PERCEPTION OF FORMANT TRANSITIONS IN SYNTHESISED VOWEL PAIRS

W.A.Ainsworth
Dept. of Communication and Neuroscience, Keele University, Staffordshire, UK

ABSTRACT

Speech analysis shows that the second formant transitions in vowel-vowel utterances are not always of the same duration as those of the first formant transitions nor are they always synchronised. Moreover the formant transitions often move initially in a different direction from their final target. In order to investigate whether these deviations from linearity and synchrony are perceptually significant a series of listening tests have been conducted with the vowel pair /ɑ/-/i/.

INTRODUCTION

Speech is produced by a series of articulatory gestures which give rise to formant transitions in the spectrograms of the resulting sounds. It has long been known that that formant transitions are important cues for the perception of many speech sounds, especially voiced consonants [1,2]. These early perceptual studies provided the basis for speech synthesis-by-rule systems [3,4].

These systems were based on formant synthesisers incorporating rules involving simultaneous formant transitions. More recent synthesis systems based on vocal tract models [5,6] generate sounds whose formants change in a nonlinear manner. Careful analysis of natural speech also demonstrates nonlinear and nonsynchronous formant transitions.

The question arises as to whether these departures from linearity and synchrony are perceptually significant. In order to investigate this question a series of experiments have been performed to measure the perceptual tolerances of formant transitions.

ANALYSIS OF VOWEL-VOWEL TRANSITIONS

The six vowels /i, ɛ, ɑ, o, u, ɜ/ were spoken as vowel pairs V_1V_2 in all possible combinations by a male speaker. With many of the formant tracks F1 and F2 changed simultaneously. These give rise to fairly linear transitions in the F1-F2 plane. Other formant tracks, however, show systematic departures from linearity. In the case of /ɛ/-/o/, for example, the path first moved parallel to the F1-axis then to the F2-axis.

Such departures from linearity can arise in two ways. In the case of /ɛ/-/o/ the transition of the first formant is completed in the first 130 ms of the sound whilst the transition of the second formant begins at about this point. In other cases, /ɛ/-/ɑ/ for example, both formants begin and end at approximately the same time but for a time they move in a different direction from their final target.

EXPERIMENTAL PROCEDURE

All the experiments were carried out with the vowel pair /ɑ/-/i/.Four listeners took part in these experiments. Four series of experiments were carried out: the first three involving temporal properties of the transitions and the fourth involving the shapes of the transitions.

The stimuli were two-formant synthetic sounds generated by passing a sequence of pulses with a repetition frequency of 120 Hz through digital resonators.

The stimuli were stored in digital files and presented to listeners via headphones by means of a 16-bit PC sound system. Nine sounds were generated for each

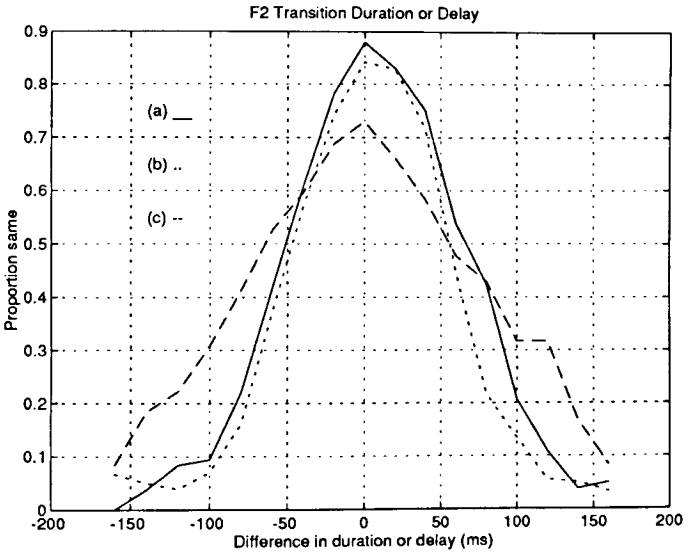

Figure 1. Proportion of stimuli judged the same as a function of the difference between (a) the duration of the F2 transition for simultaneous starts or ends, (b) the delay of the F2 transition with respect to the F1 transition and (c) the duration of the F2 transition for symmetric transitions with respect to the transition of F1 of stimulus B and the duration or delay of the F2 transition of stimulus A.

experiment. They were presented in pairs five times in random order with an ISI of 2 s. Listeners were asked to press 'S' on a keyboard if the two sounds of the pair were judged to be the same and 'D' if they sounded different.

EXPERIMENTS

Transition duration

The stimuli in the first experiment consisted of the /ɑ/-/i/ sounds with an initial steady frequency of F1 of 900 Hz for 100 ms, a linear transition for 100 ms, then a final steady frequency of 250 Hz for another 100 ms. The second formant frequency, F2, was 1100 Hz for 100 ms, then a transition whose duration varied from 20 ms to 180 ms in 20 ms steps, followed by a final steady frequency of 2500 Hz for sufficient time to make the total duration of the sound 300 ms.

It was found that when the F2 transition durations differed by less than about 70 ms half or more of the sounds were judged to be the same.

