Effects of fundamental frequency contrast on discrimination and identification of dichotic CV syllables at various temporal delays

BRUNO H. REPP

Department of Oral Biology, University of Connecticut Health Center, Farmington, Connecticut 06032

Dichotic CV syllables (identical and nonidentical pairs) were presented at nine temporal offsets between 0 and 500 msec. One task consisted in judging quickly whether the syllables in a pair were phonetically the same or different; the other task was to identify both syllables. The fundamental frequency (pitch) of the synthetic stimuli was either the same or different, and either predictable or unpredictable. The pitch variable had surprisingly little effect on the latencies of "same"-different judgments, and the expected "preparation" effect of pitch predictability was barely present. Instead, there were strong effects on the frequencies of errors at short temporal delays, which suggests shifts or biases in the phonetic "same"-different criterion with context. A comparison with analogous errors in the identification task revealed identical patterns. Further analysis of identification errors showed no overall "feature sharing advantage": The direction of this effect depends on the kind of error committed. Also, a lag effect was found only in nonidentical pairs that received two identical responses. The results are discussed in the framework of a two-stage information-processing model. Effects of pitch are tentatively explained as biases from implicit (pitch) decisions at the auditory level on phonetic decisions in the presence of uncertainty. Four sources of errors are identified: fusion at the auditory level; "integration," confusions, and transpositions at the phonetic level.

The experiment to be reported investigated the latencies and the accuracy of "same"-different judgments, as well as identification accuracy, for dichotic syllables at various stimulus onset asynchronies (SOAs). Several interesting results have emerged from earlier research employing related paradigms. For example, Cole and Scott (1972) demonstrated that the speed of "different" judgments is positively related to the number of distinctive features separating simultaneous dichotic syllables. Weiner and Singh (1974), Reed (Note 1), and Repp (1975b, Note 2) have shown that the same effect is obtained when the syllables are temporally separated (by up to 1 sec), suggesting that the "perceptual distance" resulting from feature contrasts plays a role in the comparison process in memory. At short SOAs (< 150 msec), discrimination errors are frequent. Incorrect "same" responses are positively related to stimulus similarity (i.e., more errors when there are fewer feature contrasts), while incorrect "different" responses appear to be inversely related to the intelligibility of the stimuli in an identical pair (Repp, 1975b, Note 2).

This research was conducted while the author was a research associate at the University of Chicago. It was supported by NSF Grant GB31127X2 and by NIE Grant C-74-0031, both to David McNeill whose support and encouragement made this research possible. Franklin Cooper, president of Haskins Laboratories, generously supported a visit to this excellent research institution and provided the facilities for the preparation of stimulus tapes, whereby Jim Cutting and Gary Kuhn provided valuable help.

The present experiment was prompted, in part, by the second of two studies reported by Cole and Scott (1972). They presented dichotic (simultaneous) syllables in different voices—a male voice in one ear and a female voice in the other—and obtained the same effect of feature contrasts on "different" reaction times (RTs) as with syllables in the same voice. However, average RTs increased by about 400 msec relative to the same-voice condition, while error rates remained very low. Halwes (1969) has shown that differences in fundamental frequency facilitate selective attention to dichotic syllables in an ear-monitoring (identification) task. This increased perceptual separation of the two "channels" may explain the low error rates obtained by Cole and Scott. However, the relative increase in RT in the different-voices condition suggests at least two possibilities: Either the subjects had to perform a mental transformation on the stimulus traces to bring them into a common "format" before they could be compared phonetically, or, more trivially, the increase in RT simply reflected the fact that the subjects were forced to compare the syllables at the phonetic level, while an auditory criterion (complete vs. incomplete binaural fusion) was available with identical voices.

In the second case, the differences in RT between the same-voice and different-voice conditions should disappear with temporal separation of the syllables, since this eliminates binaural fusion. However, studies of Springer (1973) and Cole, Coltheart, and

Allard (1974) have shown that "same" judgments, at least, are consistently faster when two syllables are spoken in the same voice than when they are spoken in different voices (or pitches), and this occurs at temporal separations of up to 8 sec. These findings support the hypothesis that the listeners have to perform some "normalizing operation" on stimuli in different voices before they can determine that they are phonetically the same or different, or, in other terms, that speaker characteristics such as fundamental frequency enter the perceptual space on which the subjects' phonetic decisions are based.¹

There are other studies that suggest that fundamental frequency (or pitch, for short—a variable with little phonetic relevance) is processed prior to phonetic dimensions (e.g., place of articulation). Using an adaptation of Garner's (1974) speeded classification paradigm, C. Wood (Day & Wood, 1971; Wood, 1973, 1975a,b) has shown that the two-alternative choice RT for syllables that contrast in place of articulation (e.g., /ba/ vs. /da/ increases (relative to a fixed-pitch condition), when pitch varies randomly between two levels, and decreases when pitch is correlated with place. Classification of the same syllables with respect to pitch (high-low), on the other hand, was hardly affected by phonetic variation. In other words, it seems that phonetic information can be largely ignored when processing pitch, but not the other way around. This suggests that these dimensions are processed in an at least partially serial fashion. (Note, however, that Springer, 1973, did find a clear effect of phonetic sameness and difference on the speed of "same"-"different" judgments for pitch, which is in contradiction to C. Wood's results.)

The persistence of the effect of fundamental frequency on "same"-"different" RTs at long time separations (Cole et al., 1974; Springer, 1973) suggests that some speaker characteristics are encoded and retained together with phonetic features in a single more or less "integral" (Garner, 1974) code. One purpose of the present experiment was to investigate whether this code is fixed and unchangeable or whether the listeners can modify their memory representations of speech sounds in anticipation of a comparison stimulus, i.e., whether they can "prepare" for a certain pitch. Springer (1973) and Cole et al. (1974) used a completely randomized design in which pitch was never predictable, so that the observed effect on RT may have been a function of this uncertainty. On the other hand, when the syllables in a pair are simultaneous (Cole & Scott, 1972), there is no time for preparation to occur (except for both stimuli at once). In the present study, a random condition was compared with a blocked condition in which the fundamental frequency of both syllables remained fixed and predictable. If the subjects were able to manipulate their own memory codes to optimize performance, the effect of pitch on "same"-"different" RTs should be reduced or

eliminated in the blocked condition. Moreover, the amount of reduction should depend on the time elapsing between the two stimuli, since any transformation or generation of memory codes should take some time. To this end, a range of SOAs between 0 and 500 msec is used. Posner, Boies, Eichelman, and Taylor (1969), Tversky (1969), and L. Wood (1974), among others, have demonstrated such preparation effects between modalities in paradigms very similar to the present one. However, the operation required here was intramodal and unidimensional in nature, which makes it perhaps more comparable to the mental rotation of visual figures, for which preparation effects have also been demonstrated (Cooper & Shepard, 1973; Metzler & Shepard, 1974).

Cole and Scott (1972) contrasted the low error rates of their experiments with the high error rates typically found in dichotic identification studies (e.g., Studdert-Kennedy & Shankweiler, 1970). This comparison is not quite valid, since most dichotic identification studies in the literature did not include any identical pairs and thus did not involve any implicit judgments of sameness or difference. Therefore, identification accuracy for a set of nonidentical pairs cannot be compared with the accuracy of "same"-"different" judgments for a mixed set of identical and nonidentical pairs.² Only Repp (1975a) included identical pairs at various SOAs in an ear-monitoring task. He found no dichotic interference with these pairs, but this does not necessarily mean that they were explicitly or implicitly recognized as identical pairs, since intrusions from the other ear may have compensated for loss of information from the attended ear. Moreover, there is evidence that high rates of "same"-"different" errors occur when dichotic syllables are separated by short SOAs (Repp, 1975b, Note 2), including a substantial number of incorrect "different" responses to identical syllable pairs. These latter errors ("false alarms") are of special theoretical interest, since they seem to be based on a comparison of two different memory codes for two identical stimuli. It would be useful to find out what the subjects actually perceive when they judge identical stimuli to be different.

To this end, and for the general purpose of comparing the error patterns under different response requirements, the present study included an additional condition in which identical and nonidentical pairs (in fact, the same experimental tapes as in the RT task) were presented for identification of both syllables. It may be that, in the RT task, the listeners adopt a strategy whereby they rely more heavily on comparisons of auditory, "precategorical" information at an earlier processing stage (cf. Pisoni & Tash, 1974), while they may not make use of such information in an identification task where the time pressure is less severe. Different processing strategies may be reflected in the error frequencies.

In addition to its comparative purpose, the identifi-

cation task provided an opportunity to examine, for the first time, the effects that variations in fundamental frequency and the particular composition of the stimulus set (both identical and nonidentical pairs) may have on the error pattern and on standard dichotic listening phenomena such as the "feature sharing advantage" (Pisoni & McNabb, 1974), the "lag effect," and the right-ear advantage (cf. Repp, 1975a).

