

## Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study<sup>1</sup>

MICHAEL STUDDERT-KENNEDY,<sup>2</sup> DONALD SHANKWEILER,<sup>3</sup> AND DAVID PISONI<sup>4</sup>  
*Haskins Laboratories, Inc., New Haven*

The distinction between auditory and phonetic processes in speech perception was used in the design and analysis of an experiment. Earlier studies had shown that dichotically presented stop consonants are more often identified correctly when they share place of production (e.g., /ba-pa/) or voicing (e.g., /ba-da/) than when neither feature is shared (e.g., /ba-ta/). The present experiment was intended to determine whether the effect has an auditory or a phonetic basis. Increments in performance due to feature-sharing were compared for synthetic stop-vowel syllables in which formant transitions were the sole cues to place of production under two experimental conditions: (1) when the vowel was the same for both syllables in a dichotic pair, as in our earlier studies, and (2) when the vowels differed. Since the increment in performance due to sharing place was not diminished when vowels differed (i.e., when formant transitions did not coincide), it was concluded that the effect has a phonetic rather than an auditory basis. Right ear advantages were also measured and were found to interact with both place of production and vowel conditions. Taken together, the two sets of results suggest that inhibition of the ipsilateral signal in the perception of dichotically presented speech occurs during phonetic analysis.

Current accounts of speech perception emphasize process and divide the process into a hierarchy of stages: auditory, phonetic, phonological, and so on (see for example, Fry, 1956; Chistovich *et al.*, 1968; Studdert-Kennedy, in press). The distinction between phonetic and higher levels is commonly accepted in linguistic theory and is readily demonstrated in behavior. But the distinction between auditory and phonetic levels is less easily demonstrated and is not widely recognized. The auditory stage (or stages) refers to transformation of the acoustic waveform into a set of time-varying psychological dimensions (pitch, loudness, timbre,

<sup>1</sup>This research was supported by a grant to Haskins Laboratories from NICHD. The main results were reported at the 81st Meeting of the Acoustical Society of America, Washington, DC April 1971.

<sup>2</sup>Also Graduate Center and Queens College, City University of New York.

<sup>3</sup>Also University of Connecticut.

<sup>4</sup>Also Indiana University.

duration) roughly corresponding to dimension measurable in a spectrogram. The phonetic stage refers to transformation of psychological (auditory) dimensions into phonetic features. We have argued elsewhere (Studdert-Kennedy & Shankweiler, 1970) that, while the auditory transformation may be accomplished by the general auditory system common to both cerebral hemispheres, the phonetic transformation is accomplished largely, if not exclusively, by specialized mechanisms in the language-dominant hemisphere.<sup>5</sup>

We will not repeat the argument here. But among the reasons for positing a single phonetic processing system is an interaction between left- and right-ear inputs repeatedly observed in dichotic experiments: the initial stop consonants of dichotically presented CV (or CVC) syllables, differing only in those stops, are more accurately identified if the two segments have a phonetic feature in common (Shankweiler & Studdert-Kennedy, 1967). Figure 1 (based on Table IV of Studdert-Kennedy & Shankweiler, 1970) displays the effect. The probability that both initial stops will be correctly identified is greater if the two segments have the same value on the phonetic features of place (e.g., /ba-pa/) or voicing (e.g., /ba-da/) than if they have neither feature in common (e.g., /ba-ta/).

We have interpreted this interaction as evidence that dichotic speech inputs converge on a single cerebral center before the extraction of phonetic features. We suggested, further, that "duplication of the auditory information conveying the shared feature value gives rise to the observed advantage" (Studdert-Kennedy & Shankweiler, 1970, p. 589). However, there are at least two stages at which the advantage might arise: (1) during extraction of phonetic features from the auditory transforms (the interpretation quoted above); or (2) during output of a response from the phonetic system.<sup>6</sup> The first interpretation attributes

<sup>5</sup> The asymmetrical representation of language in the brain (the left hemisphere being the dominant one in most right-handed persons) is one of the most securely known facts about the physiology of language. Whether all linguistic processes are so lateralized or whether only some are is a matter which is only beginning to be investigated. In two earlier papers (Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970), we have described dichotic rivalry experiments in which lateral differences in perception of minimally differing nonsense syllables were discovered. The consistent finding of a net advantage for the right ear input in most right-handed subjects is evidence that perception of language at the level of phonetic processes is lateralized on the left.

