Journal of Experimental Psychology:
Human Perception and Performance
1975, Vol. 104, No. 2, 105-120

Aspects of Phonological Fusion

James E. Cutting
Yale University and Haskins Laboratories, New Haven, Connecticut

Phonological fusion occurs when the phonemes of two different speech stimuli are combined into a new percept that is longer and linguistically more complex than either of the two inputs. For example, when PAY is presented to one ear and LAY to the other, the subject often perceives PLAY. The present article is an investigation of the conditions necessary and sufficient for fusion to occur. The rules governing phonological fusion appear to be the same for synthetic and natural speech, but synthetic stimuli fuse more readily. Fusion occurs considerably more often in dichotic stimulus presentation than in binaural presentation. The phenomenon is remarkably tolerant of differences in relative onset time between the to-be-fused stimuli and of relative differences in fundamental frequency, intensity, and vocal tract configuration. Although phonological fusion is insensitive to such nonlinguistic stimulus parameters, it is sensitive to linguistic variations at the semantic, phonemic, and acoustic levels.

Most of the dichotic listening literature has dealt with the phenomenon of perceptual *rivalry*. Given a different stimulus presented to each ear at the same time, the subject typically reports hearing one or both of them. The different information contained in each stimulus is not combined into a single percept. Thus, for example, given the dichotic digits FOUR/SIX, the subject never reports hearing SORE or FIX. Perceptual *fusion* does occur when certain variables are taken in account. In several types of fusion phenomena, the stimulus variables that facilitate fusion appear to be psycholinguistic in nature. For example, given the dichotic pair PAHDUCT/RAHDUCT, the subject often reports hearing PRODUCT (Day, 1968). In this type of fusion, segments of both stimuli are combined to form a new percept that is longer and linguistically more complex than either of the two inputs.

This phenomenon is called *phonological* fusion because it conforms to phonological rules of English: Given BANKET/LANKET, the subject reports hearing BLANKET, not LBANKET (Day, Note 1). Nonword fusions also occur; for example, GAB/LAB yields GLAB, not LGAB (Day, 1968). According to phonological rules of cluster information in English, initial stop consonant + liquid clusters are allowed, but initial liquid + stop clusters are not. Day (1970) found that when stimuli are presented dichotically with these constraints removed, fusion occurs in both directions. Given the stimuli TASS/TACK, for example, the subject reports hearing TASK on some trials and TACKS on others. Both /sk/ and /ks/ clusters are permissible in final position in English.

Phonological fusion cannot be explained as a response bias for acceptable English words. Day (1968) found that when different productions of PAHDUCT are presented to each ear, subjects report hearing PAHDUCT. That is, they do not report hearing the acceptable English word that corr sponds most closely to the nonword inr Likewise, RAHDUCT/RAHDUCT

This research was supported by National Institute of Child Health and Human Development Grant HD-01994 to the Haskins Laboratories, and was based in part on a PhD thesis submitted to the Department of Psychology of Yale University. Thanks go to Ruth S. Day, with whom I have never had a colorless conversation; Franklin S. Cooper, with whom I have never had a fruitless conversation; and the members of the New Haven Dance Ensemble, with whom I never talked about this at all.

The author is now at Wesleyan University, Middletown, Connecticut and Haskins Laboratories. Requests for reprints should be sent to James E. Cutting, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

RAHDUCT. Only when the stimuli are PAHDUCT/RAHDUCT does fusion occur, and it occurs regardless of which stimulus is presented to which ear. In the present article, I will consider a fusion response to occur when a stimulus pair, each item of which has n phonemes, yields a percept of $n + 1$ phonemes (following Day, 1968).

The studies presented here are an analysis of phonological fusion and the conditions necessary and sufficient for fusion to occur. Of primary interest are the modes of presentation that optimize the occurrence of fusion responses and the stimulus variables under those conditions that enhance and detract from the fused percept. Both linguistic and physical (nonlinguistic) variables are of interest.

To undertake such a project, one must be able to control the stimuli in a precise manner, and such control is only available with a computer-driven speech synthesizer. However, synthetic speech has a characteristic "cold-in-the-nose" quality which may effect phonological fusion. To insure that the present studies are not concerned with artifacts inherent to synthetic speech, fusion was first observed for synthetic and natural speech tokens of the same items.

