



Published in final edited form as:

Percept Psychophys. 1974 ; 15(2): 285–290.

Reaction times to comparisons within and across phonetic categories*

DAVID B. PISONI and JEFFREY TASH

Indiana University, Bloomington, Indiana 47401

Abstract

Same-different reaction times (RTs) were obtained to pairs of synthetic speech sounds ranging perceptually from /ba/ through /pa/. Listeners responded “same” if both stimuli in a pair were the same phonetic segments (i.e., /ba-/ba/ or /pa-/pa/) or “different” if both stimuli were different phonetic segments (i.e., /ba-/pa/ or /pa-/ba/). RT for “same” responses was faster to pairs of acoustically identical stimuli (A-A) than to pairs of acoustically different stimuli (A-a) belonging to the same phonetic category. RT for “different” responses was faster for large acoustic differences across a phonetic boundary than for smaller acoustic differences across a phonetic boundary. The results suggest that acoustic information for stop consonants is available to listeners, although the retrieval of this information in discrimination will depend on the level of processing accessed by the particular information processing task.

It has been reported that when listening in the *speech mode* a S can identify the phonetic category of a speech sound but cannot discriminate between acoustically different speech sounds selected from the same phonetic category (Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Lieberman, 1970; Pisoni, 1971). This phenomenon is known as “categorical perception,” and it appears to be unique to certain classes of speech sounds (Lieberman, 1970). In the idealized case of categorical perception, a listener can discriminate between two speech sounds only to the extent that he can identify the two sounds as different phonetic segments on an absolute basis (Studdert-Kennedy, Lieberman, Harris, & Cooper, 1970). These findings have frequently been compared to the perception of nonspeech continua such as frequency or intensity. In this case, a listener can discriminate between two nonspeech stimuli much better than he can identify them in an absolute sense (Pollack, 1952, 1953).

The complexity of speech sounds as acoustic events and the status they have as phonetic segments in natural language may force listeners to respond to speech sounds in a categorical or phonetic mode rather than continuous or auditory mode of perception. Since the discrimination tasks typically employed in previous speech perception experiments require Ss to retain information over several seconds, it is reasonable to suppose that a phonetic representation or phonetic code would be maintained in short-term memory in preference to the auditory trace or auditory code of the complex acoustic signal. For example, in the ABX test, which has been employed in a number of speech perception experiments, the S is presented with three successive stimuli separated by about 1 sec. The first two stimuli are always physically different (e.g., A, B), while the third stimulus is

*This research was supported in part by a grant from NICHD to Haskins Laboratories, a PHS Biomedical Sciences grant (S05 RR 7031) to Indiana University, and NIMH Research Grant MH-24027 to the first author. We are very grateful to Professor Lloyd Peterson for the use of his computer and to R. M. Shiffrin, M. Studdert-Kennedy, A. M. Liberman, and I. Pollack for their interest and advice on this project. An earlier report of these findings was presented at the Midwestern Psychological Association meeting, Chicago, May 1973.

identical to either the first or second stimulus (e.g., A or B). The S is required to determine whether the third sound is most like the first or second sound. To arrive at a correct decision, the S must compare Stimulus X with Stimulus B and Stimulus X with Stimulus A and then make a response. The categorical discrimination performance observed in previous speech perception studies employing the ABX test may not be based on comparisons of the specific acoustic properties of the stimuli but rather may involve a comparison of the phonetic features of these stimuli in short-term memory. Indeed, the success of the original categorical perception model proposed by Liberman and his associates (Liberman, Harris, Hoffman, & Griffith, 1957; Studdert-Kennedy et al, 1970) supports such an assumption. It was assumed explicitly that if a listener identified or categorized two stimuli as the same phonetic segments, he could only discriminate these sounds at chance. Hence, discrimination could be predicted on the basis of the identification probabilities alone (see also Pollack & Pisoni, 1971). Ss could discriminate between two stimuli that were identified as different phonetic segments but could not discriminate between two stimuli that were identified as the same phonetic segments even though the acoustic (physical) differences were comparable.

Recently, Pisoni (1973a) has suggested that auditory information from the earliest stages of perceptual processing for speech sounds tends to be lost more rapidly than information from later stages of analysis. Thus, the acoustic information underlying certain classes of speech sounds such as stop consonants appears to be unavailable for use in subsequent discrimination tasks.

However, it is possible that information from the earliest stages of processing might be available to a listener, at least for a short period of time. For example, Pisoni (1973a) found that in an A-X delayed comparison recognition memory paradigm, vowel discrimination decreased as the interval between the test and comparison stimulus increased. Pisoni and Lazarus (1974) have shown that the discrimination of stop consonants may be improved if Ss are explicitly instructed to attend to small acoustic differences between synthetic speech sounds before the discrimination test. A similar result has also been reported by Barclay (1972).

