EFFECTS OF FORMANT FREQUENCY MODULATION ON VOWEL IDENTIFICATION

Peter J. Bailey, Kim Bevan and Tracy Burr
Department of Psychology, University of York, York YO1 5DD, U.K.

ABSTRACT

Two experiments are reported concerned with the perceptual effects of modulating the frequency of formants in synthesised vowels. The first experiment showed that modulated vowels tended to be identified in noise more accurately than unmodulated vowels, and that this tendency was more marked for hearing-impaired listeners than for those with normal hearing. The second experiment explored this finding with a different set of vowels and modulation conditions.

INTRODUCTION

Sinusoidal modulation of the centre frequency of a formant-like spectral peak in a periodic signal can lead to a significant reduction in the difference limen for peak frequency, compared to the difference limen for an unmodulated peak. This effect is particularly marked when the modulated spectral peak is partially masked by the presence in the stimulus of a second lower-frequency spectral peak. [1].

We have inferred from this increase in discriminability found for modulated peaks that when the energy that contributes to a spectrum envelope peak is modulated independently of other peaks and background noise, the perceptual salience of the peak is enhanced. We take this to be analogous to the disambiguating consequences of object movement for figure-ground segregation in visual scenes.

One motivation for measuring the effects of spectral peak modulation was an interest in signal-manipulation strategies that might increase the discriminability of speech in noise for listeners with a hearing impairment. Difficulties with speech perception in noisy or reverberant environments are commonly reported by people with sensori-neural hearing impairments, and are probably due in part to the impairments in spectro-temporal resolution that often accompany loss of auditory sensitivity. Amplification can improve recognition of speech in quiet but leads to relatively less benefit for speech in noise [2]. Enhancement of spectral contrast can lead to reliable increases in speech recognition accuracy in noise, but the improvements are modest [3].

It appeared meet, therefore, to explore the efficacy of modulation as an additional enhancement procedure that might be combined with others designed to augment the salience of informationally-rich parts of the speech signal. Here we report tests of the hypothesis that the accuracy of identification of vowels whose discrimination was dependent on resolution of spectral detail might be improved by formant frequency modulation.

EXPERIMENT 1

We have argued that peak frequency modulation increases discriminability by facilitating perceptual segregation of the modulated peak from other spectral peaks. Complete segregation could well turn out to be non-optimal for signals like speech, in which relationships between spectral prominences are important for segment identification. Rather than increasing speech identification accuracy, modulation of formant frequency could be counter-productive if its major effect is to disrupt the perceptual coherence of the formants. However, modulation may prove to be of overall benefit if it can be applied to a formant frequency so that the formant becomes perceptually more salient without there being a concomitant exclusion of the formant from the overall format pattern. In the first experiment we measured identification accuracy for a small set of synthesised vowels presented in noise to normal-hearing and hearing-impaired listeners. The independent variable was the amount of modulation applied to the frequency of the second formant, selected because of its key role in many phonetic constrasts.

Stimuli

Three-formant tokens of the four vowels /a/, /ɔ/, /u/ and /i/ were created using a parallel-formant synthesiser. The synthesis parameters are given in Table 1.

Table 1. Synthesis parameters for the vowels in Experiment 1. Formant frequencies (F) and 3 dB bandwidths (B) are given in Hz, formant amplitudes (A) are given in dB relative to the maximum.

	/a/	/ɔ/	/u/	/i/
F1	750	475	325	285
F2	1100	750	935	2375
F3	2600	2625	2325	3540
B1	65	75	70	80
B2	70	90	110	160
B3	80	140	150	60
A1	0	0	0	0
A2	-10	-13	-20	-20
A3	-15	-40	-45	-30

The vowels were 540 msecs in duration, with amplitude envelopes shaped at onset and offset by 60 msec linear ramps (except for /a/, which had a 90 msec offset ramp). The fundamental frequency was constant at 125 Hz.

Three versions of each vowel were synthesised: the second formant frequency was either unmodulated, or modulated at 5 Hz or at 10 Hz. When modulated the modulation depth was set to give a fixed frequency excursion of 150 Hz centred around the nominal formant frequency.

Subjects

Data were collected from four normal-hearing listeners and three listeners with mild to moderate unilateral hearing impairments.