Synthetic Versus Natural Speech

Experiment 1: Baseline Data

General Method

Tape preparation. Four sets of stimuli of the same general pattern were selected: the PAY set (PAY, LAY, RAY); the BED set (BED, LED, RED); the CAM set (CAM, LAMB, RAM); and the GO set (GO, LOW, ROW). Two versions of each stimulus were prepared, one in synthetic speech and one in natural speech. Synthetic stimuli were prepared on the Haskins Laboratories' parallel-resonance synthesizer. All stimuli were identical in pitch contour, rising from 116 to 124 Hz and then falling to 82 Hz. Within each set all stimuli were identical in duration (350, 450, 400, and 350 msec for the four sets, respectively), and all had the same acoustic structure except for the initial 150 msec. Liquid stimuli within each set (those beginning with /l/ and /r/) differed only in the direction and extent of the third-formant transition. This acoustic cue distinguishes the two liquids (O'Connor, Gerstman, Liberman, Delattre, & Cooper, 1957). Each liquid stimulus began with 50-msec steady state prerelease formant resonances, followed by 100-msec transitions in each formant gliding to the resting frequency of the following vowel. Stop consonants began with 50 msec of formant transitions gliding into the following vowel. Voice onset time was 0 msec for voiced consonants and +50 msec for voiceless consonants. Natural speech versions of the same items were spoken by the author and recorded on audio tape. An effort was made to match utterances in terms of pitch, intensity, and duration, but some variation was unavoidable. Both synthetic and natural speech stimuli were digitized and stored on disk file for preparation of diotic and dichotic tapes (Cooper & Mattingly, 1969).

Subjects. Sixteen Yale University undergraduates participated in identification and fusion tasks. Each received course credit for his or her efforts. As in all experiments in this paper, subjects were right-handed native American English speakers with no history of hearing trouble and with no experience listening to synthetic speech. They listened in groups of four to tapes played on an Ampex AG-500 dual-track tape recorder. Auditory signals were sent through a listening station to matched Telephonics earphones (Model TDH39). Gains on the tape recorder and listening station were adjusted so that stimuli were presented at approximately 80 db. (re 20 $\mu N/m^2$). These criteria and procedures were used for all experiments in the present article. No subject served in more than one experiment. The fusion task in the present study preceded the identification task, but it is more instructive to consider them in reverse order.

Task 1: Identification of Individual Stimuli

All stimuli, natural and synthetic, were recorded on a diotic identification tape. The tape consisted of a random sequence of 120 trials: (4 sets of stimuli) × (2 versions of each set: natural and synthetic) × (3 stimuli per set) × (5 observations per stimulus). Subjects wrote down the entire word that they heard presented. There was a 3-sec interval between items.

Results

All stimuli were highly identifiable. Each synthetic and natural speech item was correctly identified on better than 93% of all trials. Cluster responses (such as PLAY) were reported on less than 2% of all trials. Identification tasks were a part of each experiment presented in this article. Since the stimuli and results generally did not differ from those in Experiment 1, the identifications are not discussed in all experiments.

Task 2: Identification of Fusible Pairs

All stimuli used in the identification task were also used in the fusion task. However, instead of presenting one stimulus at a time, two stimuli were presented, one to each ear. The dichotic pair consisted of a stop stimulus and a liquid stimulus from the same set; for example, PAY/LAY or PAY/RAY. All stimuli and possible fusions were monosyllabic, high-frequency English words: PAY/LAY → (yields) PLAY, PAY/RAY → PRAY, BED/LED → BLED, BED/RED → BREAD, CAM/LAMB → CLAM, CAM/RAM → CRAM, GO/LOW → GLOW, and GO/ROW → GROW. Most of these items and fusions have Thorndike-Lorge (1944) word frequencies of at least 100 per million. No alveolar stop-consonant stimuli were chosen because /tl/ and /dl/ clusters do not appear in initial position in English.

Two dichotic tapes were prepared, one using synthetic stimuli and one using natural speech stimuli. Three lead times were selected: The plosive portion of the stop initial stimulus began 50 msec before the onset of the liquid initial stimulus, the stop and liquid began simultaneously, or the liquid began 50 msec before the stop. Day (1968) and Day and Cutting (Note 2) found that within such a restricted range, lead time has little effect on fusion. The pulse code modulation system at the Haskins Laboratories allows temporal alignments to within .5-msec accuracy (Cooper & Mattingly, 1969). There were 96 randomly ordered items per tape: (4 sets of fusible stimuli) × (2 liquids per set) × (3 lead times) × (2 channel arrangements per pair) × (2 observations per dichotic item). All subjects listened to both types of stimuli: Half listened first to the synthetic speech pairs and then to the natural speech pairs, while the other half listened in the reverse order. Subjects were instructed to write down what they heard on each trial (one word or two, real words or nonsense) and to respond after *each* trial. Before the task began, they listened to several practice pairs and wrote their responses in order to familiarize themselves with the task and the stimuli.

Results and Discussion

Fusion responses occurred readily, but were more frequent for synthetic stimuli than for natural speech stimuli. Subjects fused synthetic pairs on 61% of all trials, while fusing on only 31% of all natural speech trials. This 2:1 ratio was highly significant by a sign test: All 16 subjects showed results in this direction ($z = 3.75$, $p < .0001$). The order in which subjects listened to synthetic and natural speech pairs was not a significant factor.