In the first experiment all the transitions began 100 ms from the beginning of the sound. A further experiment was conducted in which the stimuli were the same as those in the first experiment except that the second formants all ended at the same point as the end of the F1 transition. Similar results were found. The combined results are shown in Figure 1.

Transition delay

The experiments on transition duration confounded two possible cues: the length (or slope) of the F2 transition and the synchrony, or lack of it, between the F2 transitions and the start or end of the F1 transition. In order to explore the effects of this synchrony further experiments were performed in which the durations of the transitions remained constant but the delay between the start of the F2 transition and the F1 transition was varied.

In the first experiment F1 consisted of three segments: a steady segment, a transitional segment and a further steady segment all of 100 ms duration as before. F2 also consisted of three segments with the middle transitional segment remaining constant at 100 ms duration. The first segment, however, varied from 20 ms to 180 ms and the last was adjusted to make the total duration 300 ms.

The average results are also shown by in Figure 1. It can be seen that if the delay between the start of the F2transition and that of the F1 transition was less than about 50 ms half or more of the sounds were judged to be the same.

The experiment was repeated with the durations of both the F1 and F2 transitions at 50 ms or 150 ms. It appeared that there were no systematic differences so that two sounds have a 50% or greater chance of being judged the same if their formant transitions are synchronised to within 50 ms.

Symmetric transitions

In the next series of experiments an attempt was made to measure the effect of transition duration or slope independent of transition synchrony. This was done by generating a set of stimuli whose F2 transition durations varied but whose mid-point remained at the mid-point of the F1 transition.

In the first experiment F1 consisted of three 100 ms segments. The middle F2 segment consisted of the transition and varied from 20 ms to 180 ms. The initial and final F2 segments were equal and chosen so that the total duration of the sound was 300 ms.

This experiment was repeated twice: once with an F1 transition duration of 50 ms embedded between two 100 ms segments and once an F1 transition duration of 150 ms between two 50 ms steady segments. The duration of the F2 transition duration varied from 20 ms to 180 ms as before and the durations of the initial and final segments were chosen to make the total duration 250 ms.

The results of these experiments are also shown in Figure 1. Once again it appears that the duration of the F1 transition, at least in the range 50-150 ms, has little effect on the averaged judgements.

Transition shape

Finally the effect of F2 transition shape was examined. F1 consisted of three 100 ms segments. However the F2 transition was divided into two 50 ms segments with the frequency of F2 at the boundary between them varying from 1000 Hz to 2600 Hz in 200 Hz steps. At the centre of this range the mid frequency of F2 is 1800 Hz giving a linear transition which matches those employed in the previous experiments.

The results showed that a deviation of about 750 Hz in the mid point of the F2 transition can occur before a listener reliably distinguishes between two sounds with this structure.

DISCUSSION

Experiments have been performed to estimate the tolerance of the perceptual system to the duration and delay of transitions in vowel-vowel utterances. The effect of transition shape has also been examined.

A lead or lag of some 50 ms is required for two sounds to be reliably distinguished. Greater differences are required for transitions of different durations. If the F2 transitions are symmetric with respect to the F1 transitions a difference of about 80 ms is required, but if the F2 transitions begin or end simultaneously with the F1 transitions a difference of only about 70 ms is required.

There is evidence that formant transition duration and shape are important in consonant-vowel transitions [1,7]. It therefore remains to be seen whether similar perceptual tolerance values apply to consonant sounds.

CONCLUSIONS

A number of experiments have been performed to estimate the difference limens for the duration, synchrony and shape of formant transitions in a two-formant synthesised vowel pair /ɑ/-/i/. It seems unlikely that the deviations from linearity and synchrony observed in natural vowel-vowel pairs have any perceptual significance.

ACKNOWLEDGEMENTS

The work was supported by EC Science Contract SC1-CT92-0786.

REFERENCES

[1] Liberman, A.M, Delattre, P.C., Gerstman, L.J. & Cooper, F.S (1956) Tempo of frequency change as a cue for distinguishing classes of speech sounds, J.Exptl.Psych., 52, 127-137.

[2] Lisker, L. (1957) Minimal cues for separating /w,r,l,y/ in intervocalic position, Word, 13, 256-267.

[3] Holmes, J.N., Mattingly, I.G. & Shearme, J.N. (1964) Speech synthesis by rule, Language and Speech, 7, 127-143.

[4] Klatt, D.H. (1980) Software for a cascade/parallel formant synthesiser, J.Acoust.Soc.Am., 67 (3), 971-995.

[5] Maeda, S. (1990) Compensatory articulation during speech; evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model, in Speech Production and Speech Modelling (W.J.Hardcastle and A.Marchal, eds.), NATO ASI Series, Kluwer Academic Publisher, Dordrecht.

[6] Mrayati, M., Carré, R. & Guérin, B. (1988) Distinctive regions and modes: a new theory of speech production, Speech Communication, 7, 257-286.

[7] Ainsworth, W.A. (1968) First formant transitions and the perception of semivowels, J.Acoust.Soc.Am., 44, 698-694.