Procedure

On each trial in the experiment subjects identified a single presentation of one of the vowels by circling one of four alternatives on a response sheet. In each modulation condition (unmodulated, 5 Hz modulation & 10 Hz modulation) subjects were presented with 20 practice trials, during which they heard each of the four vowels 5 times in random order, followed by 40 experimental trials involving a further 10 presentations of each vowel in random order. As far as possible the order of the three conditions was counterbalanced across subjects.

All vowels were presented monaurally via headphones at 70 dB SPL in a continuous broad-band background noise set individually for each ear at a level corresponding to a 10 dB sensation level for the vowels. Noise levels were in the range 71-79 dB SPL for normal-hearing subjects and 60-73 dB SPL for hearing-impaired subjects. Stimuli were presented to the right ear of the normal-hearing subjects. Hearing-impaired subjects ran through the procedure twice, first using their "good" ear, and again in the same session (with a different order of conditions) using their impaired ear. All listening tests were carried out in a sound-attenuating room.

Results

The overall numbers of errors in the three modulation conditions are shown for normal-hearing subjects in Table 2, and for hearing-impaired subjects in Table 3.

Table 2. Normal-hearing subjects: number of errors out of 40 presentations per vowel per condition.

RIGHT EAR	/a/	/ɔ/	/u/	/i/	total
unmodulated	10	17	18	23	68
5 Hz modulation	7	15	14	15	51
10 Hz modulation	8	11	15	15	49

Table 3. Hearing-impaired subjects: number of errors out of 40 presentations per vowel per condition per ear.

"GOOD" EAR	/a/	/ɔ/	/u/	/i/	total
unmodulated	5	6	10	13	34
5 Hz modulation	1	4	6	10	21
10 Hz modulation	1	2	5	11	19

IMPAIRED EAR	/a/	/ɔ/	/u/	/i/	total
unmodulated	7	6	13	11	37
5 Hz modulation	0	0	6	6	12
10 Hz modulation	0	0	4	9	13

Discussion

Taken together the data suggest that: (i) modulation of the frequency of the second formant tends to decrease the error rate below that found for unmodulated vowels, (ii) the effects of modulation, when present, are not strongly dependent on modulation rate, and (iii) to the extent that the perceptual coherence of vowel formants is indexed by identification errors, there is no indication in the data that modulating a

single formant causes it to be excluded from the vowel percept. The trends seen in the means were evident in the data of all subjects, except for one of the hearing-impaired listeners whose mean identification accuracy was apparently not affected by modulation.

A comparison of the data from good and impaired ears of the hearing-impaired subjects suggests that the tendency for modulation to reduce identification errors may be more marked for the impaired ear than for the good ear. The finding that the error rate for hearing-impaired listeners was lower overall than that for normal-hearing listeners is attributable to the technique used for setting the level of the background noise. Noise levels at masked thresholds were lower for hearing-impaired listeners, with the consequence that in the main part of the experiment the spectrum level of the noise was lower relative to those of the spectral peaks in the vowels than was true for the normal-hearing listeners.

It appears that, at least for these vowel stimuli at the unfavourable signal-to-noise ratios we have used and for these subjects, second formant frequency modulation tends to improve vowel identification performance.

EXPERIMENT 2

In an attempt to establish the reliability of the trends demonstrated in the first experiment we carried out a second experiment, using a different and larger set of subjects, vowels and modulation conditions.

Stimuli

Three-formant tokens of the six vowels /ɪ/, /ɑ/, /ɔ/, /ɪ/, /u/ and /ɛ/ were created using a parallel-formant synthesiser. The formant frequencies are given in Table 4.

Table 4. Formant frequencies (Hz) for the vowels in Experiment 2.

	F1	F2	F3
/i/	280	2250	2890
/ɑ/	710	1100	2540
/ɔ/	590	840	2540
/ɪ/	325	1920	2560
/u/	310	870	2250
/ɛ/	550	1770	2490

Formant bandwidths for the first, second and third formants were constant at 50 Hz, 70 Hz and 100 Hz respectively for all vowels, and formant amplitude parameters were set equal for all formants in all vowels. Vowels were 250 msec in duration, with 5 msec linear onset ramps and 25 msec offset ramps. Pitch fell linearly from 130 Hz to 110 Hz through the vowels. The overall levels of the vowels varied by less than 2 dB.