In general, fusion was more frequent for stop + /l/ pairs than for stop + /r/ pairs (e.g., PAY/LAY versus PAY/RAY), as shown in Table 1. This pattern was reported previously by Day (1968). Anomalous fusion responses frequently occurred for both synthetic and natural speech pairs. The most common was a stop + /l/ fusion response for a stop + /r/ stimulus pair: for example, PAY/RAY → PLAY. Whereas /l/ was substituted for /r/ quite frequently, the reverse substitution was much less frequent. Day (1968) and Cutting and Day (1975) have reported this /l/-for-/r/ substitution phenomenon. It cannot be accounted for by the frequency of these clusters in English. In this study, /l/-for-/r/ substitutions occurred for both synthetic and natural speech stimuli at approximately the same rate. The relative high frequency of these responses in the present study seems unusual, since Day (1968) found them to occur on only 4% of all fusions of stop + /r/ pairs. The phenomenon will be considered in more detail in Experiments 8 and 10.

TABLE 1

Percent Fusion Responses for Each Pair in Experiment 1, in Both Synthetic Speech and Natural Speech Versions

Fusible pair	Percent fusion		
	Synthetic	Natural	*M*
PAY/LAY	78	41	59
PAY/RAY	70	33	51
			(55)
BED/LED	48	33	40
BED/RED	46	19	33
			(37)
CAM/RAM	63	29	46
CAM/LAMB	57	17	37
			(42)
GO/LOW	68	41	54
GO/ROW	60	35	47
			(51)
Stop + /l/	64	36	50
Stop + /r/	58	26	42
M	(61)	(31)	

Note. Values in parentheses are the means of the means.

Frequency of fusion responses differed across sets of stimuli. Table 1 shows the fusion rates for the four sets in both synthetic and natural speech versions. Fusion rates for the PAY and GO sets were consistently higher than those for the BED and CAM sets. These differences may be related to the frequency of occurrence of the fused responses as words in English; according to Thorndike and Lorge (194 and Carroll, Davies, and Richman (1971), the words PLAY, PRAY, GLOW, and GROW occur more frequently in general publications than BLED, BREAD, CLAM, and CRAM. Day (1968) has shown that fusions occur more frequently when the fused outcome is an acceptable English word than when it is not (although nonword fusions do occur, e.g., GORIGIN/LORIGIN →GLORIGIN). Perhaps within the domain of words, frequency of occurrence in the natural language also plays a role in fusion. There were no ear effects nor channel effects in the present experiment or in any experiment described in this article. There were no significant effects of lead time configuration here, but see Experiment 3.

Overview of Experiment 1

Although there were differences in fusion between synthetic and natural speech stimuli, the rules that govern their fusibility appear to be comparable. This conclusion is based on two distinct results. First, the pattern of fusions for the four sets of stimuli was quite similar for both natural and synthetic speech stimuli. In both cases the PAY and GO sets fused more readily than the CAM and BED sets. Second, the pattern of "misperceptions" for the two types of stimuli was almost identical. The phoneme /l/ was quite frequently substituted for /r/ on a fusion response; for example, PAY/RAY → PLAY. The reverse substitution was much less common. Thus, it appears that generalizations can be made about aspects of phonological fusion from the results of tasks using synthetic stimuli.

Aspects of Stimulus Presentation

Experiment 2: Dichotic Versus Binaural Presentation

Method

The synthetic stimulus tapes used in Experiment 1 also were employed here. Twenty-four different Yale University undergraduates listened to dichotic pairs, in which each item was presented to a separate ear, and to binaural pairs, in which items were electrically mixed and presented to both ears. Eight subjects listened first to a sequence of dichotic trials, then to a sequence of binaural trials (Group 1), while eight others listened in reverse order (Group 2). For these subjects binaural pairs were presented by mixing the two channels of the dichotic tape. The remaining eight subjects listened to a new tape of dichotic and binaural trials randomly intermixed (Group 3). This tape was exactly twice as long (192 items) as the tapes of Experiment 1, and the dichotic and binaural pairs were constructed digitally by computer at the time of recording.

Results and Discussion

Fusion was more frequent for dichotic presentations than for binaural presentations: 45% and 15%, respectively. This 3:1 ratio was highly significant ($z = 4.8$, $p < .0001$): All 24 subjects yielded fusion results in this direction. Other than this main effect, there were no statistically reliable differences in fusion. For example, the dichotic fusion rates for Groups 1, 2, and 3 were 53%, 38%, and 45%, respectively; corresponding binaural fusion rates were 16%, 8%, and 23%. Large variances in fusion rate within groups kept these differences from revealing any other statistically reliable pattern. As in Experiment 1, /l/-for-/r/ phoneme substitutions occurred readily, and there was only a slight effect of lead time. Such differences are considered in more detail in Experiment 3. Differences in fusion rate among stimulus sets were similar to those in Experiment 1.

