MECHANISMS OF VOWEL PERCEPTION: EVIDENCE FROM STEP VOWELS

F. GOODING

Dept. of Linguistics, Univ. of Wales, Bangor

ABSTRACT

"Step tones" were constructed in which a series of equal intensity harmonics substituted for the upper formants of synthetic English front vowels. The number of HF harmonics, and thus the lower "edge" of the HF step was varied, along with the relative level of the HF step. Since it was already known from previous studies that step tones were perceived as vowels, the present experiments were designed to explore systematically the effects of edge frequency and level in ordsr to determine the role of these attributes in determining vowel quality.

1. INTRODUCTION

Earlier work by the author has provided evidence on vowel perception that cannot be accounted for by traditional formant frequency based theories. First, experiments with two-formant vowels [3] demonstrated that continuous changes in phonetic quality can be achieved by altering relative formant amplitudes. Secondly, a wide range of highly recognizable vowel qualities can be be elicited by stimuli with no formant peaks [4,5]. The stimuli used in the latter experiments had auditorily flat spectra: a single LF series of loudness-equalized harmonics ('step') eliciting back vowels, and both low and high frequency steps together eliciting central and front vowels. For a given LF step, phonetic quality was dependent upon both the LF edge of the HF step and the relative LF/HF step amplitudes. In general, the findings with single step tones [4] were similar to those with single formant vowels, in that a large range of satisfactory back vowels were produced, while the range and naturalness of the front vowels were considerably less (indeed, it has been found that [i] was the only front vowel that could be elicited by single formant stimuli [2]).

The results with two-step stimuli [5] showed that highly identifiable, natural sounding vowel qualities, including the troublesome front vowels, can be elicited by such stimuli. In addition, the relative amplitude of the low and high frequency steps in these stimuli contributed to vowel quality in a fashion analogous to that of the formant amplitudes in two formant vowels of [3].

2. PRESENT STUDY

2.1 Aim and Rationale

The primary aim of the present study was to provide more evidence that would help to choose between alternative explantions of the earlier data. To this end, it was decided to examine more closely the role of the frequency of the lower edge of the HF step in tones with energy in both the F1 and upper formant regions, and any possible interaction of edge frequency with level. Theories to accomodate the earlier data must account for the following: (1) the lack of formant peaks (2) the "edge effect" -- vowel quality dependence on the edge frequencies of the LF and HF steps (3) amplitude dependence -- quality dependence on the relative LF/HF level.

It would seem that two types of theory could account for the edge effect (restricting our attention here to the LF edge of the HF step): (a) the edge hypothesis (EH) -- the *lower* edge frequency is extracted and used directly. (b) a "center of gravity" (CoG) hypothesis (CGH) --- something akin to the local CoG of the whole HF step (or upper formant region in natural vowels) is taken. Changing LF edge would have the effect of changing the CoG as well. CGH would predict that quality would depend on both the HF as well as the LF edge of the HF step (though to reduce the number of stimuli only the LF edge was manipulated in these experiments. This can still distinguish between the competing hypotheses). CGH would predict that identifications would take place by matching the CoG of the HF step (roughly the mid point on a pitch scale for a flat series of harmonics) against the CoG of the upper formant region in the S's internal reference vowel (perhaps roughly equivalent to F2 prime). In short, EH would predict best identifications with edge frequencies at or slightly below F2, while CGH would predict identifactions with edges well above F2. As a guide, midpoints for the HF steps ranged from 2554 Hz to 3057 Hz when measured on the ERB-rate scale [7].

To account for the amplitude effect, it would seem that a mechanism involving some form *global* spectral balance, or pehaps CoG, is implicated. In the case of the latter, it would involve operation over a distance of greater than the 3.5 Bark limit originally suggested by Chistovich and her colleagues [1].

2.1 Stimuli

The stimuli were produced by digital harmonic syntheis with a sampling rate of 10 KHz, and LP filtered at 4 KHz (filter cutoff rate 180 dB/Octave). All had a fundamental frequency of 125 Hz, and duration was 300 ms, with 20 ms linear onsets and offsets. The stimuli were modified from those in [5] in that in the F1 region a formant appropriate to one of four RP front vowel was substiuted for the LF step of loudness-equalized harmonics. This was done in order to reduce the strong sensation of nasality that accompanied some of the earlier stimuli. It was assumed that this was caused by the apparent broad bandwidth of the F1 region, known to be a associated with the secondary feature of nasality. Four F1 values, 272, 380, 525, and 713, were used, appropriate to the RP phonemes /i/, /I/, / / and /ae/, respectively [6].