METHOD

Subjects

Sixteen right-handed subjects, mostly students at the University of Chicago, were recruited through an advertisement in the campus newspaper and were paid \$10 for the whole experiment. Eleven of the subjects were female. Only one subject had previously participated in dichotic listening experiments (she actually turned out to be the poorest subject), while all others were experimentally naive. No subject had any known hearing deficiency. Two prospective subjects were excused during the first session because of their apparent inability to discriminate dichotic syllables at short SOAs.

Stimuli

A master tape was prepared at Haskins Laboratories (New Haven, Connecticut). It contained all possible ordered pairs of the eight syllables, /be/H, /de/H, /pe/H, /te/H, /be/L, /de/L, /pe/L, /te/L; the subscripts "H" and "L" denote high and low pitch, respectively. The syllables were recorded on separate channels at each of nine SOAs (0, 50, 75, 100, 125, 150, 250, 350, 500 msec).

The syllables were synthetic three-formant patterns of 300-msec duration, produced on the Haskins parallel formant synthesizer. They started with formant transitions which led to a steady state vowel portion with a linearly falling pitch contour. The fundamental frequency of "L" syllables fell from 130 to 90 Hz and that of "H" syllables from 190 to 150 Hz. The latter were generated from the former by changing only the pitch parameter in the synthesis specifications, leaving all other parameters unchanged.

The experimental tapes were dubbed from the master tape. There were two sets of tapes, one for the *blocked* condition and one for the *random* condition. The blocked tapes contained four blocks of stimuli, each representing one of the four possible pitch combinations: H-H, H-L, L-H, L-L. Within each block, the nine SOAs were also blocked in groups of three, each preceded by three "warm-up" pairs. The sequence of SOAs was ascending from 0 to 500 msec in the L-L and L-H blocks and descending from 500 to 0 msec in the H-H and H-L blocks. The 24 pairs at each SOA consisted of 12 identical pairs (each of the four individual pairs occurred three times) and all 12 nonidentical pairs, in random order and separated by interpair intervals of 3 sec. The random tapes were exactly identical to the blocked tapes in structure, except that the various combinations of H and L syllables were distributed in a carefully balanced fashion over all four blocks (which now became four "parallel" sets).

Design

The basic design of the "same"-different RT part of the experiment can be described as a complete $2^1 \times 9$ factorial. There were three counterbalanced between-subject factors: "hands" (the two possible hand-response assignments), "ears" (the two possible ear-lag assignments), and "order" (the two possible orders of the blocked and random conditions). In addition, there were eight within-subject factors in the stimulus design: "conditions" (blocked vs. random), "SOA" (nine levels), "pitch of leading syllable" (H vs. L), "pitch of lagging syllable," "voicing of leading syllable" (voiced vs. voiceless), "voicing of

lagging syllable," "place of leading syllable" (labial vs. alveolar), and "place of lagging syllable." In the analysis some of these factors were collapsed (see below).

Since the sequence of SOAs within blocks was fixed, it was confounded with "pitch of leading syllable" in the blocked condition; in the random condition, it simply acted as a balancing factor. The sequence of blocks within conditions was balanced between subjects in two replications of two 4 by 4 Latin squares. For each individual subject, this sequence was the same in both conditions.

The identification task was interpolated between the two RT sessions. The design of this part of the experiment was completely yoked to that of the first session, so that each subject effectively repeated the first session with only a change in response requirements. Thus, the "conditions" factor was split between subjects.

The total data pool consisted of 27,648 RTs (1,728 per subject) plus an equal number of identification responses (since there were two per stimulus pair).

Procedure

The three sessions took place on different days, separated by variable intervals. Each session lasted approximately 90 min. The subject was seated in a sound-isolated booth facing a small table furnished with either answer sheets and pencil or with two telegraph keys mounted side by side and labeled "=" and "≠," according to the hand-response assignment prescribed by the design.

At the beginning of the first session, the subject was told the eight stimuli and subsequently listened to a random list of 75 syllables (a single channel from the random tape, presented binaurally). While familiarizing herself with the synthetic stimuli, the subject also tried to identify them by writing down one response for each syllable.

Subsequently, the subject received instructions for the RT task and a few minutes of practice with stimuli taken from the random tape (at long SOAs). If the subject responded too slowly, she was encouraged to respond faster during practice. Before the start of the actual experiment, the subject was instructed in detail about the structure of the stimulus tapes. The property of identical syllables of the same pitch to fuse into a single sound image at SOA = 0 was pointed out. Both speed and accuracy were emphasized. The subject was told to "correct" any errors she realized by immediately pressing the other key, and to keep her fingers resting on the keys during the experiment.

In the identification task, the subject was provided with double-columned answer sheets and instructed to write down two responses for each pair, one for the left ear in the left column and one for the right ear in the right column. A subject whose left ear lagged in time was asked to write down the right ear first. (No subject experienced any difficulty with this right-to-left writing.) The subject was encouraged to write down her responses quickly and to avoid corrections, to guess when uncertain and to use a shorthand notation (a dash in one column) to indicate when she heard two identical consonants.

The subject listened to each block without interruption (ca. 15 min), with short breaks between blocks. The stimulus tapes were played back by a Magnecord 1022 tape recorder to Grason-Stadler TDH-10 earphones. The intensity of the syllables at the earphone was about 83 dB SPL for L syllables and 88 dB SPL for H syllables (peak readings on a sound-level meter). This difference was apparently a consequence of acoustic interactions within the syllables (between formant frequencies and harmonics of the fundamental frequency), since all syllables had precisely the same amplitude parameters during synthesis. Playback intensity was set prior to each experimental session by means of a calibration sine wave on the tapes. The calibration signals had been adjusted prior to the experiment so that identical stimuli from different channels yielded a perfect match on a two-channel storage oscilloscope. Leading and

lagging channels were interchanged electronically as required. The onset of the lagging syllables triggered a Grason-Stadler Model 1270 level/zone detector which in turn started a Gerbrands G1270 digital millisecond clockcounter. The clock was stopped by depressing either of the response keys, and at the same time one of two indicator lights lit up. The latency and origin of each response were written down by the experimenter, who also monitored the tape over earphones.

Analysis

Error percentages (arcsine-transformed) and mean latencies (untransformed) from the "same"-different judgment task were analyzed separately in a seven-way analysis of variance with three between-subject factors (hands, ears, and order) and four within-subject factors (conditions, pitch—same vs. different pitch, feature contrasts—see below, and SOA). Because of the large size of the design, the SOA factor was broken up into *three regions* (I, 0-50-75; II, 100-125-150; III, 250-350-500) which were analyzed separately. (Thus, the analysis did not reflect effects taking place between SOA = 75 and SOA = 100 or between SOA = 150 and SOA = 250; this omission was considered to be of little importance.)

The three linear contrasts of the four-level "feature contrasts" factor were always considered separately, since they are relevant to different theoretical questions. The three contrasts were (a) the *identity contrast*, identical vs. nonidentical pairs; (b) the *similarity contrast*, single- vs. double-feature contrasts; (c) the *voicing-place contrast*, voicing-only contrasts vs. place-only contrasts. When referring to the analysis of variance results, these three contrasts will be abbreviated as FC(1), FC(2), and FC(3), respectively.

Bock's (1975) pseudo-multivariate analysis of variance for repeated measurements was used, with the residual (after taking out the grand mean and the order, hands, ears, and hands by ears effects) and its respective interactions with the within-subject effects serving as error terms.

RESULTS: "SAME"-DIFFERENT JUDGMENTS

Initial Familiarization

Most subjects had little difficulty in identifying the synthetic syllables at first hearing. Eleven subjects scored between 90%-100% correct, three between 80%-90%, and two between 70%-80%. Of the individual syllables, the dentals were more identifiable than the labials. The poorest stimulus was /be_H/ (70% correct), while /be_L/ was much more intelligible (89% correct). This change in intelligibility (of /be/ only) with a change in fundamental frequency was unexpected. It indicates either some highly specific acoustic interaction within the syllable or, perhaps, a change in perceptual boundaries with changes in pitch.

Error Frequencies

Feature Contrasts and SOA. Figure 1 illustrates the effects of feature contrasts and SOA on "same"-different error rates, averaged over all other factors. The average error rates, which were quite substantial, decreased steadily with SOA (I, $p < .0003$; II, $p < .002$) and reached an asymptote at SOA = 250 (III, $p > .06$); the slight decrease between SOA = 250 and the two longer SOAs was still significant when tested separately ($p < .02$).