More frequent fusion in dichotic than in binaural presentation demonstrates that phonological fusion is an active and central process. "Pre-fused," or electrically mixed, synthetic stimuli are not perceived as a phonological mixture of the two items. The reason fusion is so infrequent in the binaural case presumably stems from the fact that the electrically mixed stimuli beat against one another, creating a very "dirty" signal. No such contamination occurs for dichotic pairs. Unique left-ear and right-ear stimuli appear to be perceptually combined at some central locus.

Experiment 3: Variation in Lead Time

Day (1970; Note 1) and Day and Cutting (Note 2) demonstrated that variation in the alignment of natural speech fusible stimuli has little effect on fusion rate. Fusion occurs almost as readily when the liquid stimulus (e.g., LANKET) leads the stop stimulus (BANKET) by as much as 150 msec, as when the stop leads the liquid by the same extent. Moreover, fusion rate in both instances is nearly identical to that for the simultaneous onset case. This result is particularly surprising when one considers that neither the segment /b/ in BANKET or /l/ in LANKET was longer than 90 msec in duration. In Experiments 1 and 2, lead times of 0 msec, + 50 msec (when the stop leads), and −50 msec (when the liquid leads) yielded nearly identical fusion rates. The present experiment was designed, in part, to investigate relative onsets (lead times) of longer than 150 msec to discover the interval at which fusion disintegrates.

In addition, Experiment 3 explores the relationship between fusion and backward masking. Consider the dilemma in the comparison of two hypothetical dichotic pairs. Imagine first the nonfusible stimulus pair PAY/DAY. If one staggers their onsets by 50 msec, the second item will most likely mask the first (see Pisoni, 1973, and Studdert-Kennedy, Shankweiler, & Schulman, 1970, among many others). Thus, if DAY leads PAY, the subject may report hearing only PAY. In a fusible pair, PAY/LAY, the later-arriving item will not mask the first. Even if LAY begins 50 msec before PAY, the subject seldom reports hearing PAY alone, but usually fuses them into PLAY. This response in this particular situation is the crux of a dilemma. The segments /p/ and /l/ are not only combined into a cluster but also perceptually reordered in time: /l/ began before /p/ in the stimuli, but in the response they are reversed. The rules of language dictate the perception of PLAY, whereas the rules of backward masking and central processing in general predict the perception of PAY. Perhaps varying relative onset times beyond 50 and even 150 msec in the phonological fusion situation would strain fusion and favor backward masking.

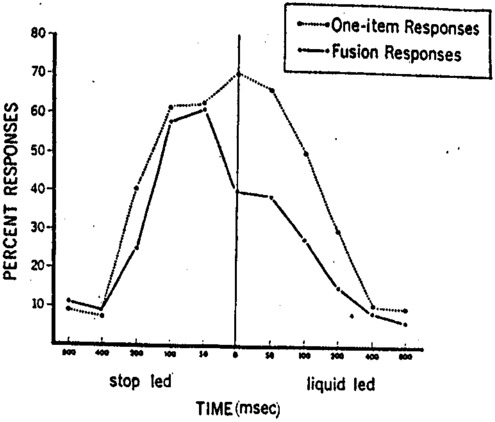

FIGURE 1. Comparison of fusion rate and number of one-item responses at 11 lead times in Experiment 3.

Method

The PAY set (PAY, LAY, RAY) was again employed. In addition, the KICK set (KICK, LICK, RICK) was generated on the Haskins' synthesizer. The PAY set stimuli were 350 msec in duration, and the KICK set stimuli 325 msec. Within a set the items were identical in all respects except for the initial formant transitions requisite for the perception of a stop consonant versus /r/ versus /l/. Dichotic pairs were assembled for all combinations of fusible items within a set: PAY/LAY → PLAY, PAY/RAY → PRAY, KICK/LICK → CLICK, and KICK/RICK → CRICK. Eleven lead times were chosen: 0, ±50, ±100, ±200, ±400, and ±800 msec. Since all stimuli were 350 msec or less in duration, the 400- and 800-msec relative onset items involved temporally nonoverlapping stimuli. All pairs were represented in two tapes, each with a different random order. Each tape consisted of 88 pairs: (2 sets of stimuli) × (2 stop/liquid pairs per set) × (11 lead times) × (2 channel arrangements per pair). Eight Yale University undergraduates served as subjects for course credit. Each subject listened to both tapes, yielding 176 trials per subject.