The LF edge of the HF step varied from 1750 Hz to 2500 Hz, and the HF limit was fixed at 3750 Hz. This value was chosen because HF energy above this frequency added a fricative-like or whistling sound to some stimuli, which, while distracting, was clearly heard as separate from the vowel. This might be seen as casting doubt on the CGH, since altering the HF edge is thus shown not to effect vowel quality. However, it can be claimed that since the energy above ca. 3750 Hz is not integrated into the vowel percept, this does not constitute a real test of that theory.

HF amplitudes varied in 10 dB steps from 0 to -40 dB for the earlier stimulus sets and from -10 dB to -30 dB for the final

set.

2.3 Procedure

A matching experiment (not reported on here) and an *identification experiment* were carried out. This was computer controlled, the stimuli being presented on-line and responses entered via the keyboard. A total of 22 Ss listened to four different randomized blocks of the stimulus set (ultimately 39 stimuli, though some subjects heard supersets of 52 and 86 stimuli). Randomizations were different for each S. Ss could listen to the stimulus as often as they wished by pressing a key. For each stimulus, Ss were asked to enter a score representing the English (RP) vowel it most resembled, identified by 13 key words shown at the top of the screen and identified by number. They also entered a confidence score (0-9) for their choice.

3. RESULTS

3.1 Edge Effect

The results, interpreted through the use of stimulus and response profiles, clearly support the EH and contradict the CGH. The stimuli most identified as the front vowels were in all cases those with edge frequency close to the F2 of the natural vowel, as predicted by the EH. /i/ unsurprisingly proved to be the most identifiable vowel. /ae/ was the least identifiable, with stimuli designed to elicit it achieving only a 24% score for the first 10 Ss. This is probably due to the lack of the RP value for most Ss in the own speech, in favor of North and Midlands [a]. These stimuli were hence dropped from the final set for the last 12 Ss.

3.2 Amplitude Effect

For a given F1 value and a given response category, identification scores were not always a monotonic function of HF level (though scores for /i/ with F1 = 272 most closely approximate this), but rather showed evidence of a trading relation between edge frequency and level. This needs to be examined more closely, but if verified, it could be taken as evidence for the operation of a *global* spectral balance or CoG mechanism. With the exception of stimuli eliciting the most /i/ responses (F1 at 272 and edge frequency at 2250 and 2500 Hz), virtually all responses for stimuli with levels of -40 dB were back or central vowels. The /i/ stimuli, by contrast, achieved scores over 60% at -40 dB, compared to 90% at -10 dB. Below -20 dB there was a clear shift to central and back vowel responses. The main point is that the responses for the 4 different levels were significantly different, indicating *phonetic* change with level. This is borne out by the judgements of two professional phoneticians to the whole range of stimuli.

Table 1 shows a brief summary of the pooled results with rounded scores. Note that except for /e/ (which is phonetically [ei] in RP) the F1 of the stimuli corresponded to the F1 of the response vowel. No stimuli were designed to elicit /e/, but the stimulus eliciting the most /e/ reponses had an F1 of 380 Hz (appropriate for /I/).

TABLE 1
Summary of Pooled results

Vowel	F2	Edge Freq	Score	Level
/i/	2361	2500	90%	-10 dB
		2250	83%	-10 dB
/I/	2085	2000	42%	-10 dB
/e/	(2000)	2125	39%	-10 dB
/ /	1943	2000	47%	-10 dB

4. DISCUSSION

The implication of the finding of a spectral edge feature in synthetic step vowels for the perception of natural vowels is that F2 in front vowels must serve as marker of the edge of the upper formant region. This feature appears to be used in conjunction with global amplitude information. The evidence reported here is not consistent with a local CoG mechanism operting over upper formant region.

5. REFERENCES

[1]CHISTOVICH, L.A. and LUBLINSKAYA, V.V. (1979), "The 'centre of gravity' effect and the critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli", *Hearing Res.* 1, 185-195.

[2]DELATTRE, P., *et al.* (1952), "An experimental study of the acoustic determinants of vowel color", *Word*, 8, 195-210.

[3]GOODING, F., (1986), "On the role of formant amplitudes in vowel perception", *IEE Conf. Pub. 258*.2, 287-291.

[4]_____ (1986b), "Formantless vowels and theories of vowel percep tion" *JASA* 80:Sup.1, S126

[5]_____ (1988), "Formantless vowels: the next step", in WA Ainsworth and JN Holmes (eds.s) *Speech '88. Proc. 7th FASE Symposium*, Edinburgh, 747-754.

[6]HENTON, C., (1983), "Changes in the vowels of Received Pronunciation", *J. Phon.*, 11, 353-371.

[7]MOORE, B., and GLASBERG, B., (1986), "The role of frequency selectivity in the perception of loudness, pitch and time", in B.Moore, ed., *Frequency Selectivity in Hearing*, Academic, 251-308.