FC(1), the "identity contrast" (filled circles vs. open symbols in Figure 1), was significant only at the

Figure 1. "Same"-different reaction time task: Error percentages for four kinds of feature contrasts (FCs) at nine SOAs. The FC groups are: identical pairs (0-FC), D/S errors; voicing-only contrasts [1-FC(V)]; place-only contrasts [1-FC(P)]; and voicing-and-place contrasts (2-FC), S/D errors.

shortest SOAs (I, $p < .001$): There were far fewer incorrect "different" (D/S) responses (to identical pairs) than incorrect "same" (S/D) responses (to nonidentical pairs), especially at SOA = 0. The interaction between FC(1) and SOA, i.e., the decrease of this identity contrast with SOA, was highly significant in the region including SOA = 0 (I, $p < .0001$). S/D errors tended to decrease somewhat faster than D/S errors up to SOA = 150 (Figure 1), but this trend did not reach significance.

FC(2), the "similarity contrast" (triangles vs. squares in Figure 1), was extremely pronounced at all SOAs, which confirmed the well-known negative relation between the number of feature contrasts and S/D error rates (I, $p < .0001$; II, $p < .0001$; III, $p < .004$). Although the effect appears to be most pronounced at short SOAs, interactions with SOA did not reach significance.

FC(3), the "voicing-place contrast" (upright vs. inverted triangles in Figure 1), was pronounced only at SOA = 0, where there were substantially more S/D errors for place contrasts than for voicing contrasts (I, $p < .04$). Since this effect was specific to simultaneity, it was mainly reflected in an interaction of FC(3) by SOA (SOA = 0 vs. 50 and 75, $p < .0002$). At SOAs other than SOA = 0, voicing and place contrasts did not differ significantly in their error rates, although voicing contrasts tended to produce more errors at long SOAs (III, $p > .09$).

Conditions (C) and Pitch (P). Effects of these factors on the latencies were hypothesized, but there were no specific expectations with respect to the error rates (except perhaps slightly lower error rates for different-pitch pairs). However, Figures 2 and 3 show that there were quite dramatic changes in error frequencies from one condition to the other, particularly at SOA = 0. For convenience, the three SOA regions will be discussed separately.

At short SOAs (Region I), there was a significant

Figure 2. RT task: D/S error percentages (identical pairs) in four experimental conditions. B = blocked, R = random, P_s = same pitch, P_d = different pitch.

main effect of pitch (I, $p < .0001$) which reflected higher overall error rates for same-pitch pairs, a significant main effect of conditions (I, $p < .0009$) which reflected higher error rates in the random condition, and a significant interaction of C by P (I, $p < .0001$) which indicated a larger effect of pitch in the random condition. These effects must be further qualified, however, by considering their interactions with feature contrasts and SOA. Clearly, the most striking and theoretically important effect was the interaction of P by FC(1) (I, $p < .0001$): Differences in pitch had opposite effects on phonetically identical and non-identical pairs, increasing error rates for the former (D/S errors) and decreasing error rates for the latter (S/D errors), relative to same-pitch pairs. Mainly due to SOA = 0, there were further complex interactions: C by P by FC(1) (I, $p < .0001$), P by FC(1) by SOA (I, $p < .0001$), and C by P by FC(1) by SOA (I, $p < .0001$). At SOA = 0, nonidentical syllables with the same pitch were rarely discriminated in the random condition, while in the blocked condition discrimination accuracy was quite good for these pairs, indicating that binaural fusion of identical same-pitch pairs aided decisions in the blocked situation (Figure 3). Error rates were virtually zero for identical (completely fused) same-pitch pairs at SOA = 0, regardless of condition, but there were many D/S errors with (phonetically) identical different-pitch pairs which did not fuse (Figure 2). Moreover, nonidentical pairs that contained a voicing contrast showed much lower error rates at SOA = 0 than pairs that contrasted only on place (Figure 3). This was especially pronounced in the blocked same-pitch condition (filled circles) and reflects the fact that precisely synchronized syllables that differ only in the direction of their formant transitions tend to be heard as a single syllable, i.e., they fuse almost as completely as identical stimuli. There were several significant interactions reflecting this phenomenon, especially P by FC(3) by SOA (SOA = 0 vs. 50 and 75, $p < .0003$).

In the middle range of SOAs (Region II), there were significant effects of P (II, $p < .02$), P by FC(1) (II, $p < .002$), and C by P by FC(1) (II, $p < .0006$), among

others. In the blocked condition (circles in Figures 2 and 3), pitch had virtually no effect on D/S error rates and only a slight effect on S/D error rates. In the random condition, on the other hand (triangles in Figures 2 and 3), pitch still had a large effect at intermediate SOAs, and this effect was in opposite directions for phonetically identical and nonidentical pairs, as described above. In identical pairs (Figure 2), there was a significant reduction in the random pitch effect between SOA = 100 and SOA = 125—reflected in C by P by FC(1) by SOA (100 vs. 125) ($p < .0006$)—while it remained fairly constant for nonidentical pairs (a reduction occurred there between SOA = 150 and SOA = 250, which is not reflected in significance tests because these SOAs fell in different “regions”).

At long SOAs (Region III), there was still a marginally significant pitch effect (III, $p < .05$) and an interaction of P by FC(1) (III, $p < .02$), and slight effects of pitch on error rates can indeed still be discerned even at SOA = 500 (cf. Figures 2 and 3).

Practice and Laterality Effects. The order (O) factor represents the sequence of the blocked and random conditions; hence, the O by C interaction represents the effect of practice. It was significant in all three SOA regions (I, $p < .02$; II, $p < .003$; III, $p < .006$): The average error rates in the second RT session (actually the third session) were considerably lower than in the first session. This strong practice effect came as a surprise, since two earlier studies (Repp, 1975b, Note 1) had shown virtually no change in errors with practice. Also, the main effect of pitch—but not its interaction with FC(1)—disappeared at longer SOAs with practice, as shown in a significant interaction of O by C by P (II, $p < .04$; III, $p < .005$).

Figure 3. RT task: S/D error percentages (nonidentical pairs) in four experimental conditions. The three different feature contrasts are shown at SOA = 0 only; at other SOAs, average error rates are shown. (See Figures 1 and 2 for symbols.)

There were no interpretable effects connected with the laterality factors, hands and ears.

Intelligibility and Temporal Order Effects. Among the factors not included in the statistical analysis were the effects of fundamental frequency per se in same-pitch pairs (H-H vs. L-L), of the temporal order of different pitches (H-L vs. L-H), of the intelligibility of individual consonants (particularly in identical pairs), and of the temporal order of individual syllables in nonidentical pairs. Some of these effects are of great theoretical interest.

H-H pairs generally showed higher error rates than L-L pairs. This can be explained by the lower average intelligibility of high-pitch syllables (especially of /be/H). There were no consistent differences between H-L and L-H sequences. The expected relation between stimulus intelligibility and D|S error rates (Repp, 1975b) was confirmed: Identical pairs of dental consonants showed generally low error frequencies, while /be/-/be/ had the highest error rates, particularly at high pitch. Also, nonidentical pairs of labial consonants had higher S|D error rates than pairs of dental consonants, although this difference emerged only beyond SOA = 125.

Temporal order effects occurred in pairs that contrasted in place: between SOA = 50 and SOA = 150, /be/-/de/ and /pe/-/te/, i.e., labial-dental sequences, showed much higher S|D error rates than /de/-/be/ and /te/-/pe/, i.e., dental-labial sequences. This effect cannot be related to intelligibility, however, since the double-feature contrast pair /de/-/pe/ exhibited more errors than /pe/-/de/, which, of course, is precisely the opposite effect. Asymmetries of this sort have been observed earlier by the author; they remain puzzling.

Individual Differences. Five of the sixteen subjects (three males, two females) stood out as remarkably accurate. Their error pattern was examined separately but was found to be similar to that of the other, much less accurate subjects.

Corrections. Of all responses, 2.8% were "corrections," with individual rates ranging from .3% to 6.4%. They were not included in the error rates reported above, since they probably were due to response anticipations (see below). Correction frequencies did not covary with the frequencies of "true," uncorrected errors, and seemed not to be related to any of the major independent variables; in particular, they did not decrease with SOA.

Latencies³

Feature Contrasts and SOA. Figure 4 shows the average latencies of correct judgments. RTs decreased steadily and almost linearly with SOA (I, $p < .004$; II, $p < .0001$) and reached their asymptote at about SOA = 300 (by extrapolation). The decrease beyond SOA = 250 was still highly significant ($p < .0001$), but there was no significant change between the two longest SOAs ($p > .84$).

Figure 4. Average latencies of correct responses for four kinds of feature contrasts. Filled circles are "same" responses; open symbols are "different" responses.

At SOA = 0, "same" latencies (filled circles in Figure 4) were clearly faster than "different" latencies (I, $p < .002$), but there was no significant identity contrast at other SOAs.

