Interactive Speech Synthesis in the Study of Normal Perceptual Development

V. Hazan and A.J. Fourcin
London, United Kingdom

1. Introduction

If normal speech perceptual development is largely dependent on learning processes in which auditory analysis plays an important role, three aspects of speech contrastive development in the normal child are likely to be of special consequence.

First, the earlier stages of speech perceptual development will be characterised by the influence of auditorily dominant speech pattern elements. Second, development will tend to proceed from the simple to the complex in auditory - rather than articulatory - terms. Third, individuals may differ markedly in their ability to use different speech pattern elements in the identification of phonetic contrasts.

The stages of development of a child's acquisition of perceptual ability in the processing of speech patterns have yet to be fully investigated. A normative study of the ages at which the child is likely to start making contrastive use of, for example, F_1, F_2 and F_3 transition cues, VOT information, aspiration cues is imperative in order to assess and facilitate the development of these abilities in hearing impaired children.

The present experiments make use of synthetic speech stimuli defining meaningful phonetic contrasts. These high quality stimuli are modelled on the utterances of a particular woman in order to provide a coherent pattern set and to minimise normalisation problems for the child.

The end-point stimuli are interpolated to provide a six step continuum. In our first work in this area, pre-recorded sequences of these stimuli were used. These were too long for young children. Now, an interactive test system is used, in which the syntheses are made on-line at a level of difficulty which is a function of the subject's response. The ongoing assessment of the subject's performance ensures that the most efficient presentation of stimuli is given, concentrated in the 75-25% labelling area, with fewer stimuli presented in those parts of the stimulus continuum where the subject is labelling with confidence. The test-retest reliability was found to be good.

The response situation has been improved by the use of touch sensitive pictures which relay the child's choice directly to the microprocessor controlling the running of the test. The whole apparatus used is relatively compact and portable so that it can be used in classroom and clinic. The tests available

range in increasing pattern complexity from simple fundamental frequency contours; a vowel contrast between two phonetically contiguous forms; a plosive consonant voicing contrast; and finally a voiced plosive alveolar-velar contrast.

2. Main features investigated

1. Fundamental frequency

English intonation is a highly complex system mainly cued by the fundamental frequency contour pattern. It cues not only the difference between question and statement but also the placement of nuclear stress. An assessment of the child's ability to contrast Fx patterns is made using a very simple 'Oh? - Oh' test in which the Fx contour is systematically varied between an extreme rise and an extreme fall.

2. Vowels

The child's ability to contrast two vowels can be assessed by systematically varying F_1 and F_2 in between two extremes corresponding to two phonetically contiguous vowels. The relative importance of F_1 and F_2 in establishing a contrast can be assessed by presenting each formant separately. This proves very fruitful when trying to assess whether the hearing impaired child is able to make effective use of F_1 information.

3. Consonant place contrasts

Important acoustic features of place contrasts in initial consonants are provided by the F_2 transition into the vowel and the frequency of the noise burst. A 'date/gate' test is used in which these two cues can be varied systematically, either separately or together.

4. Voicing contrast

The voicing contrast in initial cognate plosives is cued mainly by a rising F_1 transition in the voiced consonant and the voice onset time following the noise burst. A 'coat-goat' or 'pea-bee' opposition are used to assess the perception of these contrasts.

3. Results

Results were first obtained for normally hearing adults in order to assess how the stimuli would be labelled by subjects who can potentially make use of a complete set of speech pattern features.

Since large listener to listener variations are typically found, results are analysed individually. The labelling curves below are plotted from the percentage of responses of one label versus step number and illustrate adults' responses to the place contrast in the 'date-gate' opposition and the voicing contrast in the 'coat-goat' opposition, see fig. 1.

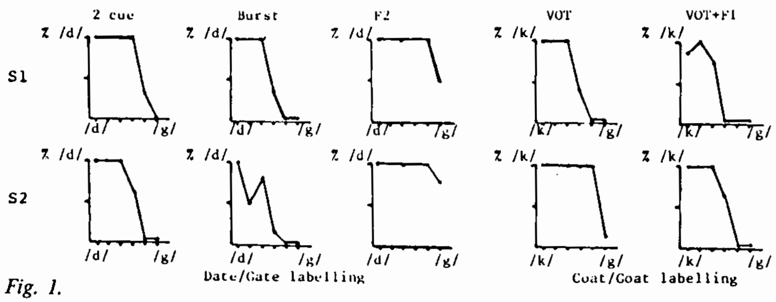

Fig. 1.

In the 'date-gate' test, results obtained for both subjects to the stimuli containing both the F_2 transition and burst frequency cues are characterised by sharp labelling. This is also the case for the stimuli in which the contrast is cued only by a change in the burst frequency value. However, both subjects had difficulties in making the contrast on the basis of F_2 transition alone. It therefore seems that they are primarily relying on the burst frequency as a main cue to this alveolar-velar contrast.

In the 'coat-goat' contrast cued by the change in VOT alone, the first subject does not seem perturbed by the absence of F_1 information and seems therefore to be making the contrast primarily on the basis of the VOT value. The second subject perceives most stimuli as voiceless in the absence of F_1 information. The labelling curve becomes more balanced when this second cue is added.

Results are needed from large numbers of children from age groups ranging from 4 to 14 in order to make a reliable estimate of the ages at which normal children are most likely to make consistent use of the various speech pattern cues. This will necessarily only be a general framework as great individual differences in strategies used by the children are to be expected.

The results given below illustrate responses obtained to the consonant contrast stimuli by two children of 6 (S1) and 9 (S2), see fig. 2.

Both children gave sharp labelling to the 'date-gate' stimuli containing the combined F_2 transition and burst frequency cues. They seemed able to make

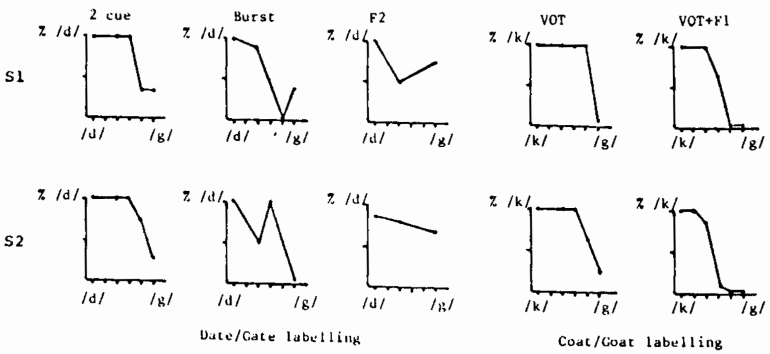

Fig. 2.

use of the burst frequency cue alone although their labelling is less sharp. There is uncertainty in their responses to the stimuli in which the contrast is cued only by the F_2 transition. The 'coat-goat' contrast seems to be more easily perceived, with evidence of sharp labelling of stimuli containing the combined VOT and F_1 transition cues. Both children also respond well to the one-cue stimuli but give a higher percentage of voiceless responses at all steps of the continuum.

4. Conclusion

Interactive speech tests may have widespread future applications for the assessment of speech perception in hearing impaired and language retarded children. While providing a rough quantitative estimate of a child's perception of speech, classical speech audiometry gives no indication as to what features the child is making use of in contrasting two sounds. Interactive speech tests, however, carried out at regular intervals, may provide a real insight into the development of perceptual abilities. Similarly, such tests may be used to assess whether the perceptual development of the language retarded child is following a 'normal' course, albeit delayed, or whether his phonological development is deviant.