Proceedings of The Institute of Acoustics

PROBLEMS OF NORMALIZING THE SPECTRAL EFFECTS OF VARIATIONS IN THE FUNDAMENTAL

R.A.W. BLADON

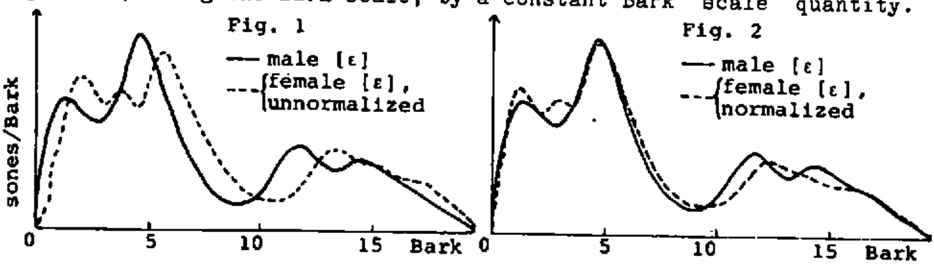
PHONETICS LABORATORY, UNIVERSITY OF OXFORD

A. INTRODUCTION

The observations in these pages are made against the background of an auditory theory of speaker normalization presented in Bladon, Henton and Pickering (1982) (henceforth BHP), a paper as yet unpublished. The relevant concepts, assumptions and findings from that work can profitably be reviewed first.

The question that BHP addressed was: how can we model the listener's normalization process whereby a male vowel and a female vowel may be judged as having 'the same' quality (even in fine phonetic detail), whilst our analysis equipment reveals substantial, apparently nonlinear acoustic differences, especially the well-known ones of formant frequency? As a baseline for developing their normalization ideas, those authors drew on the 32-year-old reasoning by Potter and Steinberg (1950) that "within limits, a certain spatial pattern of stimulation along the basilar membrane may be identified as a given sound regardless of position along the membrane." Now, BHP already had available to them a model of the auditory analysis of vowels in very much these terms, namely as patterns of excitation on the basilar membrane (Bladon & Lindblom 1981). The effect of this model is to subject an input vowel stimulus to a series of transformations reflecting the ear's nonlinear analysis of frequency and of intensity as well as its masking characteristics. Examples of two vowels treated in this way, namely as pseudo-auditory spectra calibrated in sones/Bark versus Bark units, can be seen in Fig. 1. The male vowel actually represents a spectrum averaged across 5 male speakers of Middle Northern English; the hatched spectrum is 'the same' vowel spoken by one female speaker.

BHP postulated that male/female vowel differences such as these can be normalized by a linear displacement of a vowel's auditory spectrum, along the Bark scale, by a constant Bark scale quantity.



Figs. 1 and 2. Auditory spectra, showing effects of normalization.

Fig. 2 shows the result of displacing the female vowel of Fig. 1 downward in frequency by 1 Bark: as can be seen, a good congruence of spectral shapes is achieved. By applying a distance metric to the spectral area differences, BHP determined experimentally that, for this dialect, an average shift of 1 Bark was appropriate. This finding is supported by examining other published acoustic data: e.g. calculations on Peterson & Barney (1952) yield 0.88 Bark, and on Fant (1959) 0.97 Bark. On other material, interestingly, the male/female normalizing displacement was found to be markedly different from 1 Bark: included here are our own RP English data (in excess of 1 Bark), and Dutch (Koopmans-van Beinum 1973) which averages about 0.6 Bark. To explain these striking findings, BHP argued for a learned, socially motivated, sex-linked characteristic.

B. FREQUENCY OF THE FUNDAMENTAL

The notion of a uniform displacement of a spectrum along the Bark scale, by whatever amount, provides only a first approximation to the normalization processes. A complicated nonlinearity arises from the phenomenon of interaction between the frequency of the fundamental, Fo, and the rest of the spectrum (especially F1). The general problem is summarized by Florén (1979:14) as: "When the fundamental frequency of a given vowel is increased, the timbre of the vowel will change even though the formant frequencies remain the same. To neutralize this effect demands an increase of the formant frequencies."

Two hypotheses from the recent literature have attempted to quantify this effect. Firstly, Fujisaki & Kawashima, from their study of Japanese vowel production, concluded that formant frequencies increased linearly with Fo: "The relation between Fo and the nth formant frequency Fn can be generally approximated by a linear equation" (1968:74). Their data for Japanese /e/ have been replotted in Fig. 3 so as to show the normalizing displacement implicit for

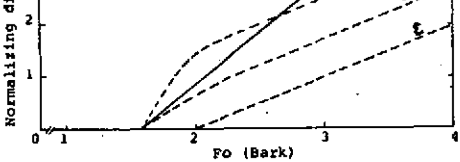


Fig. 3. Predictions of normalizing displacement in the F1 region, derived from authors shown.

