ON THE PROCESSING OF SIMPLE LINGUISTIC STIMULI

OLLI AALTONEN and PEKKA NIEMI
University of Turku

Introduction

In the 1950's, psychology underwent a change of its research paradigm, which finally resulted in what is called cognitive psychology. The latter differs from its predecessor, behaviourism, in its interest in cognition, i.e. its main area of study is the events which take place between the stimulus and the response. As is known, behaviourists argued vehemently that such things remain outside the scope of a scientific enterprise. At the same time, however, research in linguistics, artificial intelligence, neurophysiology, and experimental psychology provided evidence of something totally opposite (Chomsky 1959, Hubel and Wiesel 1965, Miller _et al_. 1960, Newell _et al_. 1958).

New psychology is sometimes called psychology of information processing. It should be noted, however, that the concepts of 'cognitive psychology' and 'information processing psychology' are not interchangeable because the latter carries an additional metaphor of man as a computer-like information processor. For the sake of precision, it should be remembered that cognitive psychology has a wider scope and a more realistic image of man than that, and it can be said to absorb the information processing approach.

A pervasive question for cognitive psychology is that of gaining reliable knowledge of the extremely fast mental processes in man's brain. A direct access, i.e. introspection, is inhibited by the very speed of these processes. At first glance, a convenient solution to the problem would be an accurate neurophysiological recording, excellent in anatomical and temporal precision. In principle, it makes it possible to trace the development of the process from the beginning to the end. Unfortunately, cognitive processes are far too complex to be investigated by neurophysiological methods although the latter have steadily been improved.

A more practical way to study mental processes and their development in the brain is to concentrate on their temporal aspect. The basic tenet in this approach is that all mental processes have a certain duration which can be measured in real time. It is worth noting that real time is quantitative in the same sense as any physiological measurement. Consequently, duration is a property of mental processes which is exploited in the same way by researchers as electrical and chemical events in the brain. In practice, it is the subjects' reaction time (RT) that is measured in a cognitive task.

The absolute RT per se is psychologically fairly uninteresting. Instead, attempts have been made to analyse the reaction process 'from within'. This research logic has an extensive tradition with the first methods having been developed in the 19th century. The approach is usually called 'mental chronometry', the modern version being the 'additive-factor method' (see Figure 1) (Sternberg 1969). The main idea is to find out processing stages which comprise different levels or phases of a process. The stages may be characterized psychologically or, if you like, linguistically. For example, the perception of speech sounds is usually assumed to consist of three successive stages: (1) auditory sensation, (2) phonetic feature analysis, and (3) phonological interpretation.

It should be noted that the concept of a processing stage does not refer to any neuroanatomically defined part of the brain. Instead, it is used to specify those logical chains of events which may be deduced on the basis of the structure of the language and the speech code. Different compo-

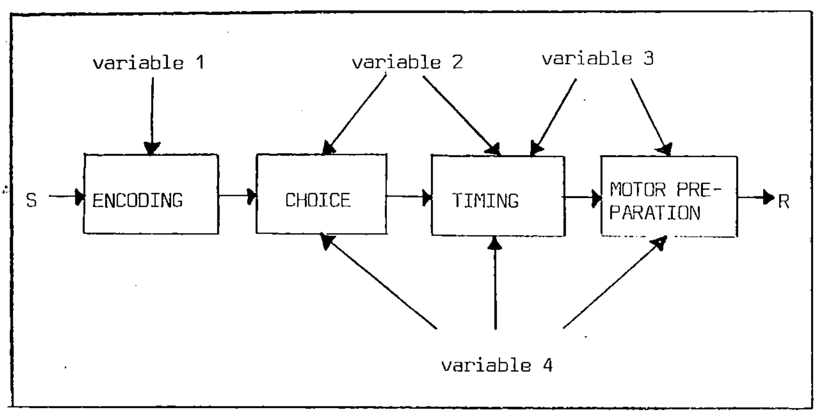

FIGURE 1. The additive-factor method. Variable 1 shows an additive relationship to variables 2, 3, and 4. The latter three variables are interacting.

nents of the chain are affected by different experimental variables. The choice of these variables is based on previous knowledge or, simply, logical inference. For example, it is evident that in a task dealing with the differentiation of acoustic stimuli, stimulus intensity affects at least the initial part of a process, whereas the manipulation of acoustically distinctive cues has an influence on the end part. The core of the research logic is in the inferences made. If two or more variables display a statistically significant interaction, they are assumed to influence the same processing stage. If they display an additive relationship, they are supposed to influence different stages. The statistical procedure is the conventional analysis of variance.