The similarity contrast, FC(2), was highly significant in all three SOA regions (I, $p < .009$; II, $p < .0001$; III, $p < .002$). As expected, "different" latencies were faster for double-feature contrasts (squares) than for single-feature contrasts (triangles). There were two exceptions, however: (a) At SOA = 0, the difference was reversed (in contrast to the error pattern), as shown in a significant FC(2) by SOA (0 vs. 50 and 75) interaction ($p < .02$). However, the latencies at SOA = 0 were unreliable because of the high error rates. (b) More interesting is the fact that, from SOA = 250 on, there was virtually no difference between the "different" RTs for place-only contrasts and place-and-voicing contrasts (inverted triangles vs. squares in Figure 4).

The voicing-place contrast, FC(3), was also significant and only slightly reduced in the middle region (I, $p < .007$; II, $p < .05$; III, $p < .002$). Voicing contrasts (upright triangles) showed consistently slower latencies than place contrasts (inverted triangles), except at SOA = 0, and the difference seemed to get even larger at long SOAs, which parallels a similar trend in the error rates (Figure 1).

Conditions and Pitch. The effects of these two factors, which were of principal interest, are shown in Figures 5 and 6. As with the error rates, the most dramatic effects were at SOA = 0, but, as noted above, the RTs at this SOA are not very representative. The faster latencies in the blocked same-pitch condition appear to be a reliable finding; the interaction of C by P by SOA (0 vs. 50 and 75) was significant ($p < .004$), reflecting again the presence of binaural fusion as a salient cue. The various other interactions at SOA = 0 need not be discussed in detail, since they largely parallel the interactions in the error rates.

Surprisingly, there were no significant effects involving conditions or pitch at intermediate SOAs. At long

Figure 5. "Same" (S|S) latencies in four experimental conditions.

SOAs, however, there was a significant interaction of P by FC(1) (III, $p < .006$): "Same" latencies tended to be faster when pitch was the same, and "different" latencies tended to be faster when pitch was different. This effect was expected for the random condition but predicted to be absent in the blocked condition. However, the crucial interaction of C by P by FC(1) was disappointingly small, though close to significance (III, $p < .06$). There are indications (cf. Figures 5 and 6) that pitch effects were somewhat larger and more consistent in the random condition, but, in general, they were irregular and much smaller than expected; nor did they completely disappear in the blocked condition. The interaction of C by P by FC(1) by SOA, which is the potential indicator of the time course of "preparation for pitch" in the blocked condition, was far from significance in all SOA regions.

Practice and Laterality Effects. Decreases in RT with practice are common and, therefore, an O by C interaction was expected. Surprisingly, it was significant only at long SOAs (III, $p < .03$). The average decrease in latencies from the first to the second RT session was only about 30 msec.

There was a consistent, marginally significant main effect of the hands factor on latencies (I, $p < .07$; II, $p < .05$; III, $p < .04$): RTs tended to be faster when the right hand responded "same," which is perhaps a more "natural" response assignment for right-handed subjects.

Intelligibility and Temporal Order Effects. Pairs of more intelligible stimuli tended to receive faster responses, in accordance with the lower error rates for these pairs. These differences were especially clear in identical pairs. There were no consistent temporal order effects.

Error Latencies. The average error latencies were based on a rather irregular sampling of conditions, stimuli, and individuals and, therefore, were not analyzed statistically. Their pattern is nevertheless

Figure 6. "Different" (D|D) latencies in four experimental conditions.

worth some attention, because of its potential implications for a model of "same"/"different" decisions: (a) Error latencies were generally longer than correct latencies; (b) D|S latencies tended to be faster than S|D latencies (except perhaps at SOA = 0); (c) S|D latencies for double-feature contrasts were slower than for single-feature contrasts (the opposite was obtained for latencies of correct responses); (d) S|D latencies tended to be slower for voicing contrasts than for place contrasts; and (e) D|S latencies decreased with SOA, while S|D latencies remained approximately constant between SOA = 0 and SOA = 150.

Correction Latencies. The latencies of incorrect responses that were subsequently corrected were significantly *shorter* than the RTs of correct responses for comparable stimuli ($p < .0001$ by sign test), while the "true" error latencies were much slower, as just mentioned. This confirms that corrections were primarily due to anticipations.

DISCUSSION

Two aspects of the data need to be accounted for: the error pattern and the latency pattern. Since error probabilities and latencies tend to be correlated, latencies can be independently interpreted only when either the error rates are near zero (i.e., at long SOAs) or their pattern diverges from that of the errors.

The principal hypothesis of the present study built upon the results of previous studies (especially Springer, 1973) that had shown that variation in fundamental frequency affects "same"/"different" RTs in a randomized design. It was hypothesized that this effect would be reduced or disappear in a blocked design, where uncertainty about pitch is eliminated. No specific effect of this design factor on the frequencies of errors was expected. The outcome was surprising: There were large and systematic variations in the error

rates, while the differences between the average latencies in different conditions were small. This first part of the discussion will deal with the latencies only.

The basic influence of fundamental frequency on RTs was present: "Same" latencies tended to be faster when pitch was the same, and there was also a slight tendency for "different" latencies to be faster when pitch differed. However, these RT differences were considerably smaller than those obtained by Springer (1973) with comparable stimulus material, and they were not consistently present at all SOAs. Most importantly, there were no significant differences between the blocked and random conditions in this respect, nor could any evidence of "preparation for pitch" be seen in the latencies of the blocked condition.

Why was the effect of pitch so weak? Only speculations are possible at this point. The most obvious difference between the present experiment and previous studies is dichotic presentation here vs. binaural presentation there. Since the effect of pitch presumably arises at a precategorical auditory level of comparison (it is difficult to conceptualize pitch as a dimension of phonetic categories), it may be that the auditory memory that preserves pitch becomes relatively less accessible when the stimuli are presented to different ears (or perhaps at different locations, if this is the relevant variable). Morton, Crowder, and Prussin (1971) have demonstrated, in a different paradigm, that "precategorical acoustic storage" is partially ear-specific (or, perhaps, location-specific). Dichotic presentation may force the listeners to reach their "same"- "different" decisions primarily or exclusively at a higher level where pitch differences are no longer represented, i.e., after phonetic categorization or feature extraction, while, in binaural or monaural presentation, such decisions may be made more often at an earlier, auditory stage (Pisoni & Tash, 1974). The same problem has recently emerged in an unsuccessful attempt to replicate the Pisoni and Tash (1974) results (influence of within-category acoustic variations on "same"- "different" RT) with dichotic stimulus presentation (Repp, Note 3). This issue is currently under further investigation by the present author.

The finding of opposite effects of pitch on "same" and "different" latencies, small as they were, clearly differs from the results of Cole and Scott (1972), and especially Cole et al. (1974), who found generally increased latencies when pitch differed. They also found "same" latencies to be faster than "different" latencies, regardless of pitch, while the present study showed no great difference between the two. Cole et al. concluded that syllables in the same voice may be compared at a precategorical level, but syllables in different voices must be converted into a higher order code before a "same"- "different" decision can be reached. The present data and Springer's, on the other hand, support a class of models that permit fast

detection of physical identity or gross physical difference at the precategorical stage, while less clear cases must be resolved after categorization of both stimuli, so that the average latencies represent a mixture of two latency distributions (Pisoni & Tash, 1974). Alternatively, a pitch comparison at the precategorical (auditory) level may precede and influence the speed of subsequent phonetic decisions. (The second model will be adopted in explaining the error pattern.) The small size of the pitch effect obtained in the present study, together with the fact that the latencies failed to mimic the error pattern at shorter SOAs (at least beyond SOA = 75), suggests that the latencies reflect primarily (or exclusively) decisions at the categorical (phonetic) level, i.e., after miscategorizations have occurred and beyond the level at which pitch is represented.

There are two important differences between the Cole et al. experiment and the present study (together with Springer's experiment). One is that Cole et al. used different *voices* and natural speech, while Springer and Repp used synthetic syllables of different *pitch* only. Different voices differ not only in pitch but also in vocal tract size (formant frequencies) and possibly other characteristics, especially in the extreme case of male-female contrast. Syllables with formant patterns appropriate to different vocal tracts may very well require a "normalization" operation before they can be compared at the phonetic level, and this may increase reaction time to such pairs. (This is in agreement with the widely accepted view that the constraints of the production mechanism somehow enter into speech perception.) Differences in pitch, on the other hand, may simply be irrelevant to the process of phonetic categorization and comparison and, therefore, will not increase phonetic decision times. This conclusion has also been reached by Summerfield and Haggard (1973), who varied both pitch and simulated vocal tract size in a choice-RT experiment. They concluded "... that normalization is determined by a formant reference frame, and that, as such, it is an automatic process that occurs whenever the reference frame changes . . ." On the other hand, "... the use of FO [fundamental frequency] information in the process of normalization is a perceptual option dependent upon the particular task, presumably entering at a different level." (p. 7) It is not quite clear which task demands may lead to "pitch normalization," since C. W (Day & Wood, 1971; Wood, 1973, 1975a,b) obtained evidence for such an effect in a choice RT paradigm very similar to that of Summerfield and Haggard. The present data, however, are in agreement with Summerfield and Haggard, who also provided evidence for opposite effects of pitch differences on choice RT in "same" and "different" stimulus sequences. (Concerning the probable equivalence of sequential effects on choice-RT and "same"- "different" judgments, see Eichelmann, 1970; Entus & Bindra, 1970.)