The extent to which unencoded auditory information can be retrieved from short-term memory and used in speech discrimination may depend on a number of factors, including the type of discrimination test, the instructions to Ss, and the particular speech sounds employed as stimuli. Recent approaches to speech perception have stressed that perceptual processing can be divided into a series of stages, each representing a different operation (Studdert-Kennedy, 1973). Thus, an examination of the stage or stages of perceptual analysis employed in a particular type of information processing task might reflect the types of information available during perceptual processing for speech sounds.

The present study deals with the relation between auditory and phonetic information in speech perception. Specifically, we were concerned with determining whether listeners can respond to acoustic differences between categorically perceived speech sounds or whether they can process these sounds only on an abstract phonetic basis. The procedure used to investigate this problem was the reaction time (RT) matching paradigm developed by Posner and his associates (Posner & Mitchell, 1967; Posner, 1969; Posner, Boies, Eichelman, & Taylor, 1969). This procedure provides an opportunity to examine the level of analysis at which comparisons are made by measuring the processing time required for different types of comparisons.

Thus, when a listener is asked to determine whether two speech sounds are the “same” or “different,” the time to arrive at a decision may reflect the level of perceptual processing

and, in turn, the type of information required for a comparison. Some speech sounds may be compared directly, based on their acoustical properties, while other stimuli may require a process of abstraction where invariant phonetic features must first be identified before being compared (Posner & Mitchell, 1967; Posner, 1969). We assume that classifying two *acoustically different* speech sounds as the “same” involves a comparison of abstract phonetic features at a higher level of perceptual analysis than does classifying two *acoustically identical* stimuli as the “same.” The latter comparison could be based on an earlier stage of analysis involving only the low-level acoustic properties of the stimuli.

Figure 1 shows a flow chart summarizing the stages of analysis involved in this type of “same”-“different” classification task. This model has been adapted from Posner’s work on letter classification (Posner & Mitchell, 1967; Posner, 1969). On every trial, a listener is presented with a pair of speech sounds and is required to determine whether the members of the pair are the “same” or “different.” We assume that stimuli are initially encoded at Stage I and that information about both the phonetic features (i.e., a phonetic code) and the auditory properties (i.e., an auditory code) of these stimuli is present in short-term memory (Fujisaki & Kawashima, 1970; Pisoni, 1973a, b). There are three types of stimulus pairs shown in Stage I: A-A, A-a, and A-B. A-A pairs represent acoustically identical pairs of speech sounds. A-a pairs represent acoustically different pairs of speech sounds selected from within the same phonetic category. These stimuli have the same phonetic features but are physically different. Finally, A-B pairs represent speech sounds selected from different phonetic categories. These pairs are acoustically different and have different phonetic features.

Figure 1 also shows two comparison stages, an auditory stage and a phonetic stage. At Stage II, the acoustic similarity of the two stimuli is evaluated against a criterion. If the two stimuli are acoustically identical (e.g., A-A), a “same” response can be executed. If the two stimuli are acoustically dissimilar (A’-B’), a “different” response may be executed. However, if a reliable decision cannot be made on the basis of acoustic similarity alone, processing will continue. It is assumed that an additional stage of processing is required for pairs of stimuli of intermediate acoustic similarity. At Stage III, the phonetic features of the two stimuli are compared. If the stimuli have the same phonetic features, the listener responds “same.” On the other hand, if the stimuli have different phonetic features, the listener responds “different.”

Depending on the particular stimulus pairs, various predictions can be made about the relative amount of time required for “same” comparisons and “different” comparisons. The stimulus comparisons of most interest in the present report are those selected from within a phonetic category, the A-A and A-a pairs. For example, if low-level acoustical information is available for a comparison, RT should be faster for a “same” response when the stimulus pairs are acoustically identical (e.g., A-A) than when they are acoustically different but have the same phonetic features (e.g., A-a). This result would be anticipated if the acoustically identical pairs (e.g., A-A) were matched as “same” at Stage II and the acoustically different pairs (e.g., A-a) were matched as “same” at Stage III. In other words, a “same” response to acoustically different pairs (e.g., A-a) would require an additional stage of analysis, involving a comparison of the phonetic features of the two stimuli. However, if only an abstract phonetic representation is used in the comparison, RTs for “same” should be identical for these two types of stimulus pairs, since they have identical phonetic features.