<sup>6</sup> The advantage might also arise during extraction of auditory information from the acoustic signal. The effect would then be due to a relatively low, perhaps subcortical, level of the perceptual process. As will be seen, this possibility, though difficult to test directly, was ruled out by implication from the results of the experiment.

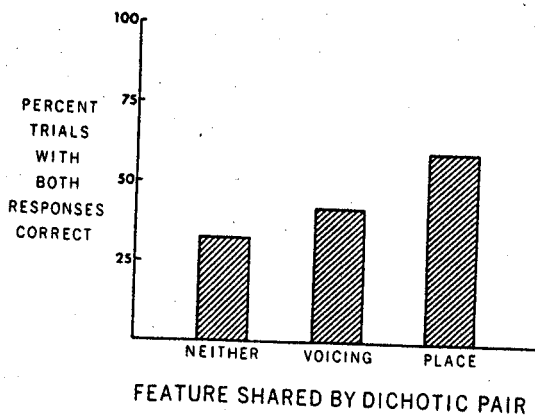


FIG. 1. The percentage of trials on which both responses were correct as a function of the consonantal feature shared by dichotic CV pairs.

the advantage to shared characteristics of the inputs (signals) to the phonetic system: phonetic analysis of the two sets of auditory parameters is facilitated if the two sets have certain auditory features in common. The second interpretation attributes the advantage to shared characteristics of the outputs (messages): correct responses from the phonetic component are facilitated by shared phonetic features rather than by shared auditory features. We should emphasize that a response from the phonetic component is conceptually a stage of the perceptual process and is to be distinguished from any observable motor response that may follow it.

The present experiment was designed mainly to distinguish between these two interpretations. We may clarify the argument by considering the set of syllables used. Table 1 lists four stop consonants (/b,p,d,t/) and their possible combinations into dichotic pairs. Note that there are two pairs sharing place (/b-p/, /d-t/), two sharing voice (/b-d/, /p-t/) and two sharing neither feature (/b-t/, /d-p/). We are most interested in the two pairs sharing place, since it is these that permit us to compare the effects of auditory and phonetic commonalty.

Figure 2 illustrates the comparison. The figure displays stylized spectrographic patterns of the eight synthetic CV syllables used in this study. They are formed from all possible combinations of the four stop consonants (/b,p,d,t/) with two vowels (/i,u/). No release burst was included in the synthesis so that all information concerning place of articulation is conveyed by the second- and, to some extent, third-formant transitions. All within column pairs share both place of consonantal articulation and following vowel: they, therefore, have identical formant

TABLE 1  
Paired Combinations of Four Stop Consonants According to Features  
of Voicing and Place of Articulation

	Place of articulation	
	Labial	Alveolar
	Voiced	b
Unvoiced	p	t
Pairs sharing		
Place alone	Voicing alone	Neither feature
b-p d-t	b-d p-t	b-t d-p

transitions. Cross-column pairs (/bi-pu, bu-pi, di-tu, du-ti/) share place of consonantal articulation but not following vowel: they, therefore, have different formant transitions. In other words, within-column (same vowel) pairs share both phonetic and auditory information; cross-column (different vowel) pairs share only phonetic information. We may now compare performance on these two types of dichotic pair. If the advantage due to sharing a feature has an auditory basis, we would expect the advantage to be greater for place-sharing dichotic pairs that also