Results and Discussion

As shown in Figure 1, fusion occurred most readily at short lead times, especially when the stop stimulus led the liquid by 50 and 100 msec. Here, fusion responses occurred on 63% and 59% of all trials, respectively. Fusions were infrequent at the −200, ±400, and ±800 msec leads, where they occurred on an average of less than 10% of the time. Intermediate fusion rates occurred at the 0, −50, −100, and +200 msec leads. Both sets of stimuli fused readily, with fusions slightly more frequent for the PAY set.

The fusion rate at the short leads (0 and ±50 msec) in the present experiment is similar to that found in Experiments 1 and 2. One would expect fusion to occur at the longer lead times where members of a pair do not overlap in time, but one might expect that they would depress fusion response rates at short leads. The absence of a decrease suggests that fusions are not a response bias resulting from lack of knowledge about the stimuli. The fusions that did occur at long leads occurred primarily at the beginning of the task when subjects were less familiar with the items. Fusions at shorter leads continued at the same rate throughout the task. Again, /l/-for-/r/ substitutions were frequent.

Fusion and Backward Masking

To consider the conflict between phonological rules and the rules of backward masking, it is necessary to observe more than just fusion responses. For example, not only are there fusion and nonfusion responses, but there are one-item and two-item responses as well. These four categories are not mutually exclusive. Consider the stimulus pair PAY/LAY: There are one-item fusions, PLAY; one-item nonfusions, PAY or LAY; two-item fusions (responses in which at least one fusion occurs), PLAY + PAY, PLAY + LAY, or even PLAY + PLAY; and of course two-item nonfusions, PAY + LAY. All of these responses occurred in the present study, but at different frequencies according to lead times. As might be expected, one-item responses predominated at the short leads while two-item responses occurred at the long leads. In Figure 1, the one-item response curve is superimposed on the fusion curve. Notice that it is nearly symmetrical, unlike the fusion response curve.

Except for the 200-, 400-, and 800-msec relative onsets, the fusion curve is primarily a subset of the single response curve. That is, between ±100 msec the subject typically wrote down a single response, and often that response was a fusion (e.g., PLAY). When the subject wrote down a single response that was not a fusion, 91% of those responses were stop stimuli alone (e.g., PAY). Thus, the area between the two curves in Figure 1 is primarily a representation of the number of PAY or KICK responses.

Between +200- and 0-msec lead times, phonological rules dominated. Most one-item responses within this lead-time domain were fusions. However, between 0- and −200-msec leads, two types of single-item responses occurred at near equal frequencies. One-item fusions, PLAY, demonstrate the influence of language rules, whereas one-item nonfusions, PAY, demonstrate backward masking and the influence of central processing constraints. Within this lead-time domain, 58% of all one-item responses are fusions and 42% are nonfusions. In Experiments 1 and 2 phonological rules always dominated over the rules of backward masking. In the present experiment, in which lead times were extended to more than double the stimulus duration, backward masking did occur, but occurred only for stimulus pairs in which the liquid stimulus led the stop by less than 200 msec; even in pairs like LAY-before-PAY, backward masking occurred less than half the time. In PAY-before-LAY pairs, there was essentially no backward masking, there were few LAY responses, and PLAY dominated at all the short leads.

Although differences in lead time of greater than about 200 msec reduced fusion to a minimum, fusion occurred within the ±100-msec domain with considerable frequency. Compare this domain with the ±2-msec tolerance of relative onset differences for sound localization, another form of auditory fusion, and one sees that the tolerance for lead-time differences in phonological fusion is several orders of magnitude greater.

Overview of Experiments 2 and 3

Mode of stimulus presentation has considerable effect on fusion; dichotic fusions of synthetic speech stimuli are much more frequent than binaural fusions, suggesting that central processes are much more important than peripheral processes in phonological fusion. The fact that fusion is relatively insensitive to lead-time variation lends credibility to this view. In visual masking, for example, central processes appear to operate over a longer time span than peripheral processes (Turvey, 1973). If phonological fusion is a central phenomenon, perhaps certain physical parameters of the stimuli would have little effect on the frequency at which fusion responses occur, just as they have little effect on central processes in vision (Turvey, 1973).