An additional feature of the method is that it reveals processing stages but not processes. The latter are analyzed through accumulating knowledge produced by several converging operations. The stages are an example of mental structures. Their functional organization for any given individual is determined by life-long processes such as learning. (The interested reader is referred to Posner (1978) for a competent account of the uses of mental chronometry in psychology.)

The purpose of the present paper is to demonstrate the usefulness of modern chronometric methodology for the study of the classical problem of phoneme perception. Traditionally, correct identification is used as an indication of phoneme differentiation. It is argued here that, besides its virtues, the method also means waste of information. The data obtained deal only with the final product of a process. Identification is easily incorporated to choice reaction procedures commonly used in mental chronometry. The same information is obtained together with additional data which deal with the processing of the stimulus phonemes reflecting the relative difficulty of processing. Typical encoding variables such as stimulus intensity and auditory channel (ear) are used throughout the study in order to make it certain that the experimental procedure provides a fair comparison of the identification and of the chronometric methods.

Method

Subjects. - The subjects of the present study were four undergraduates at the University of Turku. All were native speakers of Finnish with no known speech or hearing disorder. The subjects were right-handed and their mean age was 20 years. None had any previous experience with either synthetic speech sounds or reaction-time experiments.

Stimuli. - Two sets of steady-state stimuli were computer-synthesized by using a program which combines sine waves into complex speech signals. The constant frequencies of the tones at 250, 1800 and 2400 Hz were the basis of the stimuli. The fundamental frequency was set at 120 Hz and the bandwidths of these formants were 100 Hz. The stimuli differed on the basis of the second and third formant amplitudes. In the first set the second formant amplitude was varied -5, -10, -15, and -20 dB relative to the 0-level. The second set involved the same amplitude variations on the third formant. The constant variables (i.e. the first and the third formant amplitudes in the former set and the first and the second formant amplitudes in the latter set) were maintained at the 0-level. The duration of the warning signal was 500 msec followed by three foreperiods of 250, 500 and 750 msec. The duration of the stimuli was 200 msec and the interstimulus time interval either 2.25, 3.00, or 2.75 sec.

Apparatus. - The stimulus presentation and data collection were controlled on-line in real-time by a laboratory computer system based on a Digital Equipment Corporation LSI-11/23 central processing unit. The subjects interacted with the system through a VT 105 video terminal interfaced to the computer by means of a DEC DLVII-J serial line unit. The sampling frequency of a digital output MNCDO was 20 kHz. A 3.5 kHz LP-filter was used. A REVOX A 77 tape recorder served as an amplifier. The stimuli were delivered to the subjects via JVC Binaural Headphone-Mic HM-200 E headphones. The necessary programs were written in FORTRAN under RT-11 operating system.

Procedure. - A block of 30 randomized stimuli were presented to each subject ten times during one session. The randomization process was made automatically by the computer on the basis of stimulus type, formant amplitudes, and foreperiods. Two separate sessions (600 stimuli in total) were reserved for the presentation of the stimuli either binaurally, to the left ear only, or to the right ear only. The total number of stimuli heard by a subject was thus 1800. The pauses between the blocks were determined by the subject, who started the running of the next block. Prior to the testing, a few blocks were presented to the subjects to familiarize them with the task. The average in-lab time was 30 min for a session (range 20-40 min). Each subject had one session a day. The subjects were tested individually. The room for the tests was shielded from extraneous noises. The subject was sitting in front of the terminal resting her index fingers comfortably on the keys. The two response buttons were labeled as /i/ and /y/ and the subjects were instructed to press down the former, when they identified the stimuli as the Finnish /i/-vowel, and the latter, when they identified the stimuli as the Finnish /y/-vowel.

Results

An analysis of variance with repeated measures was carried out on mean RTs and identifications. The four factors were: ear (both, left, right), stimulus type (F2 and F3), intensity (the five amplitude values of F2 and F3), and foreperiod (the three time intervals between the offset of the warning signal and the onset of the stimulus). All RTs beyond 1.5 sec were excluded from the statistical analysis.