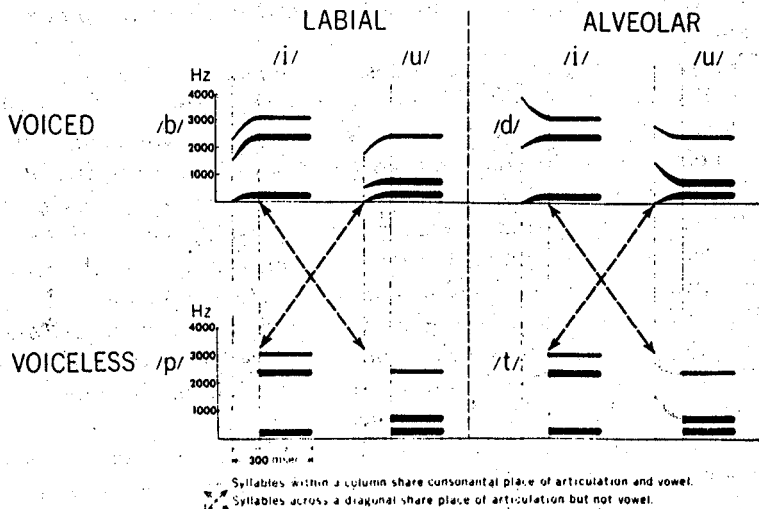


FIG. 2. Schematic spectrograms of eight synthetic CV syllables.

share the same vowel than for corresponding pairs that have different vowels. On the other hand, if the advantage has a phonetic basis, we would expect no difference in performance on these same pairs between the two experimental conditions of "vowel same" and "vowel different."

Finally, a subsidiary purpose of the present experiment was to determine the effect of auditory commonalty or contrast on the right-ear advantage typically observed for stop consonants in dichotic studies. We defer elaboration of this matter to the discussion.

#### METHOD

The eight three-formant CV syllables were synthesized on the Haskins Laboratories parallel resonant synthesizer. Each syllable had a duration of 300 msec: formant transitions lasted 40 msec, steady-state portions, 260 msec. For the voiced consonants all three formants began at the same instant; for the voiceless consonants the first formant was cut back by 70 msec, and the upper formants were aspirated over this period. The pitch contour of each syllable fell linearly from 130 to 90 Hz.

Two dichotic tapes were prepared by a computer-controlled procedure that permits precise alignment of syllable onsets. Voiced/voiceless pairs (i.e., those sharing place: /b-p/, /d-t/) were aligned so that the aspirated formants of one syllable began at the same instant as the voiced formants of the other. On one tape, the vowels of any dichotic pair were the same (either /i/ or /u/); on the other tape, the two vowels were different. There are twelve possible ordered pairs of syllables contrasting in their initial consonants (ordering refers to channel orientation). Each pair occurred 10 times in a randomized test order, with the restriction that each pair occurred five times in the first 60 trials, five times in the second 60 trials.

Sixteen university students volunteered as subjects and were paid for their work. All were right-handed, native speakers of English with no known hearing loss or speech impediment. They were run as four groups over 2 days in a balanced design, distributing all order effects equally over the two experimental conditions. On a given day, the subjects began with an 80-item monaural identification tape, 40 items to the left ear, 40 to the right. They then took a 24-item practice dichotic tape. Finally, they took the assigned test tape twice, reversing earphones after the first run to distribute channel effects equally over the ears. For the dichotic test they were told that the two consonants on any trial would always be different; they were instructed to identify both of them, drawing from the set /b,p,d,t/, to write their answers on a sheet and to give their more confident response first.

One subject scored less than 90% on the monaural identification test

and displayed a strong left-ear advantage on every data analysis. He was omitted from the group analysis, reducing the total number of trials to 1800 or 120 from each of the 15 subjects.

### RESULTS

Figure 3 displays the main results. For both experimental conditions (vowel same, vowel different), the percentage of trials having both responses correct is greater for those dichotic pairs that have a feature in common. The effect is significant by analysis of variance ( $p < .001$ ). In previous studies (cf., Figure 1) more advantage accrued to pairs sharing place than to pairs sharing voicing. Here, there is no significant difference between the two classes of dichotic pair: subjects varied in whether they gave their highest performance on place-sharing or voice-sharing pairs, so that there was significant subject-by-feature-shared interaction ( $p < .0001$ ). No subject gave his highest performance on pairs having no feature in common.