A similar outcome would also be anticipated in the case of the “different” responses. If distinct auditory and phonetic levels of analysis exist for speech sounds, then “different” responses should be faster for pairs of stimuli separated by large acoustic differences than for pairs of stimuli separated by smaller acoustic differences. Responses to pairs of stimuli

with large physical differences could be based on an evaluation of acoustic similarity at Stage II, while responses to pairs of stimuli separated by smaller physical differences would require a comparison of their phonetic features at Stage III. However, if only an abstract phonetic representation is available, RTs for “different” comparisons should also be the same, regardless of the magnitude of the physical differences between pairs of stimuli. To investigate this problem, same-different RTs were obtained to pairs of synthetic speech sounds selected from within or across phonetic categories. The finding of faster “same” RTs for A-A than for A-a pairs would be support for the notion that distinct auditory and phonetic information for speech sounds is present in STM.

METHOD

Subjects

Listeners were nine paid volunteers, all of whom were either graduate students or staff members associated with the Mathematical Psychology Program at Indiana University. The Ss were right-handed native speakers of English and reported no history of a hearing or speech disorder. Ss were paid for their services at the rate of \$1.50 per hour. All Ss had had some previous experience with synthetic speech stimuli, although they were naive with respect to the exact purposes of the present experiment.

Stimuli

A set of bilabial stop consonant-vowel stimuli were synthesized on the parallel resonance synthesizer at Haskins Laboratories. The basic set of stimuli consisted of seven three-formant syllables 300 msec in duration. The stimuli varied in 10-msec steps along the voice onset time (VOT) continuum from 0 through +60 msec, which distinguishes /ba/ and /pa/. VOT has been defined as the interval between the release of the articulators and the onset of laryngeal pulsing or voicing (Lisker & Abramson, 1964). Synthesizer control parameter values for these stimuli were similar to those employed by Lisker and Abramson (1970) in their cross-language experiments. The final 250 msec of the CV syllable was a steady-state vowel appropriate for an English /a/. The frequencies of the first three formants were fixed at 769, 1,232, and 2,525 Hz, respectively. During the initial 50-msec transitional period, the first three formants moved upward toward the steady-state frequencies of the vowel. For successive stimuli in the set, the delay in the rise of F1 to full amplitude (i.e., the degree of F1 “cutback”) and in the switch of the excitation source from hiss (aperiodic) to buzz (periodic) was increased by 10 msec. Simultaneous changes in amplitude in the lower frequency region and type of excitation source have been shown to characterize the voicing and aspiration differences between /b/ and /p/ in English (Lisker & Abramson, 1970).

Experimental Materials

All stimuli were digitized and their wave forms stored on the Pulse Code Modulation System at Haskins Laboratories (Cooper & Mattingly, 1969). Two types of audio tapes were prepared under computer control: an identification test and a matching test. A 1,000-Hz tone of 100 msec duration was recorded 500 msec before the onset of each trial. This tone served as a warning signal for the S and was also used to trigger a computer interrupt which initiated timing response latency.

Two different 140-trial identification tests were prepared. Each test contained 20 different randomizations of the seven stimuli. Stimuli were recorded singly, with a 3-sec interval between presentations. Each stimulus occurred equally often within each half of the tape.

Four different “same”-“different” matching tests were constructed. Each test tape contained 48 pairs of stimuli. Half of all the trials consisted of within-category pairs requiring a

“same” response, while the other half consisted of across-category pairs requiring a “different” response. Figure 2 shows the arrangement of the stimulus conditions employed in the present experiment.

Within-category pairs were either physically identical (A-A) or physically different (A-a). A-A trials consisted of Stimuli 1, 3, 5, and 7, each paired with itself (i.e., 1-1, 3-3, 5-5, and 7-7). The A-a trials, which were separated by two steps along the continuum, or +20 msec VOT, consisted of Stimulus Pairs 1-3, 3-1,5-7, and 7-5.

Across-category pairs (A-B), which were always physically different, were separated by two, four, or six steps along the continuum (i.e., 3-5, 5-3, 2-6, 6-2, 1-7, 7-1). These comparisons represented differences of +20, +40, or +60 msec VOT, respectively.

Each of the eight within-category comparisons appeared three times within a block of 48 trials, whereas each of the six between-category comparisons appeared four times. Thus, same and different responses were required equally often. The interstimulus interval (ISI) between members of a pair was held constant at 250 msec. Successive trials were separated by 4 sec.

Apparatus

The experimental tapes were reproduced on an Ampex AG-500 two-track tape recorder and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the earphones equivalent to 70 dB SPL re 0.0002 dynes/cm² for a 1,000-Hz calibration tone. Measurements were made on a Hewlett-Packard VTVM (Model 400) before the presentation of each experimental tape. Ss were run individually in a small experimental room. All responses and reaction times were recorded automatically, under the control of a PDP-8 computer located with the tape recorder in an adjacent room.