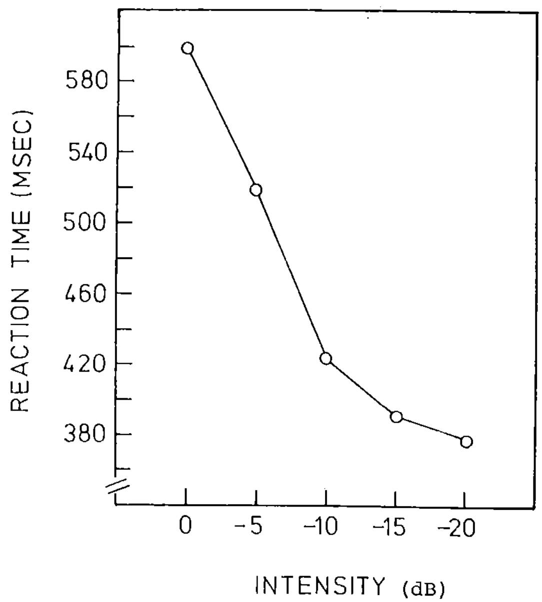

FIGURE 2. The effect of the lessening of the formant amplitude values on RT.

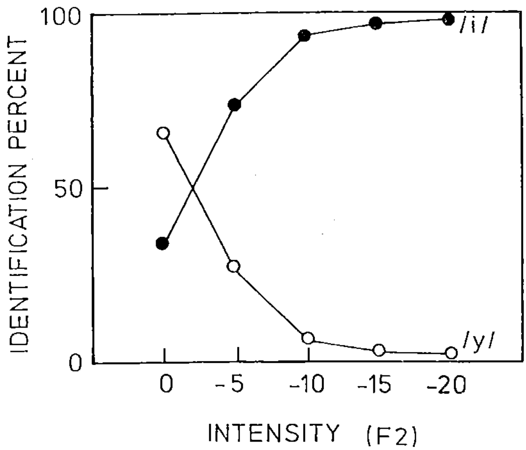

FIGURE 3. The effect of the lessening of F2 amplitude on identifications of /i/ and /y/.

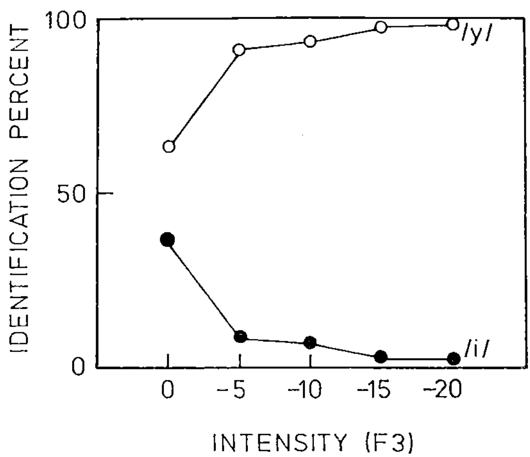

FIGURE 4. The effect of the lessening of F3 amplitude on identifications of /i/ and /y/.

The intensity and foreperiod main effects were significant on the basis of the mean RTs: $F(4,12)=14.43$; $p < .001$ and $F(2,6)=8.10$; $p < .025$, respectively. As is shown in Figure 2, the mean RTs decreased quite steeply when the amplitudes of either F2 or F3 were decreased. The interaction between the type of stimulus and intensity was nonsignificant. The mean RTs differed according to the durations of the foreperiods as follows: 483 msec for 250 msec, 464 msec for 500 msec, and 460 msec for 750 msec. The ear by stimulus type by foreperiod interaction turned out to be significant ($F(4,12)=4.52$; $p < .025$). The most visible reason for this significant interaction was within the ear factor in the case of the shortest foreperiod: the mean RTs differed between F2 and F3, when the stimuli were presented monaurally but not in binaural hearing. No other main effect or interaction reached statistical significance in this analysis.

The type of stimulus was the only factor which proved significant in an analysis of variance based on the identifications ($F(1,3)=151.32$; $p < .005$). The main effect revealed only that the identifications depended on F2 and F 3 in a systematic way. This became explicit in the interaction between stimulus type and intensity, which was highly significant ($F(4,12)=118.71$; $p < .001$). As is shown in Figure 3, the decreasing of the F2 amplitude increased the /i/-identifications, while the /y/-identifications were correspondingly decreased. The decreased F3 amplitude had an opposite effect (Figure 4): the decreasing of the F3 amplitude increased the number of the /y/-identifications and decreased that of the /i/-identifications. The ear by intensity interaction turned out to be almost significant ($F(8,24)=2.39$; $p < .05$). The identifications differed within the ear factor when all the formants (F1, F2, F3) were at 0-level. The stimuli were better identified when they were presented monaurally than when presented binaurally. This difference levelled down immediately when F2 and F3 were out of balance in intensity.