Turning to the result of most interest for the present study, we note that there is no significant effect of the following vowel. The slight advantage for place-sharing pairs that precede different vowels was present for both labial (5%) and alveolar (6%) pairs but was not significant.

Finally, we consider the ear advantages. Table 2 displays the distribution of correct responses over the ears for trials on which only one response was correct (the only trials on which an ear advantage has an opportunity to occur).<sup>1</sup> The columns headed (R - L/R + L) 100 provide

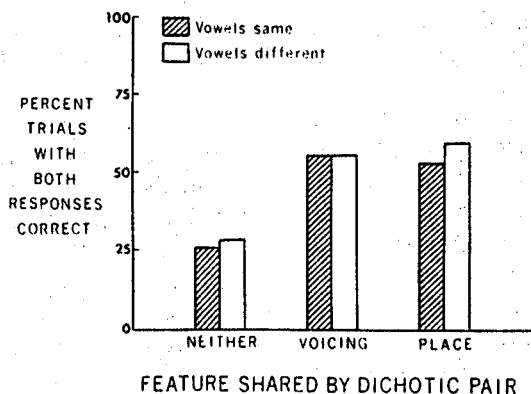


FIG. 3. The percentage of trials on which both responses were correct as a function of the consonantal feature shared by dichotic CV pairs under two vowel conditions.

<sup>1</sup>The greater number of such trials when neither feature was shared is entailed by the smaller number of both correct trials under that condition.

TABLE 2  
Distribution of Correct Responses Over Ears for One-Correct Trials Only

Feature shared	Vowel same		Vowel different				Total		
	R <sup>a</sup>	L <sup>b</sup>	$\frac{(R-L)}{(R+L)} 100$		R	L	R	L	
								$\frac{(R-L)}{(R+L)} 100$	
Place	326	208	22		281	193	607	401	20
Voicing	289	236	10		279	255	568	491	7
Neither	439	329	14		421	331	860	660	13
Total	1054	773	15		981	779	2035	1552	13

<sup>a</sup> R = Number correct on right ear.

<sup>b</sup> L = Number correct on left ear.

TABLE 3  
 Distribution of Correct Responses Over Ears and Place Values for One-Correct Trials Only

Feature value shared	Vowel same		Vowel different				Total	
	R <sup>a</sup>	L <sup>b</sup>	R	L	R	L	R	L
Alveolar	161	87	122	93	13	283	180	21
Labial	165	121	159	100	23	324	221	19
Total	326	208	281	193	19	607	401	20

<sup>a</sup> R = Number correct on right ear.

<sup>b</sup> L = Number correct on left ear.



a measure of the ear advantage: the index ranges from  $-100$  to  $+100$  with negative values indicating a left-ear advantage, positive values a right-ear advantage. All indices are positive and the ear effect is highly significant by analysis of variance ( $p < .001$ ); its variation across feature conditions falls short of significance at the .05 level. There is no reliable difference in the ear effects for the two vowel conditions: the tendency toward a larger laterality index when vowels are the same than when they are different is not significant.

However, analysis of the one-correct data into separate place values reveals complexities: there is significant, three-way interaction between ears, vowel condition, and place value ( $p < .05$ ). Table 3 shows that for alveolar pairs the laterality index is greater when vowels are the same than when vowels are different; for labial pairs the reverse is true. We may note, further, that the alveolar ear-by-vowel interaction is primarily due to a drop in right-ear performance when vowels are different, while the labial ear-by-vowel interaction is largely due to a rise in left-ear performance when vowels are the same. Summing over vowel conditions, we note no significant difference in the laterality effect for the two place values.