Physical Aspects of Stimuli

Experiments 4–6: Pitch, Intensity, and Vocal Tract Size

Method

The same stimulus sets were used as in Experiment 3: the PAY set (PAY, LAY, RAY) and the KICK set (KICK, LICK, RICK). For Experiment 4, two versions of each stimulus were synthesized, one at a relatively high fundamental frequency (pitch) and one at a relatively low pitch. All stimuli had a falling pitch contour. High-pitch stimuli began at 140 Hz and fell to 120 Hz, whereas low-pitch stimuli began at 120 Hz and fell to 100 Hz. For Experiment 5, all the low-pitch stimuli of Experiment 4 were employed, but they were synthesized at two relative intensities, 80 db. and 65 db. re 20 $\mu N/m^2$. For Experiment 6, all low-pitch high-intensity stimuli were employed, but they were synthesized as if spoken by two different speakers: one a normal adult male and another a speaker whose apparent vocal tract size was only 5/6 as large as the adult male. Because these stimuli had a low pitch, the tokens from the smaller vocal tract configuration sounded as if they were produced by a male midget. Variation in apparent vocal tract size was accomplished by elevating the formant frequencies of the PAY and KICK set stimuli by a factor of 6/5. All stimuli were highly identifiable, as determined by identification tests.

Dichotic pairs were assembled from all possible combinations of fusible items within each set. Consider the pair PAY/LAY in Experiment 4. The members of the pair shared the same pitch, PAY-high (for high pitch)/LAY-high and PAY-low/LAY-low, or they differed in pitch, PAY-high/LAY-low and PAY-low/LAY-high. The same pattern was followed in Experiment 5, except that relative intensity was substituted for relative pitch. In Experiment 6 the pattern was similar. Paired items either shared vocal tract size (PAY-large/LAY-large and PAY-small/LAY-small) or they differed in vocal tract size (PAY-large/LAY-small and PAY-small/LAY-large). Two dichotic tapes with different random orders were prepared for all three experiments. Each tape contained 96 items: (2 sets of stimuli) × (2 stop/liquid pairs per set) × (4 pitch, intensity, or vocal tract combinations) × (3 lead times) × (2 channel arrangements per pair). The leads used in all three experiments were 0 and ±50 msec. Twelve different Yale University undergraduates served as subjects in each of the three experiments.

Results

Fusion occurred readily for all stimuli pairs. In Experiment 4 the frequency of fusion response rates was identical for pairs that shared the same pitch and those that differed in pitch: 50% each. Fusions for all four types of pairs were within a few percentage points of one another. The pattern was the same in Experiments 5 and 6: Pairs with the same intensity and pairs with different intensities all fused at a rate of 36%, and pairs with the same vocal tract configuration and those with different vocal tract configurations fused at a rate of 59% each. In other words, within each experiment there was no contribution nor decrement in fusion rate attributable to variation of pitch, intensity, or vocal tract size.

Lead time had a nominal effect on fusion, but there were no interactions with any other factor. The fusion rates for the +50, 0, and −50 msec lead times were, respectively, 55%, 48%, and 48% in Experiment 4; 40%, 35%, and 34% in Experiment 5; and 65%, 56%, and 57% in Experiment 6. The /l/-for-/r/ substitutions were frequent in all three studies, and the PAY set fused at a slightly higher rate than the KICK set.

Experiment 7: Multidimensional Stimulus Variation

Method

The PAY set and the KICK set were synthesized in eight different versions: (2 pitches, those frequencies used in Experiment 4) × (2 intensities, those used in Experiment 5) × (2 vocal tract configurations, those used in Experiment 6). One dichotic tape consisted of items within a set that shared *none* of the values in the dimensions of pitch, intensity, and vocal tract size. Thus, for example, PAY-low(pitch)-low(intensity)-small (vocal tract) was paired with LAY-high-high-large. This tape consisted of 96 items: (2 sets of stimuli) × (2 liquids per set) × (3 lead times) × (8 different mismatched pairings, across the dimensions of pitch, intensity, and vocal tract size). Channel arrangements of the stimuli were arranged such that stops and liquids were equally represented on both channels. A second 96-item dichotic tape was prepared to consist of fusible items that were all low-pitch, high-intensity, large vocal tract stimuli: (2 sets of stimuli) × (2 liquids per set) × (3 lead times) × (2 channel arrangements) × (4 observations per pair). Twenty-four Yale University undergraduates listened to both fusion tapes. Half the subjects listened first to the multidimensionally varied tape and then to the unvaried tape (Group 1), whereas the other half listened in reverse order (Group 2).

Results

Differences in pitch, intensity, and vocal tract size among the fusible stimuli had remarkably little effect on fusion. Overall fusion rate was 47% for the multidimensionally varied pairs and 44% for the unvaried pairs, a result that shows the reverse trend of what might have been expected. The two groups of subjects differed somewhat in fusion rate for each class of stimulus pairs. Group 1 fused on 46% of all varied trials and on only 38% of the subsequent unvaried trials. This trend was not significant; only 7 out of 12 subjects showed this pattern. Group 2 fused on 49% of the unvaried trials and on 45% of the subsequent varied trials. This trend was not significant either; again, 7 out of 12 subjects yielded results in this direction. The accustomed nominal effect of lead times and the "usual" amount of /l/-for-/r/ substitutions were found in the data. Again, PAY-set pairs fused somewhat more often than KICK-set pairs.

