ISCA Archive
http://www.isca-speech.org/archive

EUROSPEECH '97
5th European Conference on Speech
Communication and Technology
Rhodes, Greece, September 22-25, 1997

PRELIMINARY EXPERIMENTS ON THE PERCEPTION OF DOUBLE SEMIVOWELS

W.A.Ainsworth and G.F.Meyer
Centre for Human and Machine Perception Research
Department of Communication and Neuroscience
Keele University
Keele, Staffordshire ST5 5BG, England
w.a.ainsworth@keele,ac.uk

ABSTRACT

A number of previous studies have shown that it is possible to recognise two vowel sounds spoken simultaneously if their pitches, onsets or the spatial locations of their sources differ sufficiently. Experiments have been carried out to explore the perception of isolated syllables containing the glides /w/ and /j/ spoken at the same time. It was found that consonants in concurrent syllables which contained the same vowel but were spoken with different pitches could not be reliably identified. However if the vowels and their pitches differed, the glides could be recognised about 70% of the time. This suggests that the neuronal mechanisms underlying the separation of simultaneous consonants employ other features as well as pitch differences

1. INTRODUCTION

It has long been known that a conversation can be followed in the presence of another equally loud conversation. This was called the cocktail party effect by Cherry [1]. However the only formal experiments which have been performed to measure the effects of one speech sound on the perception of another speech sound spoken simultaneously have been carried out with pairs of vowel sounds [2, 3, 4, 5, 6]. Real speech consists of sounds with constantly changing spectra so it is of interest to study the effects with formant transitions spoken simultaneously.

Perhaps the simplest of such sounds are the approximants /w/ and /j/ which can be synthesised by three formants rising in frequency for /w/ and a rising first formant and falling second and third formants for /j/ [7, 8]. These were combined with the vowels /i/ and /a/ to produce the syllables /wi/, /wa/, /ji/ and /ja/ and played simultaneously to a number of listeners for identification.

2. STIMULI

The stimuli consisted of 3-formant sounds with a 100 ms segment in which the frequency and amplitude of the formants changed followed by a 100 ms segment during which the frequency and amplitude remained constant. In order to produce a /w/-like sound F1 began at 250 Hz, F2 at 750 Hz and F3 at 1500 Hz. To produce a /j/-like sound F1 again began at 250 Hz but F2 began at 2500 Hz and F3 at 3500 Hz. An /i/-syllable was formed with F1 of the steady segment at 250 Hz, F2 at 2500 Hz and F3 at 3000 Hz, whereas an /a/-syllable was formed with F1 at 900 Hz, F2 at 1100 Hz and F3 at 2500 Hz. The formant tracks for all combinations of these syllables are shown in Figure 1. The four syllables were easily recognisable in isolation.

3. EXPERIMENTS

Eight listeners took part in the experiments. In each experiment the listeners heard two syllables played simultaneously and were asked whether the consonant they heard at the beginning of the syllable was /w/, /j/ or whether they heard both. In each experiment /w/-syllables were added to /j/-syllables in the ratios 1:9, 3:7, 5:5, 7:3 and 9:1 (approximately -20, -7, 0, +7 and +20 dB). Various combinations of vowel (/i/ or /a/) and fundamental frequency (100 or 150 Hz) were employed in the different experiments.

3.1 Same vowel, same pitch

In this experiment /wa/ was combined with /ja/ and //wi/ with /ji/ with both syllables in the pair at

