of single occurrences of a phoneme are covered by a CBW-F1 window, but the distribution is larger especially in /u/ and in /o/ (data from [3]). The [u] and [o] variants to the left of the CBW-F1 represent mainly short vowels and dental contexts.

ACKNOWLEDGMENT

I am grateful to Eli Fischer-Jørgensen for inspiring discussions.

REFERENCES

[1] Flanagan, J. L. (1955) A difference limen for vowel formant frequency. *JASA*, vol. 27, pp. 613-617.
[2] Iivonen, A. (1987) The Critical Band in the Explanation of the Number of Possible Vowels and Psychoacoustical Vowel Distances. Mimeogr. Series of the Dept. of Phonetics, Univ. of Helsinki 12.
[3] Iivonen, A. & Laukkanen, A.-M. (1993) Explanations of the qualitative variation of Finnish vowels. *Publications of the Department of Phonetics, University of Helsinki*, Series B: 5, 29–54.
[4] Ladefoged, P. & Maddieson, I. (1990) Vowels of the world's languages. *J. Phonetics*, vol.*18*, pp. 93–122.
[5] Lindblom, B. (1986) Phonetic universals in vowel systems. In Ohala, J. J. & Jaeger, J. J. (eds.), *Experimental Phonology*, pp. 13–44. Orlando: Academic Press.
[6] Nagakawa, T. Saito, T. & Yoshino, T. (1982) Tonal difference limens for second formant frequencies of synthesized Japanese vowels. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, vol. 16, pp. 81–88.
[7] Schwartz, J.-J., Beautemps, D., Abry, C. & Escudier, P. (1993) Inter-individual and cross-linguistic strategies for the production of the [i] *vs.* [y] contrast. *J. Phonetics*, vol. 21, pp. 411–425.