Overview of Experiments 4–7

Within the relatively large ranges of values explored in the dimensions of pitch (20 Hz), intensity (15 db.), and vocal tract size (a ratio of 5:6), the physical aspects of the signal have remarkably little effect on fusion. Even when the three dimensions vary, such stimulus pairs as PAY-low-low-small/LAY-high-high-large fuse as readily as PAY-low-low-small/LAY-low-low-small. Taken together with the results of Experiments 2 and 3, the results of these experiments strongly suggest that the perceptual combination of the stimuli occurs after the linguistic information is extracted from each of the signals. Otherwise, physical variation would surely affect the rate at which fusion occurs.

Linguistic Aspects of Stimuli

Three linguistic levels are considered in the remaining experiments: the level of semantics, the level of the phoneme, and the level of acoustic structure as it pertains to language. The phoneme level might appear to be the primary linguistic level at which phonological fusion occurs, since phonemes are fused into clusters. However, this need not be the case. Therefore two other levels were chosen, one higher (semantics) and one lower (acoustics) than the phoneme level. The experiments begin at the semantic level and move "downward" to the phoneme and acoustic levels.

Experiment 8: Semantic Level: Sentence Contexts

Day (1968) found that semantic cues at the word level influenced fusion rate. Fusion rates were higher when the fused outcomes were real words than when they were nonwords (PAHDUCT/RAHDUCT → PRODUCT vs. PAHLOW/RAHLOW → PRAHLOW). Nonword fusions did occur (GORIGIN/LORIGIN → GLORIGIN), although at a reduced rate.

The present experiment was designed to observe the effects of semantic cues at the sentence level on fusion rate. Since Experiments 1–7 found that /l/ was frequently substituted for /r/ in fusion responses, the present study was also designed to observe the effect of semantic context on /l/-for-/r/ substitutions.

Method

Two sets of stimuli were used, the PAY set (PAY, LAY, RAY) and the GO set (GO, LOW, ROW). Fusible pairs were presented in isolation, as in previous experiments, and imbedded in sentence contexts. The PAY set appeared in the context THE TRUMPETER ———S FOR US (Sentence Frame 1) and THE MINISTER ———S FOR US (Sentence Frame 2), whereas the GO set appeared in THE COALS ARE ——ING AGAIN (Sentence Frame 3) and THE TREES ARE ——ING AGAIN (Sentence Frame 4). These sentences were made into dichotic pairs such that THE TRUMPETER PAYS FOR US, for example, was presented to one ear and THE TRUMPETER LAYS FOR US to the other. All fusible pairs appeared in both semantically probable and improbable contexts. Thus, when PAY/LAY appears in Sentence Frame 1, the response might be THE TRUMPETER PLAYS FOR US, a reasonable sentence; but when the same pair appears in Sentence Frame 2, the response might be THE MINISTER PLAYS FOR US, a semantically less probable sentence. A reverse pattern of expected results follows from PAY/RAY pairs, and Sentence Frames 3 and 4 with GO-set stimuli parallel Sentence Frames 1 and 2.

Two tapes were prepared, one with fusible targets imbedded in sentences and the other with target pairs presented in isolation. The sentence tape consisted of 64 items: (4 sentence frames) × (2 stop/liquid pairs per set) × (2 channel arrangements) × (4 observations per sentence). Dichotic sentence pairs were presented at a simultaneous onset with 12 sec between trials. Subjects wrote down the entire sentence they heard. The isolated-pair tape also had 64 pairs: (2 sets of stimuli) × (2 stop/liquid pairs per set) × (2 channel arrangements) × (8 observations per pair). Again, only the simultaneous onset time was used, but the intertrial interval was 4 sec. Subjects wrote down "what they heard"—one word or two, acceptable words or nonsense. Half of the 16 subjects listened first to the sentence tape and then to the isolated-pair tape, while the others listened in reverse order.

Results

Fusion was significantly greater in the sentence condition than in the isolated-pair condition. All subjects showed this trend ($z = 3.8$, $p < .001$). Fusion rate was 89% for sentence trials and 63% for isolated-pair trials. The order in which subjects listened to the tapes was not a significant factor. Fusion rates were comparable for stop + /r/ and stop + /l/ items, as well as for both sets of stimuli in each condition.