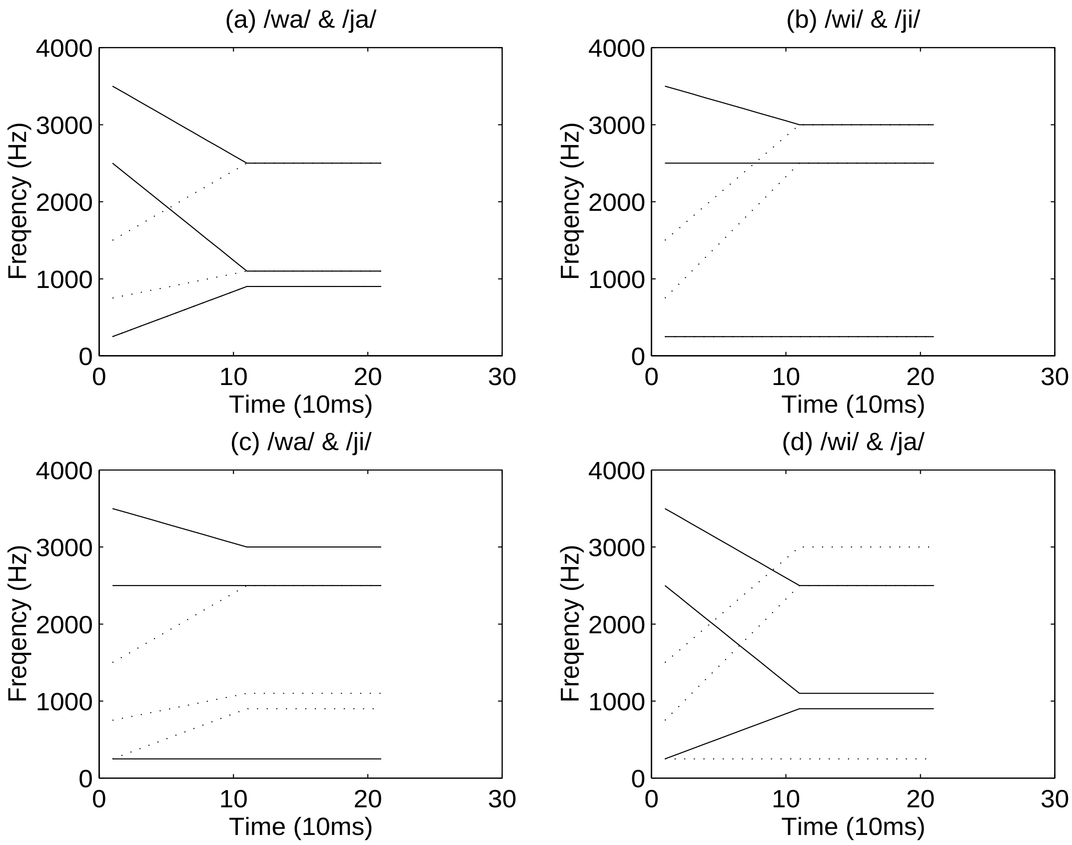

Figure 1. Formant frequency as a function of time for (a) /wa/ & /ja/, (b) /wi/ & /ji/, (c) /wa/ & /ji/ and (d) /wi/ & /ja/. The formant tracks for the /w/ syllables are shown by dotted lines and those for the /j/ syllables by solid lines.

100 or 150 Hz. The results are shown in Figure 2(a). As expected, for ratios of 1:9 and 3:7 more syllables beginning with /j/ were heard and for 7:3 and 9:1 more /w/ syllables. Averaged over all the listeners 15.0% of the stimuli were identified as containing both /w/ and /j/. These were fairly independent of the mixture ratio.

3.2 Same vowel, different pitch

In this experiment the same pairs of syllables were employed except one syllable had a fundamental of 100 Hz and the other one of 150 Hz. The results are shown in Figure 2(b). Some 16.1% of the stimuli were identified as containing both /w/ and /j/. There was a slight peak in the proportion of dual responses when the two syllables were equally intense.

3.3 Different vowel, same pitch

In this experiment /wi/ was combined with /ja/ and/wa/ with /ji/. Both syllables in the pair had a fundamental frequency of either 100 Hz or 150 Hz. The results are shown in Figure 2(c). The proportion of stimuli estimated to contain both consonants rose to 23.5%

3.4 Different vowel, different pitch

The same syllable combinations were employed as in the fourth experiment but one syllable of the pair had a fundamental frequency of 100 Hz and the other one of 150 Hz. The results are shown in Figure 2(d). This condition increased the proportion of stimuli perceived as containing both consonants to 33.3%.

3.5 Different vowel, different pitch again

In experiments 3 and 4, although listeners were asked to decide if both consonants were present, they could have been hearing both of the vowels but not the consonants. Experiment 4 was therefore repeated but the listeners were asked to indicate whether they heard either /wi/ or /wa/ alone, /ji/