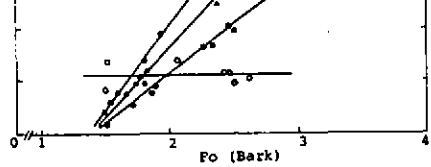


Fig. 4. Normalizing displacement required for female vowel data of this study. Three vowels show good fit to harmonically related lines.

the F1 region of the spectrum, as Fo increases. The linear dependence is apparent. A similar view of the Fo-dependence of F1 is suggested in Carlson, Fant & Granström (1975).

A second theory is that of Traunmüller (1981) who argues that, as Fo increases, F1 frequency increases linearly with it, maintaining a constant F1-Fo difference; and that judged vowel openness depends on the size of that difference, the most open vowels having the largest F1-Fo. What values will this predict for our Bark scale normalizing displacements? They are shown, also in Fig. 3, as vowel-specific lines which are approximately parallel.

However, neither of these theories can account for our data, seen in Fig. 4. These comprise RP English speakers of both sexes who each produced 11 vowels monophthongally on a falling pitch. The figure illustrates by how much the F1 region of the spectrum must be displaced, on a Bark scale, in the individual female vowels shown, to achieve coincidence with the average male F1 region. (The control measures taken, and the measurement procedures, have been detailed in BHP.) Clearly, there is not the uniform linear relationship proposed by Traunmüller or by Fujisaki & Kawashima. A markedly distinct behaviour is apparent in [ɑ] (also shown by [æ] and [ʌ], not plotted here) indicating a substantially constant F1 independent of any variation in Fo. Moreover, the remaining vowels differ in their slopes. Indeed, it turns out that these slopes reflect the trajectories of different harmonics: [ɜ] with H4, [e] with H3 and [i] with H2. It seems that, under appropriate circumstances, we need to normalize for a physical F1 peak which undergoes a shift in frequency following the nearest harmonic. These requirements can be met by proposing a harmonic-dependent normalizing displacement of F1 x (Fo-1.4)/Fo Bark. Values given by that expression are shown in Fig. 4 as unbroken lines. A good match with the observed data is obtained.

Why should open vowels such as [ɑ] not show this behaviour? Acoustic and perceptual explanations can both be advanced: acoustically, these vowels have a high F1/Fo ratio ensuring that several harmonics normally fall within a formant's bandwidth - thus their F1, to maintain amplitude when harmonics shift with Fo, does not need to shift in tandem. Then perceptually, these are the vowels whose F1 and Fo are most consistently separated by more than the 'critical distance' (Chistovich et al. 1979) of 3.5 Bark, implying separate auditory-spectral analysis of F1 and Fo: the two peaks are guaranteed perceptual independence.

C. AMPLITUDE OF THE FUNDAMENTAL

In attempting to quantify the results of modelling a given displacement of a vowel along the Bark scale, our procedure has been to apply a distance metric (Bladon & Lindblom 1981) to a pair of spectral shapes, one of them a reference vowel. Such distance measures are substantially affected by differences in the amplitude of the fundamental, Ao, as can be appreciated from Fig. 5. This shows auditory spectra for a female vowel [ɛ] to whose physical spectrum increments of 5, 10 and 15 dB have been applied. Emphasizing as it does the importance of low frequencies to the total loudness of a sound, this kind of representation presents a potential problem.

First we should verify whether Ao variations of this order are realistic. Of this there seems no doubt. Fig. 6 displays Ao for ten RP speakers' open vowels, expressed in dB relative to F1 level and to H2 level. Across speakers, a range of 10 to 15 dB is certainly attested (cf. likewise Sundberg & Gauffin 1982). Interestingly, there is a clear tendency in this sample for the females to have the higher and the males the lower Ao values. Thus Ao normalization may be another requirement in male/female spectral normalization.