Procedure

The instructions for the identification test were similar to those used in previous speech perception experiments. Ss were required to identify each stimulus as either /ba/ or /pa/ and to respond as rapidly as possible. The Ss responded to each stimulus by pressing one of two labeled telegraph keys. For a given S, one hand was always used for a /ba/ response, while the other hand was used for a /pa/ response. The two keys were counterbalanced for hands across Ss.

For the matching task, Ss were told that they would hear a pair of stimuli on every trial and that their task was to decide whether the two stimuli were the “same” or “different” phonetic segments. These instructions are analogous to the “name match” instructions employed by Posner and Mitchell (1967). Ss were told that half of all pairs were the same (e.g., /ba-/ba/ or /pa-/pa/) and half were different (e.g., /ba-/pa/ or /pa-/ba/). Ss were encouraged to respond as accurately and as rapidly as possible. Ss responded to each pair by pressing one of two telegraph keys, labeled “same” and “different.” The response keys were also counterbalanced for hands across Ss.

Ss were tested for an hour a day on 2 consecutive days. Each session began with a 140-item identification test, which was followed after a short break by two 48-trial matching tests. The first day served as a practice session. Only the identification and matching data from the second session are considered in this report.

RESULTS AND DISCUSSION

Identification Task

The average identification function is shown in Fig. 3 along with the mean RT for identification. Each point represents the mean of 180 responses over the nine Ss. The filled squares and open circles show the percentage of /ba/ or /pa/ responses to each of the seven stimuli in the continuum. The filled triangles represent the corresponding latency of identification response to each stimulus.

Examination of Fig. 3 indicates that the identification function is quite consistent. Ss partitioned the stimulus continuum into two discrete phonetic segments. The phonetic boundary or crossover point in identification is at about +30 msec VOT, which is consistent with previous findings (Lisker & Abramson, 1967).

Inspection of the RTs for identification shows that Ss are slowest for Stimulus 4, which is at the phonetic boundary, and fastest for the other stimuli, which are within phonetic categories. These results are also consistent with the findings reported by other investigators who have studied reaction time in the identification of synthetic speech sounds (Studdert-Kennedy, Liberman, & Stevens, 1963). Reaction time is a positive function of uncertainty, increasing at the phonetic boundary where identification is least consistent and decreasing where identification is most consistent. In anticipation of the discrimination tests, we note that identification time is slowest for the stimulus region where discrimination is best.

Same-Different Comparisons

The main results, of the “same”-“different” classification task are shown in Fig. 4. The mean RT for each of the two types of “same” trials (i.e., A-A, A-a) is based on 216 responses, while the mean RT for each of the three types of “different” trials is based on 144 responses averaged over the nine Ss. Only RTs for correct responses were included in the analysis. The overall error rate for all types of trials was 4.0%. The “same” responses shown in Fig. 4 are based on the mean of /ba/ and /pa/ within-category comparisons for A-A and A-a trials.

An examination of the “same” responses indicates that Ss are faster for pairs of acoustically identical stimuli (e.g., A-A) than for pairs of acoustically different stimuli (e.g., A-a) which have been selected from within a phonetic category. The 41-msec difference between these two conditions is significant ($p < .01$) by a correlated t test, $t(8) = 3.20$ (one-tailed). These results are consistent with the model described earlier. Ss can employ low-level acoustic information in a comparison even though they have categorized these pairs of stimuli as the “same” phonetic segments. Thus, “same” matches to acoustically identical speech sounds are presumably based on an earlier stage of perceptual analysis than are “same” matches to acoustically different speech sounds. In the latter case, the “same” response is based on a comparison of the phonetic features of each stimulus in Stage III. We assume that the abstraction of phonetic features from the acoustic signal requires an additional amount of processing time and is reflected in the difference in RT between the two types of within-category comparisons (i.e., A-A and A-a).

It is possible that these findings could result entirely from comparisons involving Stimulus 3, which was categorized less consistently in identification and also showed somewhat longer latency in identification. However, when the A-A and A-a trials are examined separately by phonetic category (e.g., /ba/ and /pa/), the difference between A-A and A-a pairs is still obtained. Table 1 shows the mean “same” RTs for A-A and A-a trials and the error rates. The difference between A-A and A-a trials was 31.1 msec for the /ba/ category and 48.6 msec for the /pa/ category. If the difference between A-A and A-a trials were due to comparisons involving only Stimulus 3, we would expect a large difference between these

two types of trials in the /ba/ category and no difference in the /pa/ category. In fact, the difference between A-A and A-a comparisons is slightly larger in the more consistently identified /pa/ category, and this possible objection can be dismissed.