#### DISCUSSION

The main outcome is predicted by the phonetic interpretation: the gain in performance for feature-sharing dichotic pairs arises from commonality in the phonetic message rather than in the acoustic signal or its auditory transform. From this we may draw two inferences. First, since commonality in the message is defined at the level of phonetic features and takes effect whether attributable to shared voicing or shared place of articulation, we may infer that distinct phonetic feature processors are engaged during perception. Second, we may infer that activation of a feature processor for one response facilitates its activation for another temporally contiguous response. The last statement might serve to describe a short-term response bias leading to errors of feature substitution in speaking of the kind described by Fromkin (1970). However, in the present instance, repetition of a feature in successive responses is not a random, internally generated error, but is adapted to the particular pair of signals presented: voicing tends to be repeated on pairs that share voicing, place on pairs that share place. The effect, therefore, precedes the observable response and is perceptual.

At the same time, the results justify the distinction between auditory and phonetic processes upon which the experiment was based, since commonality at the two levels affects overall performance and the laterality effect differently. Phonetic feature-sharing facilitates perform-

ance but has little or no effect on the ear advantage. Auditory similarity or contrast affects the ear advantage (Table 3) but not performance (Figure 3). We conclude that phonetic and auditory transformations are indeed distinct processes. Furthermore, the mutual facilitation of responses to feature-sharing dichotic pairs suggests that phonetic transformation is accomplished by a single system of processors<sup>8</sup> to which both inputs have access.

We turn now to the ear advantages. We have argued elsewhere (Studdert-Kennedy & Shankweiler, 1970) that auditory-to-phonetic transformation may be the prerogative of the language-dominant cerebral hemisphere. At the same time, the minor hemisphere is evidently specialized for recognition of complex auditory patterns (Milner, 1962; Kimura, 1964, 1967; Shankweiler, 1966; Darwin, 1969). The interaction between ears, feature-value shared, and vowel condition in the present study (see Table 3) may reflect, in part, this functional dissociation of the hemispheres. Study of Figure 2 will show that the most marked formant transition contrast is between alveolar pairs followed by different vowels (/di-tu, du-ti/). If we assume that auditory analysis of both inputs is attempted by both hemispheres, we might expect that these pairs, with their conflicting transitions, would present the greatest analytic problem and that this problem would be more difficult for the right-ear/left-hemisphere system than for the left-ear/right hemisphere system.<sup>9</sup> The results bear this out: it is precisely these pairs that lower right-ear performance and contribute most strongly to the observed interaction.

We should be clear that we are here accounting not for a reversal of the ear advantage but for a reduction in its size due to lowered right-ear performance under one condition of this experiment. We should not confuse this reduction with the generally lower left-ear performance observed in dichotic speech studies. The latter may be attributed to loss of auditory information arising from interhemispheric transfer of the left-ear signal to the dominant hemisphere for phonetic processing (see Studdert-Kennedy & Shankweiler, 1970), while the reduced right-ear performance

<sup>8</sup> We are here conceiving a set of individual feature processors organized as a system.

<sup>9</sup> Superiority of the right hemisphere in auditory pattern recognition has so far been shown only for nonspeech patterns. The possibility of left-hemisphere superiority in the analysis of patterns peculiar to speech (such as formant transitions), due to its possession of specialized auditory feature processors, cannot be excluded. This possibility is currently being investigated experimentally at Haskins Laboratories. In the present account, we are tentatively assuming, on the basis of the cited dichotic work with nonspeech patterns, that the left hemisphere is inferior in the resolution of conflicting ipsilateral and contralateral auditory patterns.

under one condition of this experiment is here attributed to increased interference of the left-ear signal with the right-ear signal during auditory analysis in the left hemisphere.