Next, what are the perceptual implications of Ao differences? Sundberg & Gauffin (1982) succeeded in obtaining some judgements of a lowered F1, when Ao was increased by 10 dB. But these judgements appear to have been a small minority, made on very high Fo vowels, and never occurring with open vowels. We may conclude that Ao variations, at conversational pitches, do not significantly affect perceived vowel quality in languages like English and Swedish. Noteworthy also is the remark by Sundberg & Gauffin that, when presented with high Ao vowels in succession, their subjects rapidly

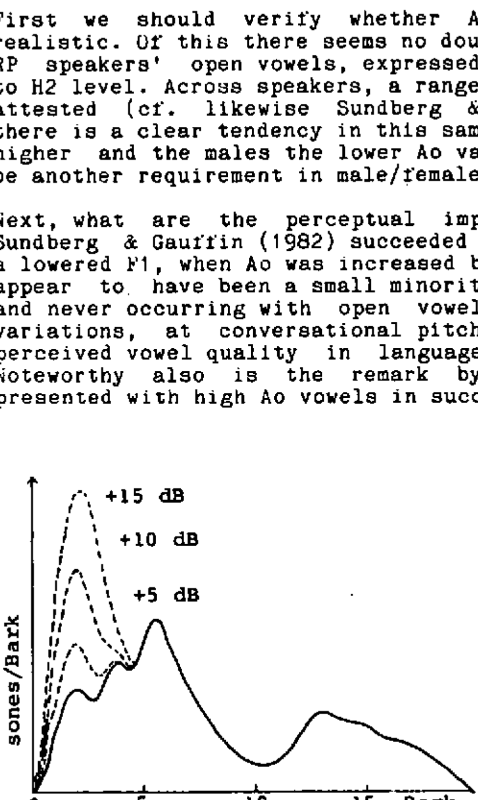


Fig. 5 (above) The effect upon an auditory spectrum of increases in amplitude of the fundamental.

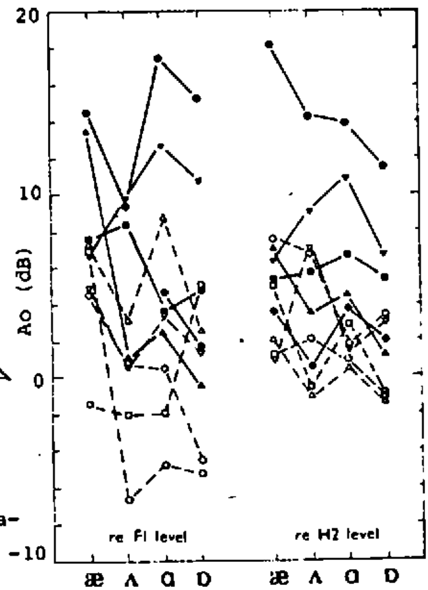


Fig. 6 (right) Observed range of variation in amplitude of the fundamental, Ao, in 10 speakers.

factored out the high Ao, and vowel quality was hardly ever judged to shift.

Thus it seems that our models should normalize Ao variations, since the perceptual system compensates for them. There is one exception however: a high amplitude fundamental has been shown to be a systematic correlate, both in production and perception terms, of a breathy-voiced vowel, as used contrastively in numerous languages (Bickley 1982). To some extent the same is true also of high amplitude sub-F1 harmonics above Fo. The ear can apparently attend to this spectral information when required. When it is not required, the perceptual effects can presumably be modelled as a downward weighting of low-frequency amplitudes which exceed a certain spectral slope. Determination of the optimum weighting remains to be carried out.

REFERENCES

C. BICKLEY 1982 ASA Conference Paper. Acoustic analysis and perception of breathy vowels.
R.A.W. BLADON and B. LINDBLOM 1981 JASA 69, 1414-1422. Modeling the judgement of vowel quality differences.
R. CARLSON, G. FANT and B. GRANSTRÖM 1975 in Fant and Tatham (eds.) *Auditory Analysis and Perception of Speech* (Academic, London), 55-82. Two-formant models, pitch and vowel perception.
L.A. CHISTOVICH and V.V. LUBLINSKAYA 1979 Hearing Research, 1, 185-195. The Center of gravity effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli.
G. FANT 1975 RIT STL-QPSR 2-3/1972, 28-52. Non-uniform vowel normalization.
A. FLORÉN 1980 Perilus, no.1. Why does [ɑ] change to [ɔ] when FO is increased?
H. FUJISAKI and T. KAWASHIMA 1968 IEEE Trans.Audio Electroacoust., AU-16 (1), 73-77. The roles of pitch and higher formants in the perception of vowels.
F. KOOPMANS-VAN BEINUM 1973 J.Phon. 1, 249-261. Comparative phonetic vowel analysis.
G.E. PETERSON and H. BARNEY 1952 JASA 24, 175-184. Control methods used in a study of the vowels.
R.K. POTTER and J.C. STEINBERG 1950 JASA 22, 807-820. Towards the specification of speech.
J. SUNDBERG and J. GAUFFIN 1982 in Carlson and Granstrom (eds.) The Representation of Speech in the Peripheral Auditory System (Elsevier), 223-240. Amplitude of the voice source fundamental and the intelligibility of super pitch vowels.
H. TRAUNMÜLLER 1981 JASA 69, 1465-1474. Perceptual dimension of openness in vowels.