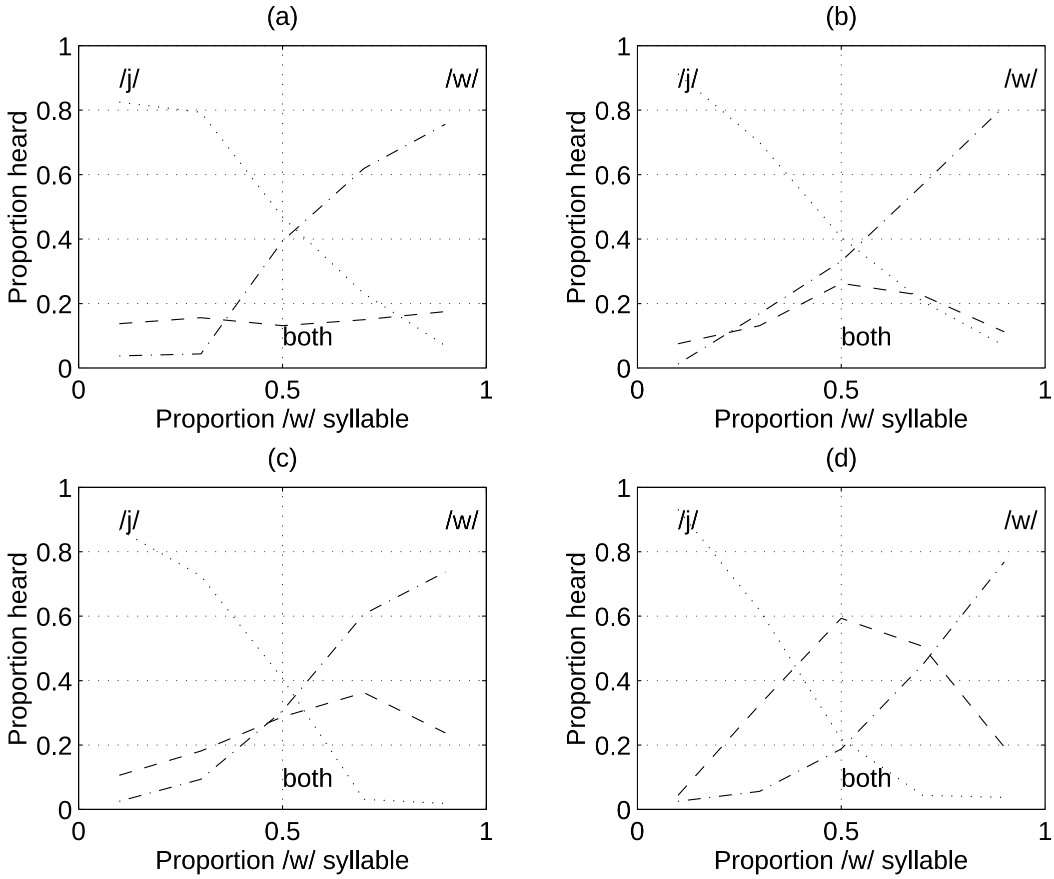

Figure 2. Proportion of /w/ (dot-dash), /j/ (dotted) and both /w/ and /j/ (dashed) syllables heard as a function of the ratio of the intensity of the /w/ syllable to the /j/ syllable for (a) same vowel and pitch, (b) same vowel but different pitch, (c) different vowel but same pitch, and (d) different vowel and pitch.

or /ja/ alone, /wa/ and /ji/ together or /wi/ and /ja/ together. It was thus possible to decide if they were actually hearing the consonants separately and not just the vowels.

The results are shown in Figure 3. Although the incorrect combination of syllables was sometimes heard the correct combination was perceived more often. On average when the listeners perceived two consonants they responded correctly 67% to /wa/ and /ji/ and 70% to //wi/ and /ja/. This shows that they were hearing both the consonants (and vowels) correctly at least some of the time.

4. DISCUSSION

It has been found that when two syllables consisting of the approximants /w/ or /j/ and the vowels /i/ and /a/ are presented simultaneously both consonants can sometimes be heard. If the two syllables contain the same vowel it is difficult to hear both consonants, but some slight advantage is obtained if the two syllables are synthesised with different pitches. If the two syllables have different vowels the two consonants are easier to identify. If the pitches of the two syllables are also different the two consonants are much easier to identify. Even in this condition, however, they are not correctly identified all of the time.

Experiments on the perception of simultaneous vowel sounds have demonstrated that they are easier to segregate if their duration is increased up to 200 ms. This implies some sort of internal analysis window of about this magnitude. The glides themselves were only 100 ms in duration so it might be expected that they would be difficult to segregate.