Eight versions of each vowel were synthesised: one was unmodulated, three had one formant modulated (F1, F2 or F3), three had two formants modulated (F1&F2, F2&F3 or F1&F3), and one had all three formants modulated. When modulation was present the modulation rate was 10 Hz and the peak-to-trough modulation depth was 10% of the nominal formant frequency. When more than one formant was modulated all modulations were in phase.

Subjects

Data were collected from 20 normal-hearing listeners.

Procedure

The task for the subjects was similar to that in Experiment 1, except that there were six response alternatives. Each experimental session involved 12 practice trials (two presentations of each unmodulated vowel), followed by 5 blocks of 48 experimental trials. In each block each vowel was presented in each modulation condition once, in random order. The stimuli were presented from tape binaurally over headphones in a continuous broad-band noise set to a sound pressure level 8 dB greater than the peak level for the least intense vowel (/u/). Overall presentation level was set to be comfortable for the subjects (approximately 75 dB SPL). Listening tests were carried out in a quiet (but not sound-attenuating) room.

Results and Discussion.

Overall errors are shown in Table 5 for all vowels and modulation conditions. The data show large differences in the mean errors across vowels, but little systematic trend in the mean errors for the modulation conditions. These observation were confirmed with a two-way within-subjects analysis of variance on the error data which showed a significant main effect of vowel

Table 5. Overall percentages of identification errors for Experiment 2.

conditions	vowel						
(formants modulated)	/i/	/ɑ/	/ɔ/	/ɪ/	/u/	/ɛ/	MEAN
none	18	2	20	56	38	5	23.1
F1	24	1	25	57	42	6	25.8
F2	29	1	19	60	35	5	24.8
F3	26	3	21	49	47	7	25.5
F1&F2	25	3	19	55	40	7	24.8
F2&F3	22	2	18	55	49	6	25.3
F1&F3	20	3	25	50	40	5	23.8
F1&F2&F3	27	0	20	53	48	3	25.2
MEAN	23.9	1.9	20.9	54.4	42.4	5.5	24.8

[$F_{(5,95)}$=20.64, $p<0.01$], but no main effect of modulation condition [$F_{(7,133)}$=0.53] and no interaction between vowel and modulation condition [$F_{(35,665)}$=1.18]. Although formant frequency modulation had no systematic effect on the accuracy of vowel identification, it is noteworthy that for three of the four vowels for which substantial numbers of errors were made, the unmodulated condition did not show the lowest mean error rate. Analyses of the confusions between vowels revealed a strong tendency for /ɪ/ to be heard as /ɛ/ or /i/, and for /u/ to be heard as /ɔ/.

There is one aspect of the modulation conditions involving simultaneous modulation of more than one formant that deserves comment. Given that all modulations were at the same rate and in the same phase, the perceptual coherence of formants accruing from common modulation may have acted in opposition to any perceptual salience that modulation may have conferred on individual modulated formants.

There were several differences between the stimuli and procedures in the two experiments that might have contributed to the differences in outcome. A likely candidate was the differences in spectrum level of the formants relative to the spectrum level of the background noise. This was particularly true for F2 and F3, whose spectrum levels for all vowels were below the spectrum level of the noise in Experiment 1, but above it in Experiment 2. We draw the tentative conclusion, to be examined in further experiments, that effects of formant frequency modulation may only be manifest when the signal-to-noise ratio is particularly unfavourable.

CONCLUSION

The limited set of data reported here suggest that in some, but not all, circumstances, modulation of formant frequency may be able to improve the discriminability of speech contrasts that are dependent on resolution of spectral detail. If the precise conditions under which such improvements are found reliably can be established, formant frequency modulation would seem to have potential as an additional item in the armoury of the hearing-aid designer.

ACKNOWLEDGEMENTS

We are grateful to Mr G. Hope and Mr. A. Grace for their help in locating suitable hearing-impaired subjects, and to the subjects themselves for their cooperation. The UK Science and Medical Research Councils provided financial support.

REFERENCES

[1] Bailey P.J. & Bevan K.M. (1991), "Frequency modulation of formant-like spectral peaks", Proc. XIIth Int. Congr. of Phon. Sci. (Aix-en-Provence), vol. 4, 46-49

[2] Plomp R. (1978) "Auditory handicap of hearing impairment and the limited benefit of hearing aids.", J. Acoust. Soc. Amer., vol. 63, 533-549

[3] Baer T., Moore B. C. J. and Gatehouse S. (1993) "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality and response times.", J. Rehabil. Res. and Devel., vol. 30, 49-72