There is another important difference between the Cole et al. (1974) experiment and the present study. Cole et al. used syllables that represented the names of letters in the alphabet, and they found no systematic effects of feature contrasts on "different" latencies. Apparently, their subjects made use of "higher order responses" (letter names) and compared them at a level (lexical or semantic, perhaps even visual) where phonetic feature composition no longer plays a role. The present study shows, in agreement with previous studies using nonsense syllables (Repp, 1975b, Note 2; Weiner & Singh, 1974), that both correct "different" and incorrect "same" latencies reflect the degree of similarity between two syllables, in terms of their phonetic or auditory features. Because of the correlation of phonetic and auditory dimensions in the stimuli used here and previously, the origin of this effect cannot be precisely located in the processing chain, but it clearly reflects the comparison of multidimensional stimulus codes in the brain. Of course, at shorter SOAs there are analogous differences in error frequencies (e.g., more incorrect "same" responses to more similar syllables). There is another finding that does not mirror the error frequencies: From SOA = 250 on, correct responses were equally fast for place-only and place-and-voicing contrasts, but slower for voicing-only contrasts. This suggests that, as soon as dichotic interference ceased, decisions with regard to the place dimension were made faster, or prior to, decisions on the voicing dimension, perhaps because of the greater saliency of the place dimension in the stimuli used here.

It should also be noted that the decreasing latency function reaches its minimum at a longer SOA (between 250 and 360 msec) than the error function (150-250 msec). This suggests that there are processing limitations that exceed the region of direct processing interference. Although other interpretations may be possible, the latency function would support a serial processing theory that postulates that the first syllable must be completely processed before the processing of the second syllable can begin or continue. Accordingly, it would take about 300 msec to process the first syllable—an estimate that is in good agreement with other estimates in the literature (e.g., Massaro, 1972, 1974; Repp, 1975a).

RESULTS: IDENTIFICATION RESPONSES

"Between-Class" Errors

The errors committed in the identification task were divided into "within-class" (S|S and D|D) errors and "between-class" (D|S and S|D) errors. This section deals only with the latter and compares their pattern to that of analogous errors in the RT task. D|S errors were defined as all cases of two different consonants as responses to an identical pair, while S|D errors were all cases of two identical responses to a nonidentical pair.

The pattern of between-class errors was found to be almost indistinguishable from the error pattern in the RT task (Figure 1). The degree of similarity was investigated in a joint analysis of variance, with task (T) as a within-subject factor and condition as a between-subject factor. Only the error data from the first RT session were considered in this comparison, so that the task factor was confounded with practice and familiarity effects but allowed a within-subject comparison. Therefore, given the large practice effect on accuracy in the RT task, it was not surprising to find a clear decrease in between-class errors from the first to the second session (I, $p < .009$; II, $p < .006$; III, $p < .02$). Although this effect may in part be due to a change in task demands, it probably reflects just general practice and familiarity with the stimulus material. Also, the T by P interaction was significant in all three SOA regions, but especially at long SOAs (I, $p < .02$; II, $p < .02$; III, $p < .0001$). It represents a larger decrease in errors for same-pitch pairs than for different-pitch pairs, leading to a reduction in the main effect of pitch. Since a similar trend was observed within the RT task (see above), this effect is likely to be a result of practice too. No other interactions reached significance. In short, there was no clear indication of any task-specific change in error patterns.

"Within-Class" Errors

Figure 7 shows the pattern of the remaining identification errors, viz., those responses that presumably would have led to correct judgments in the RT task (S|S and D|D errors). The S|S errors were defined as all cases of two identical incorrect responses to an identical pair, while D|D errors were all cases of two nonidentical responses, at least one of which was incorrect, to a nonidentical pair. The pattern in Figure 7 differs radically from that in Figure 1. First of all, S|S errors were extremely rare and did not vary with SOA. Naturally, the difference between S|S and D|D error rates was highly significant (I, $p < .0001$; II, $p < .0001$; III, $p < .002$). Second, D|D errors decreased with SOA (II, $p < .02$) but showed an inverted

Figure 7. Identification task: S|S and D|D error rates.

"similarity contrast" at short SOAs: Double-feature contrasts received more incorrect responses than single-feature contrasts ($I, p < .004$). One interesting difference between Figures 1 and 7 is that the corresponding effect on S|D errors seems to persist over the whole SOA range (with opposite sign), while it disappears at SOA = 100 for D|D errors [II, $p > .32$ for the similarity contrast, FC(2)].

As might be expected from the complementary relation between S|D and D|D errors, there were also effects of conditions and pitch on within-class errors, particularly at short SOAs. At short SOAs, there was a significant main effect of pitch ($I, p < .007$), a P by FC(1) interaction ($I, p < .004$), a C by P by FC(1) interaction ($I, p < .03$), plus several other effects. Of all these, only the interaction of C by P by FC(1) was also found at longer SOAs (II, $p < .04$; III, $p < .006$): There was no consistent effect of pitch in the blocked condition, but in the random condition different-pitch pairs exhibited higher error rates. At SOA = 0, place-only contrasts with the same pitch had very low error rates in both conditions, which again reflects the strong tendency of these stimuli to fuse and to lead to S|D errors rather than D|D errors. The majority of the few S|S errors that occurred derived from the random condition (particularly from same-pitch pairs).

Total Error Rates

A brief look at the total error frequencies (or at their complement, the double-correct responses, which would be the customary way of presenting dichotic identification data) showed that the general decrease in errors with SOA tended to asymptote at SOA = 250 except for voicing contrasts, for which accuracy increased up to SOA = 500. Accuracy for identical pairs was much higher than for nonidentical pairs and increased only slightly—but reliably—with SOA. The most interesting finding was an inversion of the "feature sharing advantage" (Pisoni & McNabb, 1974): Except at SOA = 0, where voicing-only contrasts showed the highest accuracy, overall performance with double-feature contrasts tended to be superior to performance with single-feature contrasts. The error rates in the random condition were uniformly higher than in the blocked condition. In both conditions, same-pitch identical and different-pitch nonidentical pairs tended toward lower error frequencies than stimuli with conflicting dimensions.

Transpositions

The scores in a fixed-order two-response dichotic identification task, such as the present one, may be computed with or without the restriction that the responses must have been assigned to the correct channels. (In the present case, this made little difference.) However, response transpositions are of interest in their own right. In the present task, the sequence of the channels was fixed, so that failures

Figure 8. Identification task: Relative percentages of transpositions in double-correct responses.

to assign the stimuli to their correct origin and failures to get their temporal order correct (at SOAs other than zero) led to the same outcome. The two questions of interest were (a) whether two *simultaneous* stimuli (SOA = 0) could be assigned to their respective channels with above-chance success (chance = 50% transpositions), and (b) whether the relative frequencies of transpositions were related to the similarity between the two stimuli.

Figure 8 shows the results. Transpositions are shown as the percentage of all double-correct responses (to pairs of nonidentical syllables, of course) that were in incorrect order.⁴ They can be seen to decline at the same rate as the error rates as SOA gets longer. The two results are clear: At SOA = 0, transpositions were less frequent than expected by chance, and transpositions were less frequent in double-feature contrasts than in single-feature contrasts throughout (but especially at SOA = 0). Also, there were more transpositions in voicing-only contrasts than in place-only contrasts. (In terms of *absolute* frequencies, there were actually twice as many voicing-contrast than place-contrast transpositions at SOA = 0, because of the much smaller number of double-correct responses to place-only contrasts at SOA = 0, cf. Figure 1.) Finally, there was also a tendency toward lower relative transposition frequencies in different-pitch pairs.

The Lag Effect

The lag effect, the well-documented relative dominance of the lagging over the leading syllable, has been reported in ear-monitoring (single-response) experiments (Kirstein, 1973; Repp, 1975a) as well as in two-response paradigms (Berlin, Lowe-Bell, Cullen, Thompson, & Loois, 1973; Studdert-Kennedy, Shankweiler, & Schulman, 1970). The present experiment allowed two further looks at the lag effect:

By requiring the listeners to write down the leading syllable first, a bias against the lag effect was introduced (e.g., the second response may be lost from memory while the first is being written down), and it was of interest whether the lag effect would be strong enough to overcome this bias. The present study also allowed, for the first time, examination of whether a lag effect exists in *identical* pairs (i.e., in D|S errors) or in S|D errors.

The results were fairly clear-cut and are shown in Figure 9. Nonidentical pairs, particularly place contrasts, showed lag effects in S|D errors up to about SOA = 150, while identical pairs (D|S errors) showed no lag effect at all but a slight "lead effect" (Figure 9, upper panel).⁵ Surprisingly, however, D|D errors showed no consistent lag effect (Figure 9, lower panel).