The present findings, based on “same” responses to within-phonetic-category comparisons indicate that even perception of stop consonants may not be entirely categorical, as previously supposed. Rather, the degree of categorical perception will depend upon the extent to which acoustic information can be utilized within the experimental task. Since acoustic information not only decays rapidly over time but is highly vulnerable to various types of interfering stimuli, the specific discrimination procedure may be crucial in determining the relative roles of acoustic and phonetic information in speech sound discrimination. For example, the ABX procedure may force the listener to rely almost entirely on phonetic information in discrimination because of the arrangement of stimuli in this procedure.

An additional point should be emphasized concerning the within-phonetic-category comparisons. It could be argued that, in the present experiment, Ss were responding to these stimuli as isolated acoustic events rather than as speech sounds. If so, the proportion of “same” responses should have been quite different for A-a pairs than for A-A pairs. In fact, $P(\text{“same”} | A-a) = .99$ and $P(\text{“same”} | A-A) = .96$, indicating that Ss responded to these stimuli as speech sounds.

The mean correct RTs for “different” responses (i.e., A-B comparisons) are also shown in Fig. 4. The error rates were 9.7%, 5.6%, and 1.4% for the 2-, 4-, and 6-step comparisons, respectively. Examination of these RTs provides additional support for the assumption that Ss can employ relatively low-level acoustic information in the comparison process. RT for “different” matches is slower for a 2-step difference than for a 4-step or 6-step difference across the phonetic boundary. Both differences are significant ($p < .005$) by correlated t tests, $t(8) = 4.95$ and $t(8) = 4.82$, respectively. The difference between the 4-step and 6-step comparisons was not significant, $t(8) = .93$. These findings suggest that “different” matches may not be based solely on an abstract phonetic representation. Rather, a “different” response to pairs of stimuli across category boundaries may also be based on low-level acoustic information at an earlier stage of perceptual analysis. Pairs of stimuli which are separated by large physical differences in VOT, such as 1-7 and 2-6, can be differentiated solely on the basis of their acoustic dissimilarity. Stimuli pairs separated by small differences in VOT, such as 1-3 and 3-5, cannot be differentiated on the basis of their acoustic similarity, and an additional stage of analysis is required. Since an initial decision cannot be made reliably on the basis of gross acoustic information alone, the “different” decision for Pair 3-5 must be based on a comparison of the phonetic features of the two stimuli at Stage III.

In summary, the results of this study suggest that low-level acoustical information about a speech stimulus may be available to listeners along with a more abstract phonetic representation, even in the case of stop consonants. Presumably the extent to which low-level information can be accessed will depend not only on the particular level of perceptual analysis examined, but also on the type of information processing task employed.

The present findings argue for a diversity of experimental tasks in the study of speech sound perception. On the basis of the distribution of responses alone, we might conclude that only a categorical or phonetic analysis is available for stop consonants. The addition of the RT task reveals another level of analysis. A view of speech sound perception entailing a series of interrelated stages of analysis could serve as the framework for determining quantitatively the ways in which speech perception may involve specialized mechanisms for perceptual

analysis. Moreover, such an approach may help to determine the ways in which various speech perception phenomena may conform to more general perceptual processes.

References

- Barclay JR. Noncategorical perception of a voiced stop. A replication *Perception & Psychophysics*. 1972; 11:269–273.
- Cooper, FS.; Mattingly, IG. Status Report on Speech Research (SR-17/18). Haskins Laboratories; New York: 1969. Computer-controlled PCM system for investigation of dichotic speech perception; p. 17-21.
- Fujisaki, H.; Kawashima, T. Annual Report on Speech Research (SR-17/18). Haskins Laboratories; New York: 1969. Some experiments on speech perception and a model for the perceptual mechanism; p. 17-21.
- Fujisaki, H.; Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. Annual speech mode. In: Hamburg, DA., editor. *Perception and its disorders*, Proceedings of ARNMD. Baltimore: Williams & Wilkins; 1970. p. 238-254.
- Lieberman AM, Cooper FS, Shankweiler DS, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- Lieberman AM, Harris KS, Hoffman HS, Griffith BC. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*. 1957; 54:358–368. [PubMed: 13481283]
- Lisker L, Abramson AS. A cross language study of voicing in initial stops: Acoustical measurements. *Word*. 1964; 20:384–422.
- Lisker, L.; Abramson, AS. Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. Prague: Academia; 1970. The voicing dimension: Some experiments in comparative phonetics; p. 563-567.
- Pisoni, DB. Status Report on Speech Research (SR-27). Haskins Laboratories; New Haven: 1971. On the nature of categorical perception of speech sounds; p. 101
- Pisoni DB. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*. 1973a; 13:253–260. [PubMed: 23226880]
- Pisoni, DB. Haskins Laboratories Status Report on Speech Research (SR-34). 1973b. The role of auditory short-term memory in vowel perception; p. 89-117.
- Pisoni DB, Lazarus JH. Categorical and noncategorical modes of speech perception along the voicing continuum. *Journal of the Acoustical Society of America*. 1974; 55 in press.
- Pollack I. The information in elementary auditory displays. *Journal of the Acoustical Society of America*. 1952; 24:745–749.
- Pollack I. The information in elementary auditory displays, II. *Journal of the Acoustical Society of America*. 1953; 25:765–769.
- Pollack I, Pisoni DB. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*. 1971; 24:299–300.
- Posner, MI. Abstraction and the process of recognition. In: Bower, GH.; Spence, JT., editors. *The psychology of learning and motivation*. New York: Academic Press; 1969. p. 44-100.
- Posner MI, Boies SJ, Eichelman WH, Taylor RL. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*. 1969; 79(1):1–16.
- Posner M, Mitchell RF. Chronometric analysis of classification. *Psychological Review*. 1967; 74:392–409. [PubMed: 6076470]
- Studdert-Kennedy, M. The perception of speech. In: Sebeok, TA., editor. *Current trends in linguistics*. Vol. XII. The Hague; Mouton: 1973.
- Studdert-Kennedy M, Liberman AM, Stevens KN. Reaction time to synthetic stop consonants and vowels at phoneme centers and at phoneme boundaries. *Journal of the Acoustical Society of America*. 1963; 35:1900.
- Studdert-Kennedy M, Liberman AM, Harris K, Cooper FS. The motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*. 1970; 77:234–249. [PubMed: 5454133]

