Proceedings of The Institute of Acoustics

THE INFLUENCE OF INTONATION CONTOURS ON THE RECOGNITION OF WORDS IN SYNTHESISED SENTENCES.

W.A. AINSWORTH, C.C. HUGHES, & M.P. HUBBARD.
DEPT. OF COMMUNICATION AND NEUROSCIENCE, UNIVERSITY OF KEELE.

Most of the research with synthetic speech has been concerned with the perception of segmental phonemes. A small amount has been concerned with the perception and interpretation of intonation contours (1, 2). Very little has been concerned with the influence of intonation on the recognition of phonemes, words, or sentences (3). Yet in normal speech recognition the words of a sentence are perceived without the listener being aware of the individual constituents of the sounds.

One question which arises is whether intonation has any effect on the recognition of individual words (other than resolving ambiguities where a word has different meanings depending upon the position of the stressed syllable). Isolated words spoken on a monotone can readily be identified. On the other hand, synthesised phrases sound immeasurably better when produced with an approximately natural intonation contour.

It could be that the words of a sentence are recognised by one mechanism and the intonation contour by another, and that the results of these two processes are only combined when the meaning of the sentence is derived. If this is the case it would be expected that a listener who was given the task of writing down the words of a sentence would perform equally well whether the sentence was spoken on a monotone or with natural intonation. If the word and intonation recognition mechanisms are more intimately related, however, it would be expected that the words would be more accurately recognised when the sentence was spoken with natural intonation.

In order to test this hypothesis some experiments have been performed in which listeners were asked to recognise words in synthesised sentences. Synthetic speech was employed because this allowed accurate control of the intonation contour independently of the acoustic structure of the segmental phonemes. The sentences were synthesised with a natural intonation contour and with a monotone. They were also synthesised with an inverted contour. In this condition the fundamental rose where it fell with natural intonation, and vice versa. This condition was included because the brain is more sensitive to stimulus changes than to steady-state stimuli, and it is possible that this in itself would cause words in monotone sentences to be recognised less accurately than those in natural intonation sentences. If words in inverted contour sentences are also less accurately recognised than words in natural contour sentences, then there is a good indication of interaction between the word and intonation recognition mechanisms.

Method

Fifteen sentences were selected from a list (4). These were spoken with natural intonation by one of the experimenters fitted with laryngograph electrodes (5). The waveform of the speech and the output of the laryngograph were digitised at 5kHz sampling rate and stored on the disc of a computer. From these data the waveform envelope and the fundamental frequency of the sounds were calculated for each 10 msec period, and then printed out.

Phontetic transcriptions of the sentences were also made. These were transcribed on to the print-outs using the waveform envelope as a guide. The positions of the phoneme boundaries were then estimated. In this way the value of the fundamental frequency at each phoneme boundary was determined.

The sentences were then synthesised with a synthesis-by-rule system (6, 7). A string of characters representing the phonetic transcription interspersed with others giving the fundamental frequency values was typed into a computer. The program then calculated the parameter values and sent these to a parallel formant speech synthesiser (8). Each sentence was preceded by a warning signal and then recorded with a sufficient interval of time between sentences for a listener to be able to write down the sentence.

The mean value of fundamental frequency was found to be 120 Hz. The sentences were resynthesised with a constant fundamental of this value, and then recorded in the same order to produce the monotone condition.

Finally each value of the fundamental was subtracted from 240 Hz. These values were inserted in the phonetic string and the sentences were again synthesised in the same order to produce the inverted intonation condition.

Thirty six listeners took part in the experiment. They were assigned randomly to each condition, so that 12 heard the sentences with natural intonation, 12 with monotone, and 12 with inverted intonation. They were told that the sentences had been distorted by a computer. They were asked to listen to each sentence following the warning signal, and then to write down all the words they heard, leaving gaps if necessary.

Results

The results are shown in Table I as the percentage of words correctly identified for each intonation condition. With natural intonation 55.2% were correctly heard, with monotone 45.0%, and with inverted intonation 36.8%.

An analysis of variance was performed of these results. This showed that there were no significant differences between the groups of listeners ($F(11, 22) = 0.04$, n.s.), but there were significant differences between the intonation conditions ($F(2, 22) = 6.84$, $p < 0.01$).

Condition	Mean	Standard deviation
Natural	55.2	5.2
Monotone	45.0	9.5
Inverted	36.8	13.6

Table I. Recognition scores (%) for words in sentences synthesised with various intonation contours.

THE INFLUENCE OF INTONATION CONTOURS ON THE RECOGNITION OF WORDS

The differences between the scores obtained with the different intonation conditions were examined with two-tailed t-tests. It was found that the natural intonation condition differed from both monotone and inverted intonation at the 1% level ($t = 3.28$ and 4.37, respectively) but that the monotone differed from inverted intonations only at the 10% level ($t = 1.71$).

Discussion

The results show that natural intonation contours have a significant effect in enabling a listener to recognise words in sentences. Furthermore they indicate that intonation contours which move in the wrong direction cause more interference with correct word recognition than flat intonation contours. This supports the view that intonation is actively involved in the process of word recognition.

The mechanism by which intonation aids word recognition is not known. It could be that it enables the sentence to be broken down into phrases, or smaller linguistic units, which have a regular structure. With these smaller units the range of possible words at each position is narrowed, resulting in easier word recognition. With monotone, no information is available for sub-dividing the sentence, so more mistakes are made. With inverted intonation mistakes will be made in sub-dividing the sentence into phrases, which will lead to further errors at the word recognition stage.

The main drawback of the present experiments is that synthetic speech with an intelligibility of only about 50% was employed, so that generalisations to natural speech recognition should be made with care. The use of such poor quality speech, however, did have the advantage that ceiling effects were avoided. It would probably be possible to improve the quality of the speech by adjusting the durations of the phonemes to match those of the natural speech from which the intonation contours were derived. It would also be possible to increase the intelligibility scores by exposing the listeners to more synthetic speech of this quality (6).

Conclusions

It has been found that intonation contours have a significant influence on word recognition in synthesised sentences. Natural intonation contours produce higher word recognition scores than monotones and these in turn produce higher scores than inverted intonation contours.

References

1. K. HADDING-KOCH and M. STUDDERT-KENNEDY 1964 Phonetica 11, 175 - 181. An experimental study of some intonation contours.
2. M. STUDDERT-KENNEDY and K. HADDING 1973 Language and Speech 16, 293 - 313. Auditory and linguistic processes in the perception of intonation contours.
3. L. B. LEONARD 1973 Language and Speech 16, 327 - 336. The role of intonation in the recall of various linguistic stimuli.
4. H. DAVIS and S. R. SILVERMAN 1964 Hearing and Deafness, Holt, Rinehart and Winston, New York.
5. A. J. FOURCIN and E. ABBERTON 1971 Medical and Biological Illustration 21, 172 - 181. First applications of a new laryngograph.
6. W. A. AINSWORTH 1974 Int. J. Man-Machine Studies 6, 493 - 511. Performance of a speech synthesis system.

7. W. A. AINSWORTH and J. B. MILLAR 1976 Int. J. Man-Machine Studies 8, 159 - 168. Allophonic variations of stop consonants in a synthesis-by-rule system.
8. J. N. HOLMES, I. G. MATTINGLY and J. N. SHEARME 1964 Language and Speech 7, 127 - 143. Speech synthesis-by-rule.