The Right-Ear Advantage (REA)

This experiment also allowed for the first time an assessment of the REA in identical pairs (D|S errors). The surprising result was that *only* identical pairs showed a REA (4% absolute difference) which persisted over the first 150 msec of SOA, while there was no consistent REA for nonidentical pairs (S|D and D|D errors), not even at SOA = 0. (Double-feature contrasts even tended to show a left-ear advantage.) D|S errors showed a REA regardless of which ear lagged, and the REA was greater for same-pitch pairs (except at SOA = 0 where virtually no D|S errors were committed in same-pitch pairs).

DISCUSSION

The first result that deserves comment is the close similarity of the (between-class) error patterns in the RT and identification tasks. Prior to the experiment, it was considered possible that the listeners would adopt different processing strategies in the two tasks, which would be reflected in the error frequencies. In the RT task, where the time pressure was more severe and explicit identification was not required, the subjects might have made use of fast judgments at the pre-categorical level (cf. Pisoni & Tash, 1974), while, in the identification task, such judgments are of little use. It must be conceded that all subjects experienced the RT task before the identification task and, therefore, may have maintained certain strategies, such as making implicit "same"-*"different"* judgments which constrained the subsequent identification responses. The other possibility is that the reverse was true and the subjects based their "same"-*"different"* judgments in the RT task on implicit identifications of the stimuli, i.e., that these judgments were made at the categorical level. The data permit both interpretations, except for one crucial result: the relationship between stimulus intelligibility and incorrect "different" responses to identical pairs. The significance of this relationship has been pointed out earlier (Repp, 1975b, Note 2). It is proof that at least a substantial part of

Figure 9. Identification task: (a) Upper panel, the lag effect (correct score for lagging syllables minus score for leading syllables), S|D and D|S errors only. (b) Lower panel, the lag effect in D|D errors.

"same"-*"different"* judgments is made at the categorical level. Since similarly conclusive evidence for pre-categorical judgments is absent, it is more parsimonious to conclude that both "same"-*"different"* judgments and identification responses were based on the same categorical stimulus codes. As pointed out earlier, these codes should be conceptualized as multidimensional (feature) representations.

Most likely, then, the effects of pitch and of its predictability on the error frequencies have a common explanation in the two tasks. Almost by definition, fundamental frequency—a "nonlinguistic" dimension (Day & Wood, 1971)—is not represented at the categorical level, so that its influence must derive from an earlier stage in processing. One effect that a pitch difference might have is to increase the auditory distance or "separability" of two dichotic stimuli. It clearly eliminates binaural fusion at SOA = 0, and it also increases the effectiveness of selective attention to one ear (Halwes, 1969). However, there is little reduction in the overall error rates, due to the higher probability of incorrect "different" responses in different-pitch pairs, and the small effect that was observed in the present study disappeared with practice.

A much more powerful effect is the bias that sameness and difference of pitch exert on (presumably) phonetic "same" and "different" judgments in the random condition. Of course, it presupposes a certain baseline rate of errors, i.e., there must be *uncertainty* about decisions at the phonetic level (due to dichotic interference) in order for variations in pitch to exhibit any effect. (However, in the absence of errors, the latencies may well reflect precisely the same bias.) This effect is best explained as a true decisional bias: "Same"-*"different"* decisions about pitch can be made at an earlier processing stage than phonetic decisions and, therefore, will precede the latter, unless the

pitch difference is difficult to detect. Presumably, decisions about pitch are made automatically and subsequently feed into the same response selection mechanism as phonetic decisions. Thus, the "same"- "different" response mechanism is "primed" by the pitch decisions and in turn influences the ongoing categorization process. Since pitch decisions always have a high degree of certainty, but the confidence in the phonetic decisions involved in categorization is reduced as a consequence of dichotic interference, pitch decisions will bias the phonetic response selection. In other words, uncertain phonetic decisions will be biased in favor of an identification response that satisfies the primed state of the response mechanism. (By extension of this argument, it may be speculated that differences in the magnitude of pitch effects on latencies in different studies may derive from different levels of uncertainty at the phonetic level; this applies especially to studies using synthetic speech, which often is associated with categorization uncertainty in naive listeners.)

It remains to be explained why the biasing effect of pitch on error frequencies disappeared in the blocked condition. One possibility is that the automatic precategorical decisions about pitch are suppressed or adapted when there is no variation (i.e., when all are "same" or all are "different"). On the other hand, the "same"- "different" decision criteria at the phonetic level may shift and compensate for any *predictable* bias. (This presupposes that there is uncertainty not only at the categorization stage but also in the comparison of phonetic codes. This is plausible, since phonetic comparisons may be attempted before final decisions have been reached about the identity of the stimuli, if the uncertainty is high.) These two possibilities are not mutually exclusive and may both be true. The important prerequisite is uncertainty at the categorical level. Uncertainty arises because of incomplete evidence about the identity of the stimuli, a consequence of dichotic interference. Incomplete evidence means two or more possible category judgments for a stimulus. The more equiprobable candidates there are for an identification response, the longer the process of phonetic categorization will take (since it is itself a decision process), and the more susceptible it will be to external biases. In the extreme case, there will be no evidence at all for a stimulus, and a random guess will be made. Of course, such guesses will be most strongly influenced (and perhaps completely determined) by the biasing influence of pitch decisions.

The final issue to be discussed is the nature of dichotic interference itself. The present study provides some new evidence on several aspects of dichotic interaction. Most likely, this interaction takes place at several levels. At $SOA = 0$, binaural fusion may occur in varying degrees, depending on the acoustic similarity of the syllables. Identical syllables fuse completely,

place-only contrasts almost completely, voicing contrasts to a lesser degree, and syllables of different pitch or at temporal offsets do not fuse at all. This kind of fusion is basically defined only phenomenologically (hearing a single stimulus), and the effects of pitch differences and SOA identify it as a relatively low-level auditory phenomenon. (In Cutting's, 1972, classification of fusions, these are "lower level" fusions, including "localization" and "psychoacoustic fusion.")

In other cases, although phenomenological fusion is absent and the listener is aware that two separate stimuli have been presented, they may nevertheless sound like the same syllable (consonant) to him, leading to "same" judgments (or two identical identification responses) which often may be incorrect. As SOA is shortened, incorrect "same" responses (SID errors) increase at a faster rate than incorrect "different" responses (Repp, 1975b, Note 2), and this tendency was also observed in the present experiment (cf. Figure 1). This perceptual "integration" (Repp, 1975b) or failure to differentiate may occur at temporal separations up to about 100-150 msec, as judged from the different slopes of the error functions for the two types of errors, and it may also occur for syllables of different pitch. Therefore, it is probably a higher level phenomenon, arising at the categorical (phonetic) level as a consequence of information overload, and it is basically the same as "phonetic feature fusion" or "blending" (Cutting, 1972), except that perceptual "dominance" of one stimulus should be included as a possibility. It may be operationally defined as the occurrence of a single phonetic categorization in the presence of two stimuli, the outcome of which may be a correct response to either of the two stimuli or a response combining features of the two stimuli. Its cause may be genuine integration of information prior to, or during, categorization; it also may be the result of processing interruption, i.e., one of the two categorization processes (usually the one started by the leading syllable) is aborted and results in no categorical code.

However, in the author's view, the most important source of errors in dichotic listening arises from uncertainty and confusions in the process of categorization (Repp, 1975a). Integration is only the limiting case, where only 50% of the information present (equivalent to a single stimulus) is actually processed phonetically. Between this extreme (which may be the rule at $SOA = 0$) and perfect identification of both stimuli, there will be all possible degrees of information loss. Typically, there will be more or less uncertainty about the category of one or both stimuli, and *confusions* will arise. Since confusions are responses based on partial information, they will tend to share some features with the correct response. Such confusions may be assumed to occur independently for each stimulus, as long as no lower level fusion or integration occurs, but one stimulus may

be affected more than the other one. The actual mechanism that brings about the uncertainty may be conceptualized either as a division of processing effort between two partially parallel categorization processes (cf. Kahneman, 1973; Norman & Bobrow, 1975) or, perhaps, as the consequence of decay of information held in a temporary store during a serial categorization process. In each case, "perceptual noise" (Repp, 1975b) will be introduced and lead to uncertainty and confusions. Note that especially incorrect "different" responses to physically identical stimuli must be the consequence of such perceptual imprecision.