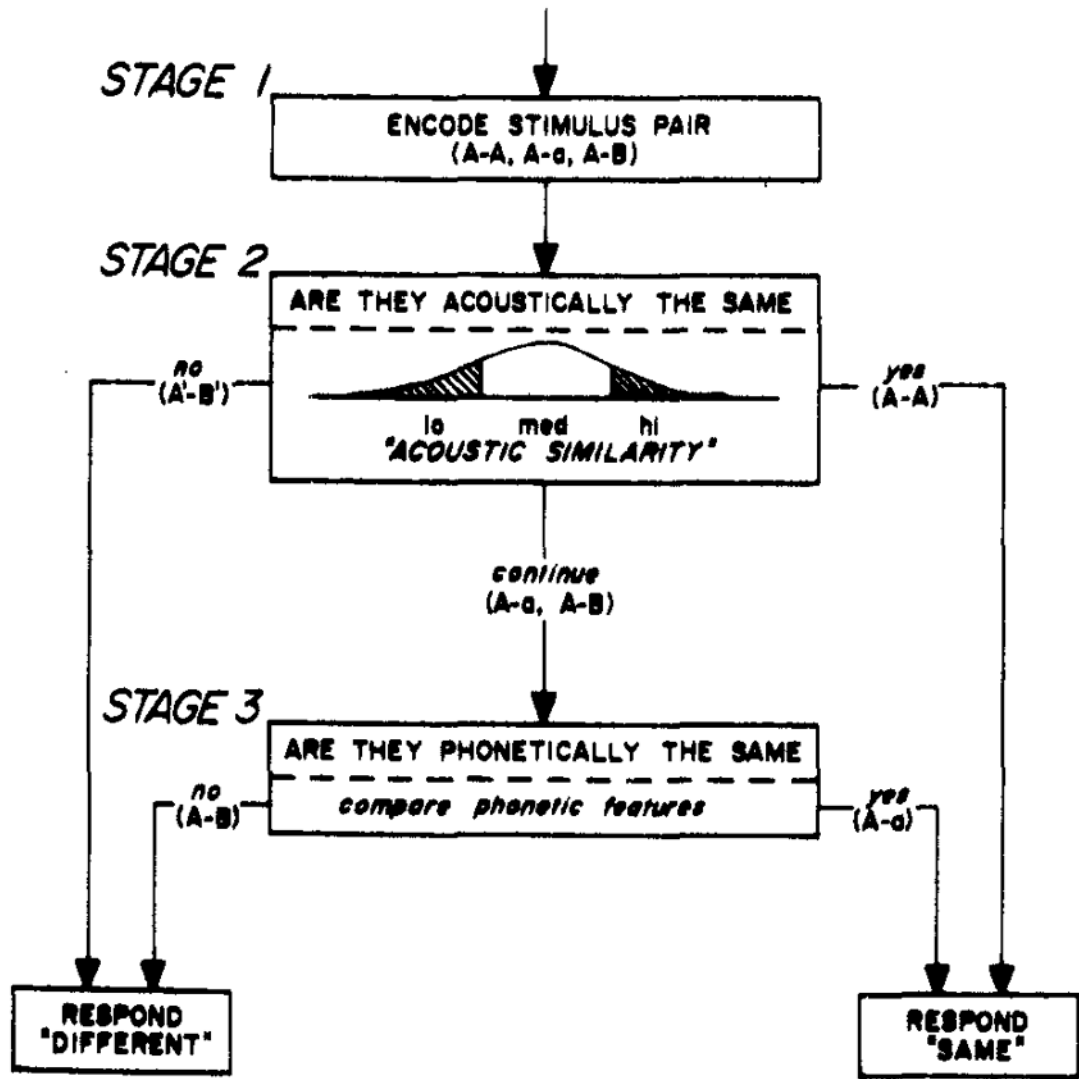


Fig. 1. Flowchart of the stages of analysis involved in the "same"- "different" classification task. In Stage I, the pair of speech sounds is encoded. In Stage II, the acoustic similarity of the stimuli is evaluated, whereas in Stage III, the phonetic features are compared.