Clearly this account is not complete, since it leaves unexplained the rise in left-ear performance on labial consonants when vowels are the same. However, detailed explanation is of less importance than the fact of the interaction. The finding that the vowel condition affects stop consonant perception differently for the two ears (Table 3) is the first evidence of central auditory interaction between dichotic speech inputs. From this we may infer that inhibition of the ipsilateral signal under dichotic stimulation (see Milner, Taylor, & Sperry, 1968) occurs not in the pathways to the cerebral hemispheres but after central auditory analysis, either at the auditory-phonetic interfacing or during phonetic analysis. Since we have already seen that the two inputs may interact during preparation of a response from the phonetic component, we may eliminate the first alternative and conclude that inhibition occurs at some stage of phonetic analysis. In this regard, we note that the laterality effect for speech is only obtained if both signals are perceived as speech: contralateral white noise (Shankweiler & Halwes, unpublished data), noise limited to the speech band (Darwin, 1971a), or pure tones (Day & Cutting, 1970) do not produce an ear advantage. This, too, would seem to implicate phonetic rather than auditory analysis as the primary level of dichotic competition.<sup>10</sup>

To sum up, this study has provided further grounds for distinguishing between auditory and phonetic levels of speech processing. The results suggest that both signals of a dichotically presented syllable pair are transmitted to a single phonetic processing system and that correct output from the system is facilitated if the two messages have phonetic features in common. At the same time, they suggest that inhibition of the ipsilateral signal in the perception of dichotically presented speech occurs during phonetic analysis.

#### REFERENCES

- CHISTOVICH, L. A., GOLINSINA, A., LUBLINSKJA, V., MALIVNIKOVA, T., & ZUKOVA, M. Psychological methods in speech perception research. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 1968, 21, 102-106.
- DARWIN, C. J. Auditory perception and cerebral dominance. Unpublished Ph.D. thesis, University of Cambridge, 1969.

<sup>10</sup> Recent findings of Darwin (1971b) and Spellacy and Blumstein (1970) support this interpretation. These studies showed that the same acoustic conditions of competition may or may not result in an ear advantage depending upon the complexity of the subjects' task, which was varied by manipulating the range of other stimuli occurring in the experiment.

- DARWIN, C. J. Dichotic forward and backward masking of speech and nonspeech sounds. Paper presented at 81st Meeting of Acoustical Society of America, Washington, D. C., April 1971a.
- DARWIN, C. J. Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, 1971b, 23, 46-62.
- DAY, R. S., & CUTTING, J. E. Perceptual competition between speech and nonspeech. *Journal of the Acoustical Society of America*, 1970, 49, 85 (A). (Text in Haskins Laboratory Status Report, SR/24.)
- FROMKIN, V. A. Tips of the slung-or-to err is human. Univ. of California at Los Angeles: Working Papers in Phonetics, 1970, 14, 40-79.
- FRY, D. B. Perception and recognition in speech. In M. Halle, H. G. Lund, H. McClean, and C. H. van Schooneveld (Eds.), *For Roman Jakobson*. The Hague: Mouton, 1956.
- KIMURA, D. Left-right differences in the perception of melodies. *Quarterly Journal of Experimental Psychology* 1964, 16, 355-358.
- KIMURA, D. Functional asymmetry of the brain in dichotic listening. *Cortex*, 1967, 3, 153-178.
- MILNER, B. Laterality effects in audition. In V. B. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance*. Baltimore: Johns Hopkins Univ. Press, 1962.
- MILNER, B., TAYLOR, L., SPERRY, R. W. Lateralized suppression of dichotically-presented digits after commissural section in man. *Science*, 1968, 161, 184-185.
- SHANKWEILER, D. Effects of temporal-lobe damage on perception of dichotically presented melodies. *Journal of Comparative and Physiological Psychology*, 1966, 62, 115-119.
- SHANKWEILER, D., & STUDDERT-KENNEDY, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, 19, 59-63.
- SPELLACY, F., & BLUMSTEIN, S. The influence of language set on ear preference in phoneme recognition. *Cortex*, 1970, 6, 430-439.
- STUDDERT-KENNEDY, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics*, Vol. XII. The Hague: Mouton, in press. (Also in Haskins Laboratories Status Report, SR/23, 1971, pp. 15-48).
- STUDDERT-KENNEDY, M., & SHANKWEILER, D. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, 48, 579-594.

(Accepted December 12, 1971)