If this confusion model (which is open to a formulation in terms of discrete features) is chosen as a heuristic, many results fall neatly into place. To begin with, effects of stimulus intelligibility are trivially accounted for, since they imply a priori variations in categorization uncertainty. S|S errors in identification are so rare because the probability that two independent confusions will occur *and* be identical is small. D|S errors are more frequent because, typically, only one of the two stimuli will be misidentified. D|D errors are the errors that are reported in most dichotic listening studies in the literature, which typically presented only nonidentical pairs to subjects who knew that all pairs contained two different syllables. Such errors show a feature sharing advantage (Pisoni & McNabb, 1974), i.e., more errors for double-feature contrasts than for single-feature contrasts, which is also demonstrated by the present study. S|D errors, on the other hand, show the inverse pattern. This is predicted by the confusion model: First, it predicts that, in pairs of "similar" syllables, the confusion response to one syllable will be more likely to be identical with the other (correct) response than in pairs of less similar syllables. Hence, S|D errors will be more frequent when there are fewer feature contrasts. (In single-response paradigms, however, there will be more *correct* responses to pairs of similar syllables because confusion-intrusions add to the number of correct responses, cf. Pisoni & McNabb, 1974; Repp, 1975a). By reciprocity, D|D errors will be relatively less frequent in pairs of more similar syllables. The same is true for the absolute error frequencies, as long as there are no gross differences in total error rates between single- and double-feature contrasts. However, an overall advantage for double-feature contrasts was found here, which suggests that their higher discriminability leads to a genuine processing advantage in dichotic competition.

This explanation suggests that whether or not a feature sharing advantage is obtained in a dichotic study will depend on the composition of the stimuli and on the kinds of responses asked for. In a typical dichotic situation, S|D errors are not permitted and the subject will guess instead. Hence, the second response (in a free-order two-response task) provides considerably

less information than the first response (Halwes, 1969; Studdert-Kennedy & Shankweiler, 1970). By permitting S|D responses, guessing is reduced and failures to discriminate are revealed. (It may depend on the purpose of the experiment whether or not this is a methodological advantage; certainly, it was one in the present context.)

One finding that is not so easily explained is the apparent disappearance of the effect of feature contrasts on D|D errors at SOA = 100 (cf. Figure 7), while the corresponding effect on S|D errors persists over the whole SOA range. It may reflect perceptual integration or "phonetic feature fusion" (Cutting, 1972), since its temporal range is similar to that of the increased frequency of S|D errors that led to the concept of "integration" in this context (Repp, 1975b). Still, from the confusion model somewhat more D|D errors would be expected for double-feature contrasts than for single-feature contrasts, even at longer SOAs. The absence of a difference, then, probably again reflects a genuine perceptual advantage for pairs of more dissimilar stimuli. This conclusion is contrary to that often derived from a feature sharing advantage which, in the present view, is a consequence of the confusion pattern and not a true perceptual advantage. It is in agreement, however, with effects of target-mask similarity in visual masking (e.g., Schiller, 1966; Turvey, 1973).

Even if two syllables are correctly identified, they may be assigned to incorrect channels, so that the responses are transposed. This source of errors is probably independent of the other interference phenomena discussed above and occurs after identification of both stimuli. The present experiment demonstrates that, even at SOA = 0, the two stimuli are correctly assigned to their respective channels with a probability that is above chance. This indicates that the incoming stimuli are "tagged" as to their origin and that these tags can be recovered with some success. The other interesting finding is that the relative frequencies of transpositions are higher in single-feature contrasts than in double-feature contrasts. It should be noted that some of these "transpositions" will actually be double confusions of the two stimuli "with each other" (i.e., both confusion responses happen to be correct responses to the other syllable in the pair). The confusion model predicts that such double confusion will be more frequent in single-feature contrasts. However, the probability of double confusions (and particularly of those that simulate transpositions) is too low to explain the large difference observed here (Figure 8). It seems, then, that the mechanism that is responsible for tagging the stimulus information with respect to its origin is sensitive to the (auditory or phonetic) similarity of the stimuli. The same is probably true for temporal order discrimination, since, at SOAs other than zero, temporal order was confounded with the origin of the two stimuli (one ear always preceded the other). An interesting parallel may be noted in

studies of short-term memory for lists of verbal material: There, too, more similar adjacent stimuli are more often transposed than more dissimilar stimuli (Conrad, 1964; Wickelgren, 1965), and this loss of order information in memory has also been considered to be independent of the loss of stimulus information (Estes, 1972), just as in the present case. Therefore, dichotic transpositions and memory transpositions may well reflect the same mechanism.

In the present identification task, a clear lag effect—more correct responses to the lagging syllable than to the leading syllable—was observed, and it was strong enough to persist when the response to the leading ear was written down first. However, this effect was present only in S/D errors; it was shown neither by D/S errors nor by D/D errors. This suggests that the lag effect may be specifically associated with *failures to discriminate* two dichotic stimuli. This is in agreement with an interruption interpretation of the lag effect (Studdert-Kennedy et al., 1970; Kirstein, 1973; Porter, 1974). The second stimulus interrupts or delays the processing of the first, so that evidence about the leading stimulus is lost and it will sound similar to the lagging stimulus, typically resulting in a “same” response. On the other hand, whenever two different stimuli are perceived (incorrect as they may be), they are likely to originate from two separate phonetic categorizations. In this view, the lag effect arises during the process of categorization. However, an alternative explanation is possible (Repp, 1975a): There may be a time gradient in phonetic feature fusion or integration that makes the lagging stimulus perceptually more dominant, perhaps of partial decay of the (not yet categorized) leading stimulus from a temporary memory (cf. Massaro, 1972, 1974). Again, however, this effect would have to be localized in the categorization process, since it presupposes the extraction of phonetic features. The second explanation fits better into the author’s view of dichotic interference, although both explanations may be formally equivalent.

Finally, the absence of a right-ear advantage for nonidentical pairs should not be overinterpreted, although it remains a puzzling result which follows several previous failures to obtain right-ear advantages in similar tasks (McNeill & Repp, 1973, and unpublished research; Repp, 1975a). No explanation has been found as yet, but the evidence from other laboratories for a dichotic right-ear advantage is so overwhelming that it cannot be challenged by these few negative results.

A MODEL OF DICHOTIC INTERACTION

Instead of summarizing the detailed experimental findings, the model of dichotic interference that has emerged from these (and earlier) results will be reviewed briefly. This model assumes the existence

of two distinct processing stages: an *auditory* or precategorical stage and a *phonetic* or categorical stage (cf. Pisoni, 1975; Studdert-Kennedy, in press). These stages are at least partially sequential.

At the auditory level, dichotic stimuli may interact by fusing with each other or by masking each other. (Auditory masking seems to play a relatively minor role.) *Fusion* may occur for simultaneous stimuli, either in the form of complete binaural fusion of identical stimuli (and especially of identical vowel portions) or as psychoacoustic (“phenomenological”) fusion of nonidentical stimuli (cf. Cutting, 1972). In each case, only a single stimulus will be heard. Temporal offset or differences in fundamental frequency (pitch) prevent these fusions from occurring.

If pitch is variable, implicit judgments will be made at the auditory level about whether this dimension is the same or different in the two stimuli. This automatic registration will “prime” the corresponding response and will *bias* the phonetic decision mechanism in the direction of this response. Pitch is assumed not to be represented at the phonetic level.

At the phonetic level, three potential sources of errors in the dichotic listening task are distinguished: integration, confusions, and transpositions. Prior to or during phonetic categorization, the information representing the two stimuli may be *integrated*, resulting in a single input to the phonetic categorizing mechanism. The resulting phonetic code may represent one of the two stimuli (“dominance”) or a mixture of them (“phonetic feature fusion”—cf. Cutting, 1972). In this case, the listener, although he is aware that two stimuli were presented, will arrive at a “same” judgment or give two identical identification responses. Integration may occur for stimuli of different pitches and over temporal separations of up to about 150 msec. Also, a time gradient may exist that favors the lagging stimulus, leading to a lag effect.

If two stimuli are not integrated, they will be *categorized* separately and independently, perhaps in parallel. Because of processing limitations, the categorization process will be slowed down and lose precision, so that *uncertainty* will exist. Through *uncertainty*, *confusions* in the selection of appropriate categories will arise. Confusions will tend to share features with the correct responses.

The categories that are finally selected will be influenced by the outcome of implicit pitch decisions at the auditory level before they are compared and a phonetic “same”-“different” decision is made. Alternatively, phonetic codes may be compared in a state of incompleteness or uncertainty, and the bias of implicit pitch decisions directly influences the phonetic “same”-“different” decision mechanism. Such a bias will exist only when pitch decisions are unpredictable; when they are predictable, they exert no bias on phonetic decisions. In the absence of errors, reaction

times may reflect similar biases. Reaction times also reflect the multidimensional nature of phonetic codes.

In an identification task, pitch variation will bias the responses to the same degree as in a phonetic discrimination task. However, there is an additional source of errors (in addition to the incidence of incorrect identification responses that would lead to correct "same"-"different" judgments): Two responses may be *transposed* (after they are categorized) and assigned to incorrect channels (temporal order discrimination, right-left discrimination). Transpositions seem to be more frequent when the two stimuli are similar to each other, just as in short-term memory.