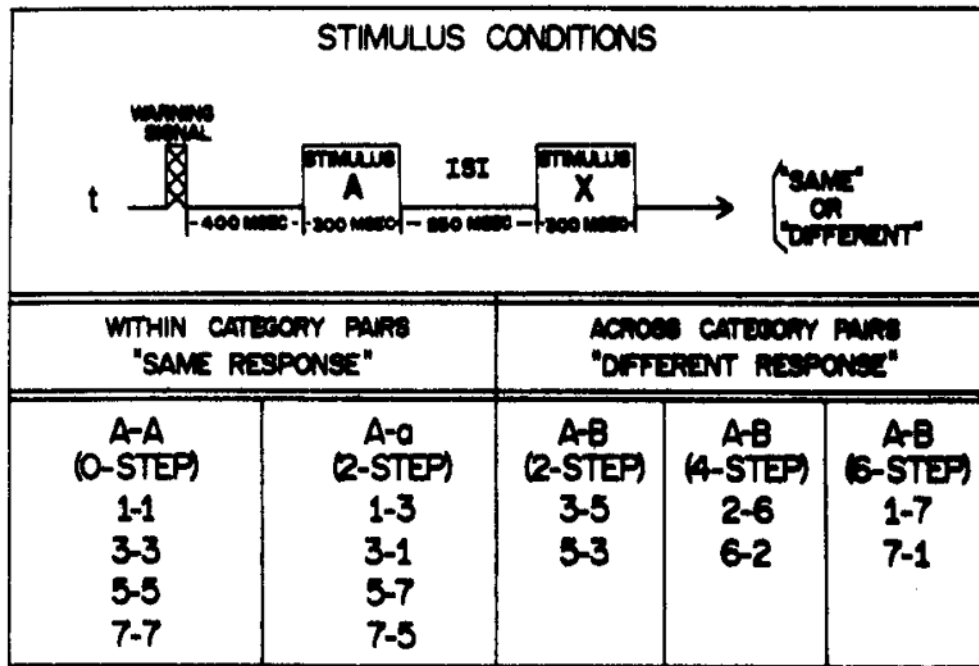


Fig. 2. Description of the stimulus conditions employed in the matching task. Pairs of stimuli requiring a "same" response are selected from within a phonetic category; pairs of stimuli requiring a "different" response are selected from across phonetic categories.