Discussion

It has been widely argued in the literature of speech perception whether or not speech perception is something 'special'. The critical acoustic features that differentiate 'speech' from 'non-speech' have long been sought and the acoustic characteristics of speech sounds are fairly well known today. However, there are questions which remain unanswered such as how linguistic information is coded in the sound wave and transferred via auditory mechanisms to the nervous system, or how the speech code is processed in the brain.

There is plenty of research data on the perception of isolated speech sounds, which is a special form of speech perception. Experimental evidence suggests that there are differences between vowels and consonants in terms of categorical perception, memory functions, and cerebral lateralization. These phenomena have also been used as evidence to the effect that speech perception entails peculiar processes which are distinct from those of non-speech auditory perception (Liberman et al. 1967, Crowder 1973, Molfese 1977). It has been claimed that there is a continuum of 'encodedness', which refers to two things. Firstly, the relationship between the acoustic signal and its phonetic category is complex, and secondly, consonants are more 'encoded' than vowels.

According to the theory of categorical perception, listeners are assumed to have a limited ability to discriminate differences between speech sounds which belong to the same phoneme class. A typical experiment on categorical perception is an identification (labeling) test, where stimuli from a physical continuum are presented in a randomized order to subjects for a classification into one or the other category. Such identification tests have shown that the perception of consonants is categorial by nature while the perception of wovels is more or less fuzzy. Repp (1982) gives an excellent review of evidence that supports categorial perception while Ades (1977) and, more recently, Hary and Massaro (1982) have presented counter-arguments.

It is also well known that vowels and consonants are processed differently in short-term memory when the stimuli are presented auditorily. The findings of different immediate-memory functions for vowels and consonants have been explained by means of a theory of acoustic memory, according to which acoustic information is stored as a transient sensory trace. Consonants, which are acoustically characterized by rapid transitions of the first two formants, are retained a shorter period of time in precategorical acoustic store (PAS) than vowels, which are acoustically characterized by static relations between formant frequencies of a considerably longer duration (Crowder 1973, Battacchi et al. 1978). Arguments against this explanation have been presented by Darwin and Baddeley (1974).

Evidence from dichotic listening studies shows that a right-ear advantage (REA) is harder to show for vowels than for consonants (Studdert-Kennedy and Shankweiler 1969). On the basis of this finding, a special speech decoding mechanism has been postulated in the left hemisphere, which is more necessary for the 'encoded' speech sounds (such as stop consonants) than for others (such as vowels). However, Darwin and Baddeley (1974) have presented counterarguments against this interpretation of REA. According to these authors, the phenomenon of REA reflects the general functions of memory and encoding. Hence, it is disputable whether or not there is any psychological evidence for supposing that different speech sounds are, in general, perceived in different ways.

Delattre et. al. (1952) have shown that the most important acoustic cues for vowel perception are the frequencies of the two lowest formants. However, it has been admitted by them and many others that the two lowest formants are not sufficient for the identification of /i/- and /y/-like front vowels. Later on, Carlson et al. (1975) have used the concept of the effective upper formant to explain the inconsistency between the acoustic structure and perception. The term effective formant refers to the combined effect of two closely spaced formants on the frequency scale. The frequencies of F2 and F3 are assumed to form an effective formant or a center of gravity in the cases of the /i/- and /y/-vowels.

However, it can be shown that F3 alone plays a decisive role, when the acoustic information is ambiguous as a result of the frequency of F2 (Aaltonen 1982a). The results of the present study show that not only the frequencies of the formants but also their amplitudes may be distinctive in the labeling along the continuum /i/-/y/. Changes in the intensity balance between two closely spaced formants produced immediate changes in the identifications by the subjects. Laine and Rahko (1982) have reported similar findings in their experiments.