Clearly, this model is vague and post hoc. However, it may lend itself to formalization (e.g., as a mathematical confusion model), it makes a number of testable predictions, and it seems to account for a large amount of dichotic listening data. Future research will have to revise and refine this preliminary attempt.

REFERENCE NOTES

1. Reed, C. M. Same-different reaction times to pairs of nonsense syllables. Paper presented at the 87th meeting of the Acoustical Society of America, New York, April 1974.
2. Repp, B. H. Discrimination of dichotic speech sounds: Effects of distinctive feature contrasts and temporal separation. Unpublished paper.
3. Repp, B. H. Categorical perception, dichotic interference, and auditory memory: A "same"-"different" reaction time study. Paper presented at the 89th meeting of the Acoustical Society of America, Austin, Texas, April 1975.

REFERENCES

- BERLIN, C. I., LOWE-BELL, S. S., CULLEN, J. K., JR., THOMPSON, C. L., & LOOVIS, C. F. Dichotic speech perception: An interpretation of right-ear advantage and temporal offset effects. *Journal of the Acoustical Society of America*, 1973, **53**, 699-709.
- BLUMSTEIN, S., & COOPER, W. E. Identification versus discrimination of distinctive features in speech perception. *Quarterly Journal of Experimental Psychology*, 1972, **24**, 207-214.
- BOCK, R. D. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill, 1975.
- COLE, R. A., COLTHEART, M., & ALLARD, F. Memory of a speaker's voice: Reaction time to same- or different-voiced letters. *Quarterly Journal of Experimental Psychology*, 1974, **26**, 1-7.
- COLE, R. A., & SCOTT, B. Distinctive feature control of decision time: Same-different judgments of simultaneously heard phonemes. *Perception & Psychophysics*, 1972, **12**, 92-94.
- CONRAD, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55**, 75-84.
- COOPER, L. A., & SHEPARD, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973. Pp. 75-176.
- CUTTING, J. E. A preliminary report on six fusions in auditory research. *Haskins Laboratories Status Report on Speech Research*, 1972, SR-31/32, 93-107.
- DAY, R. S., & WOOD, C. C. Interactions between linguistic and nonlinguistic processing. *Haskins Laboratories Status Report on Speech Research*, 1971, SR-27, 185-192.
- EICHELMAN, W. H. Stimulus and response repetition effects for naming letters at two response-stimulus intervals. *Perception & Psychophysics*, 1970, **7**, 94-96.
- ENTUS, A., & BINDRA, D. Common features of the "repetition" and "same"-"different" effects in reaction time experiments. *Perception & Psychophysics*, 1970, **7**, 143-148.
- ESTES, W. K. An associative basis for coding and organization in memory. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory*. Washington, D.C.: Winston, 1972. Pp. 161-190.
- GARNER, W. R. *The processing of information and structure*. Potomac, Md: Erlbaum, 1974.
- HALWES, T. G. Effects of dichotic fusion on the perception of speech. Unpublished doctoral dissertation, University of Minnesota, 1969.
- KAHNEMAN, D. *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- KIRSTEIN, E. F. The lag effect in dichotic speech perception. *Haskins Laboratories Status Report on Speech Research*, 1973, SR-35/36, 81-105.
- MASSARO, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, **79**, 124-145.
- MASSARO, D. W. Perceptual units in speech recognition. *Journal of Experimental Psychology*, 1974, **102**, 199-208.
- MCCNEILL, D., & REPP, B. Internal processes in speech perception. *Journal of the Acoustical Society of America*, 1973, **53**, 1320-1326.
- METZLER, J., & SHEPARD, R. N. Transformational studies of the internal representation of three-dimensional objects. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium*. Potomac, Md: Erlbaum, 1974.
- MORTON, J., CROWDER, R. G., & PRUSSIN, H. A. Experiments with the stimulus suffix effect. *Journal of Experimental Psychology Monograph*, 1971, **91**, 169-190.
- NORMAN, D. A., & BOBROW, D. G. On data-limited and resource-limited processes. *Cognitive Psychology*, 1975, **7**, 44-64.
- PISONI, D. B. Dichotic listening and processing phonetic features. In F. Restle et al. (Eds.), *Cognitive theory: Volume I*. Potomac, Md: Erlbaum, 1975.
- PISONI, D. B., & McNABB, S. D. Dichotic interaction of speech sounds and phonetic feature processing. *Brain and Language*, 1974, **1**, 351-362.
- PISONI, D. B., & TASH, J. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 1974, **15**, 285-290.
- PORTER, R. J., JR. The dichotic lag effect: Implications for the central processing of speech. *Proceedings of the Speech Communication Seminar*, Stockholm, August 1974.
- POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79**, 1-16.
- REPP, B. H. Dichotic forward and backward "masking" between CV syllables. *Journal of the Acoustical Society of America*, 1975, **57**, 483-496. (a)
- REPP, B. H. Distinctive features, dichotic competition, and the encoding of stop consonants. *Perception & Psychophysics*, 1975, **17**, 231-240. (b)
- SCHILLER, P. H. Forward and backward masking as a function of relative overlap and intensity of test and masking stimuli. *Perception & Psychophysics*, 1966, **1**, 161-164.
- SPRINGER, S. P. Memory for linguistic and nonlinguistic dimensions of the same acoustic stimulus. *Journal of Experimental Psychology*, 1973, **101**, 159-163.
- STUDDERT-KENNEDY, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. Springfield, Ill: Thomas, in press.
- STUDDERT-KENNEDY, M., & SHANKWEILER, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579-594.
- STUDDERT-KENNEDY, M., SHANKWEILER, D., & SCHULMAN, S.

- Opposed effects of a delayed channel on perception of dichotically and monotically presented CV syllables. *Journal of the Acoustical Society of America*, 1970, **48**, 599-602.
- SUMMERFIELD, A. Q., & HAGGARD, M. P. Vocal tract normalization as demonstrated by reaction times. In *Speech Perception*, 1973, Series 2, No. 2. Department of Psychology, The Queen's University of Belfast. Pp. 1-12.
- TURVEY, M. T. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, **80**, 1-52.
- TVERSKY, B. Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, 1969, **6**, 225-233.
- WEINER, F. F., & SINGH, S. Multidimensional analysis of choice reaction time judgments on pairs of English fricatives. *Journal of Experimental Psychology*, 1974, **102**, 615-620.
- WICKELGREN, W. A. Short-term memory for phonemically similar lists. *American Journal of Psychology*, 1965, **78**, 567-574.
- WOOD, C. C. Parallel processing of auditory and phonetic information in speech perception. *Haskins Laboratories Status Report on Speech Research*, 1973, SR-35/36, 39-53.
- WOOD, C. C. Levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **104**, 3-20. (a)
- WOOD, C. C. A normative model for redundancy gains in speech discrimination. In F. Restle et al. (Eds.), *Cognitive theory: Volume I*. Potomac, Md: Erlbaum, 1975. (b)
- WOOD, L. E. Visual and auditory coding in a memory matching task. *Journal of Experimental Psychology*, 1974, **102**, 106-113.
- in the studies of Cole and Scott (1972) and Cole et al. (1974), since real male and female voices were used. Springer (1973), on the other hand, presented synthetic syllables that differed *only* in fundamental frequency, which apparently was sufficient to have a strong effect on RTs. As in Springer's experiment (and as in C. Wood's studies mentioned below), the synthetic syllables of the present study varied only in pitch, holding simulated vocal tract size constant.
2. Blumstein and Cooper (1972) "tricked" their subjects into responding "same" or "different" to dichotic syllables that were in fact never identical. The subjects made a substantial number of "same" responses, especially when the syllables differed only on a single feature. But because of the deception technique employed, these error rates are not useful for our purpose of comparing two paradigms.
3. The latencies shown in the figures are weighted means (i.e., adjusted for different error rates), while the statistical analysis was necessarily conducted on unweighted means. However, the difference is likely to be negligible.
4. Of course, transpositions may also occur when one or both responses are incorrect. There was some supporting evidence in the data, but basically a mathematical model is needed to determine precisely which percentage of responses is due to transpositions. Therefore, we have restricted the present discussion to correct responses, where the probability of "pseudotranspositions" is low.
5. Theoretically, all differences at SOA = 0 should have been zero, or at least in the same direction (in the case that a slight imbalance between channels existed in spite of the careful control of the playback intensities). Nevertheless, non-identical pairs showed a slight lag effect and identical pairs a slight lead effect at SOA = 0. In view of this unexplained variation, the lag effect for place-only contrasts (SID errors) may be the only reliable one.

NOTES

1. It should be noted that the two principal dimensions that distinguish different speakers, pitch (fundamental frequency) and vocal tract size (formant frequencies), covaried

(Accepted for publication May 30, 1975.)