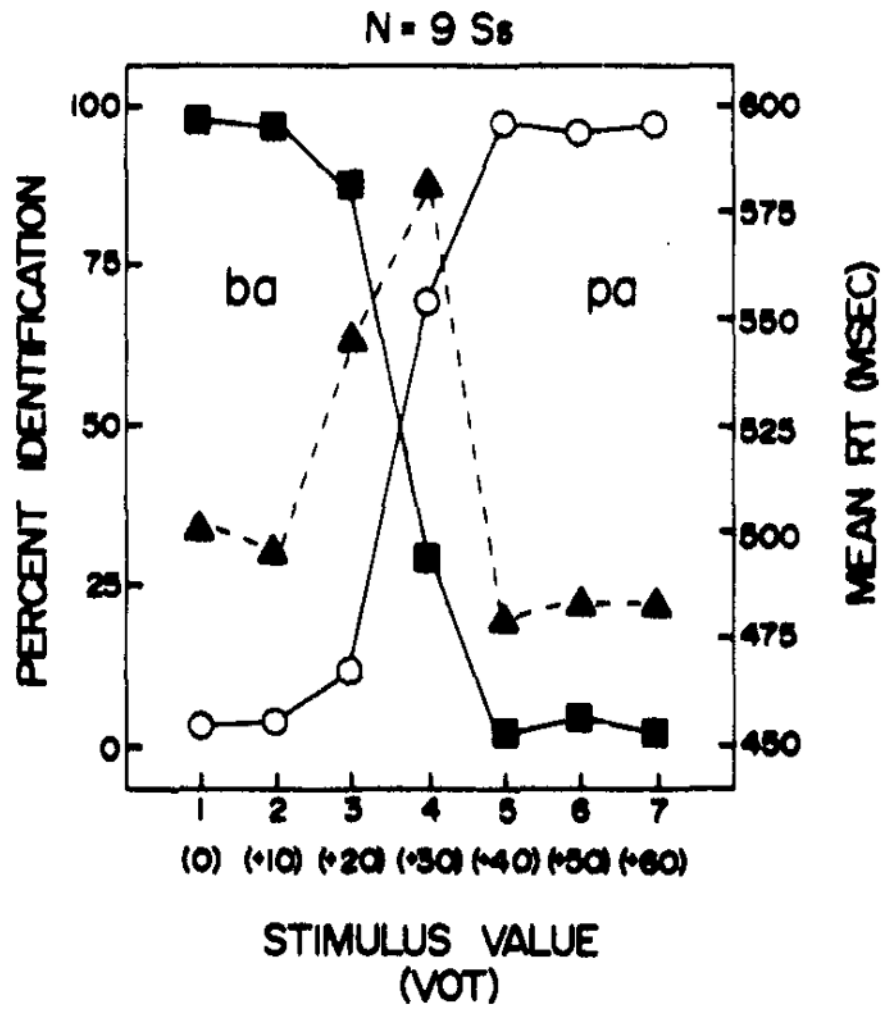


Fig. 3. Average identification function for the voice onset time continuum with the mean RT during identification.

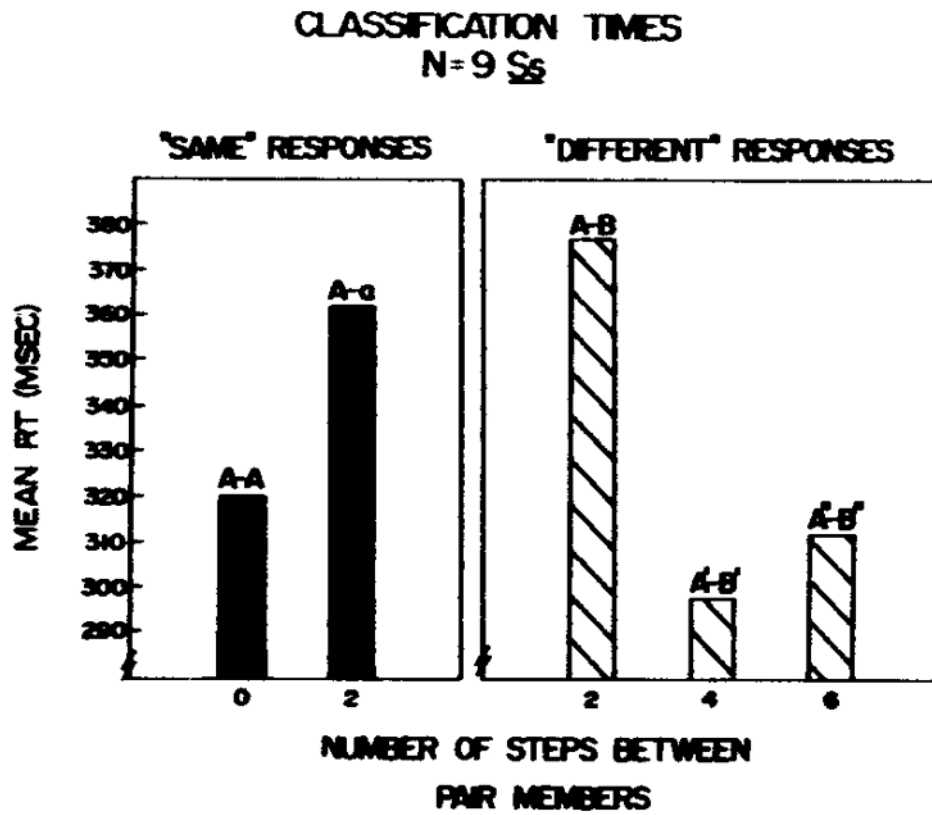


Fig. 4.
Mean RT for "same" and "different" responses to within- and across-category comparisons. The number of steps between pair members reflects the magnitude of the acoustic difference in voice onset time (VOT).

Table 1

Mean Reaction Times (in Milliseconds) and Error Rates for “Same” Responses to Within-Category Pairs Separately for /ba/ and /pa/ Comparisons

	<u>/ba/ Comparisons</u>		<u>/pa/ Comparisons</u>	
	A-A	A-a	A-A	A-a
Mean RT	359.2	390.3	284.4	333.0
Percentage of Errors	4.6	.9	2.8	.9