The steady decrease of the mean RT in the present study indicated how identification becomes easier, even though the phoneme boundary has already been passed. In this respect the perception of vowels proved to be categorical on the basis of the identifications but more continuous when the analysis was based on the processing time. A decreased intensity within the vowel-like stimuli had an effect on the identifications and decreased RT, which suggests that the phonological level of processing was influenced. Lessened overall intensity has been shown to increase RT without any effect on the identification of the stimuli, which implies that the stage of auditory sensation is influenced (Aaltonen 1982b).

The auditory channel (both, left, right) had no significant effect on RT and identifications in the present study. Obviously, a more extensive experiment is necessary to show the possible effects associated with it. Instead, there was the usual foreperiod effect on RT. It reflects the subject's momentary responsive state which is not crucial from the point of view of the identifications. Hence, there was no effect of the foreperiod on the latter.

To summarize, the use of chronometric methodology in the labeling experiment on simple synthetic vowels proved to be a sensitive tool showing how the phonetic quality of the stimuli improved even though the identifications remained unchanged.

References

Aaltonen, O. 1982a. The effect of the third formant on vowel identifications and processing: a reaction-time study of subjects with and without disturbances in speech communication, Papers in Speech Research 4, Publications from the Institute of Finnish Language and Communication, University of Jyväskylä, 60-88.

Aaltonen, O. 1982b. The effect of intensity on the processing and identification of synthetic vowel-like stimuli, in A. Iivonen and H. Kaskinen (eds.), The Eleventh Meeting of Finnish Phoneticians - Helsinki 1982, Publications of the Department of Phonetics, University of Helsinki 35, 9-19.

Ades, E.A. 1977. Vowels, consonants, speech, and nonspeech, Psychological Review 84, 524-530.

Battacchi, M., G. Pelamatti, M. Tessarollo and C. Umilta 1978. Immediate memory for natural stops and vowels, International Journal of Psycholinguistics 5, 5-19.

Carlson, R., G. Fant and B. Granström 1975. Two formant models, pitch and vowel perception, in G. Fant and M.A.A. Tatham (eds.), Auditory analysis and perception of speech, London: Academic Press, 55-82.

Chomsky, N.A. 1959. Review of "Verbal Behaviour" by B.F. Skinner, Language 35, 26-58.

Crowder, R.G. 1973. Representation of speech sounds in precategorical acoustic storage, Journal of Experimental Psychology 1, 14-24.

Darwin, C.J. and A.D. Baddeley 1974. Acoustic memory and the perception of speech, Cognitive Psychology 6, 41-60.

Delattre, P., A. Liberman, F. Cooper and L. Gerstman 1952. An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns, Word 8, 195-209.

Hary, J.M. and D.W. Massaro 1982. Categorical results do not imply categorical perception, Perception & Psychophysics 32, 409-418.

Hubel, D.H. and T.N. Wiesel 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat, Journal of Neurophysiology 28, 229-289.

Laine, U. and T. Rahko. Perceptual evaluation of naturalness of vowels from terminal analog synthesis, in R. Carlson and B. Granström (eds.), The Representation of speech in the peripheral auditory system, Amsterdam: Elsevier Biomedical Press, 229-234.

Liberman, A.M., F.S. Cooper, D.P. Shankweiler and M. Studdert-Kennedy 1967. Perception of the speech code, Psychological Review 74, 431-461.

Miller, G.A., E. Galanter and K.H. Pribram 1960. Plans and the structure of behavior. New York: Holt.

Molfese, D.L. The ontogeny of cerebral asymmetry in man: auditory evoked potentials to linguistic and non-linguistic stimuli in infants and children, in J.E. Desmedt (ed.), Language and hemispheric specialization in man: cerebral ERPs. Progress in Clinical Neurophysiology 3, 188-204.

Newell, A., J.C. Shaw and H. Simon 1958. Elements of a theory of human problem-solving, Psychological Review 65, 151-166.

Posner, M.I. Chronometric explorations of mind. Hillsdale, N.J.: Erlbaum.

Repp, B.H. 1982. Categorical perception: issues, methods, findings. Haskins Laboratories: Status Report on Speech Research SR-70, 99-183.

Sternberg, S. 1969. The discovery of processing stages: extension of Donders' method, in W.G. Koster (ed.), Attention and performance II, Acta Psychologica 30, 276-315.

Studdert-Kennedy, M., and D. Shankweiler 1969. Hemispheric specialization for speech perception, The Journal of the Acoustical Society of America 48, 579-594.