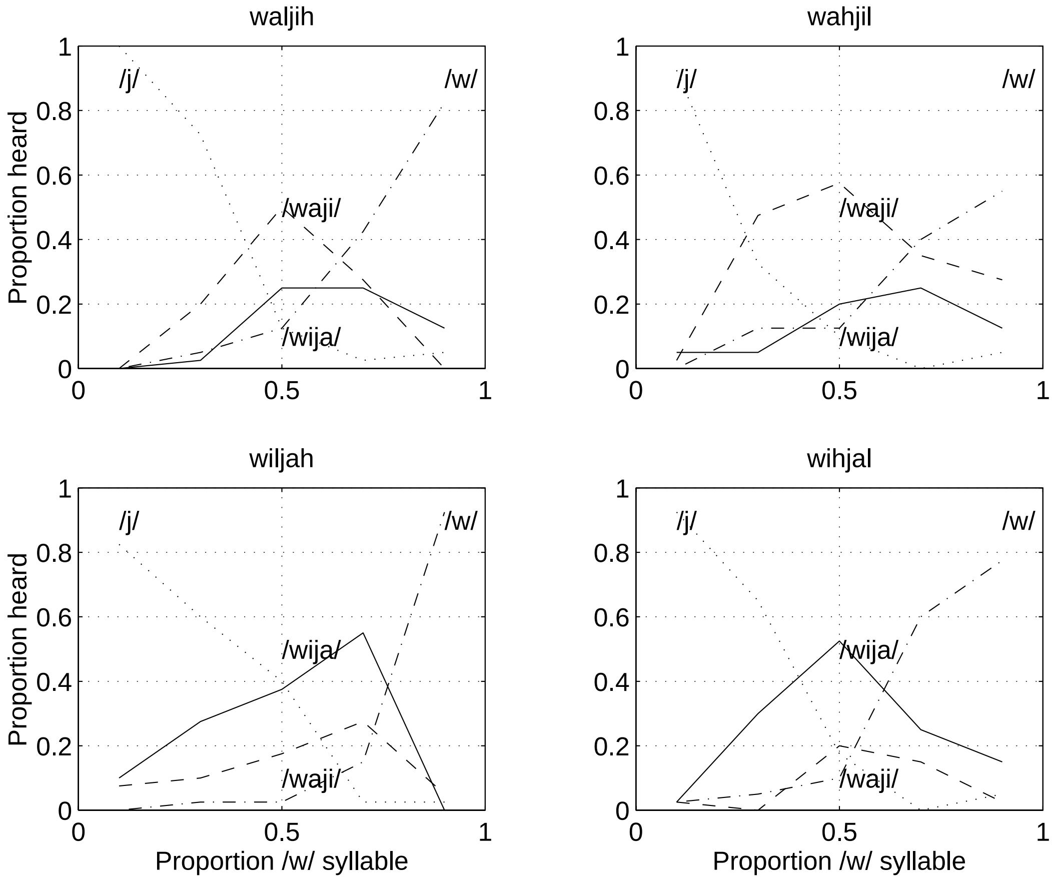

Figure 3. Proportion of /w/ (dot-dash), /j/ (dotted), /wa/-/ji/ (dashed) and /wi/-/ja/ (solid) syllables heard as a function of the ratio of the intensity of the /w/ syllable to the /j/ syllable.

5. CONCLUSIONS

It thus appears that when listening to simultaneous consonants, in order to hear both of them, it is not sufficient for the pitches of the two syllables to be different, as is the case for isolated vowels. It is also necessary for the vowels of the syllables to be different, at least for the restricted case considered here. It is likely that some sequential analysis of the formant transitions takes place which is facilitated if the formants of the two syllables are moving towards different vowel targets. Fortunately when listening to real simultaneous conversations this will almost always be the case.

6. REFERENCES

[1] E.C.Cherry, "Some experiments on the recognition of speech, with one and two ears", *J.Acoust.Soc.Am*. 25, 975-979, 1953.

[2] M.T.M.Sheffers, "Shifting Vowels", Doctoral Dissertation, Groningen University, Netherlands, 1983.

[3] P.F.Assmann and Q.Summerfield, "Modelling the perception of concurrent vowels: vowels with different fundamental frequencies", *J.Acoust.Soc.Am*. 88, 680-697, 1990.

[4] J.F.Culling and C.J.Darwin, "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0", *J.Acoust.Soc.Am*. 93, 3454-3467 1993.

[5] A.de Cheveigné, S.McAdams, J.Laroche and M.Rosenberg "Identification of concurrent harmonic and inharmonic vowels: A test of the theory of harmonic cancellation and enhancement", *J.Acoust.Soc.Am*. 97, 3736-3748, 1995.

[6] F.Berthommier and G.F.Meyer, "*Source separation by a functional model of amplitude demodulation*", Proc.Eurospeech'95, 135-138, 1995.

[7] J.D.O'Connor, L.J.Gerstman, A.M.Liberman, P.C.Delattre and F.S.Cooper, "Acoustic cues for the perception of initial /w,r,l,j/ in English", *Word* 13, 24-43, 1957.

[8] W.A.Ainsworth, 'First formant transitions and the perception of synthetic semivowels', J.Acoust.Soc.Am. 44, 689-694, 1968.