Stop + /l/ responses dominated all sentence contexts even when they were semantically inappropriate. For Sentence Frame 1, THE MINISTER PLAYS FOR US occurred on 74% of all trials. Certainly, the minister PLAYING is semantically less likely than PRAYING (which occurred on only 16% of all Sentence Frame 1 trials, regardless of the liquid stimulus presented) even in today's climate of changing roles. Likewise, in Sentence Frame 3, THE TREES ARE GLOWING AGAIN occurred on 83% of all trials, despite the fact that it is not very probable. GROWING responses occurred on only 4% of such trials, regardless of liquid presented. Sentence Frames 2 and 4 yielded a more predictable set of results: In both cases, stop + /l/ fusions are semantically probable and as fusions were very frequent. They occurred on 96% and 84% of all trials, respectively. There were virtually no THE TRUMPETER PRAYS FOR US and THE COALS ARE GROWING AGAIN responses, regardless of what liquid stimulus was presented. Other responses to sentence frames were primarily responses in which only the stop stimulus was reported; for example, THE MINISTER PAYS FOR US. Pairs of stimuli in the isolated-pair condition yielded similar liquid substitution results: For example, when PAY/RAY fused, PLAY responses were given 85% of the time, a pattern nearly identical to that for sentence trials. The reverse substitution rarely occurred.

The present experiment showed that meaning at the sentence level can not nullify the /l/-substitution effect. Relative frequency of occurrence of the fused words cannot account for the substitutions either: GLOWING, for example, is much less frequent than GROWING (Carroll, et al., 1971; Thorndike & Lorge, 1944). Word-level considerations, however, may provide a clue. Day (1968) found that, although subjects usually reported hearing GROCERY when given GOCERY/ROCERY, sometimes they reported hearing GLOCERY, a nonword. In the present series of experiments, both the stop + /r/ and stop + /l/ fusions for a given set were acceptable English words. Day (1968), on the other hand, chose stimuli that could fuse meaningfully with only one of the liquids, /l/ or /r/. She found that PAHDUCT/RAHDUCT → PRODUCT and not PLODUCT, and that GEEDY/REEDY → GREEDY not GLEEDY. Such results suggest that meaning at the word level can override the /l/-substitution effect. This effect is considered again in Experiment 10.

Experiment 9: Phoneme Level: Stops and Fricatives

Phonological fusion occurs when phonemes from different dichotic stimuli are perceived as a cluster. Experiments 9 and 10 examined the phonemic components, the stop and the liquid, to assess their importance in the fusion phenomenon.

In Experiments 1–8, stop consonants served as the first phoneme of the to-be-fused consonant–consonant–vowel cluster. The present experiment compared the fusibility of stop/liquid and fricative/liquid pairs. Both initial clusters occur very frequently in English, but Day (1968) reported results of pilot work showing infrequent fusions for stimuli involving fricative phonemes.

Method

In addition to the BED set (BED, LED, RED) and the GO set (GO, LOW, ROW), the fricative stimuli FED and FOE were synthesized. The fricative /f/ was chosen because it is the only fricative in English that clusters with both /r/ and /l/ in initial position. Fricative stimuli were identical to the stop stimuli in duration, pitch, and intensity, and differed only in the acoustic structure of the first 100 msec. BED and GO have formant transitions in this region requisite for the perception of stop consonants, plus a portion (50 msec) of steady state vowel. FED and FOE, on the other hand, began with bandpass noise (above 1,500 Hz) and aspirated transitions requisite for the perception of /f/. A given liquid stimulus such as LED was paired with both a stop consonant stimulus (BED/LED) and a fricative stimulus (FED/LED). All stimuli and possible fusions were English words or names: BED/LED → BLED, BED/RED → BREAD, FED/LED → FLED, FED/RED → FRED, GO/LOW → GLOW, GO/ROW → GROW, FOE/LOW → FLOW, and FOE/ROW → FRO.

One tape was prepared with stop/liquid pairs and another tape with fricative/liquid pairs. Each tape consisted of 120 dichotic trials: (2 sets of stimuli) × (2 consonant/liquid pairs per set) × (3 lead times) × (2 channel arrangements) × (5 observations per pair). Twelve subjects listened to both tapes: half listened first to the fricative/liquid stimuli and then to the stop/liquid stimuli, while the others listened in the reverse order.

Results

Fusions occurred more readily for stop/liquid pairs than fricative/liquid pairs. Fusion rates were 57% and 18%, respectively. This 3:1 ratio was highly significant, with all subjects showing greater fusion rates for stop/liquid items ($z = 3.18$, $p < .001$). Unlike the results of Experiments 1 and 3, fusion rate differences of this experiment cannot be accounted for by the relative frequency of the possible fusion responses in English: for example, BLED and GLOW are much less common than FLED and FLOW (Carroll et al., 1971). Furthermore, initial /f/ + liquid clusters occur at about the same frequency as initial /b/ + liquid clusters in English, and considerably more frequently than initial /g/ + liquid clusters (Denes, 1965; Hultzén, Allen, & Miron, 1964).

The order in which subjects listened to the stop and fricative test tapes was not a significant factor. Fusion rates for stop + /l/ and stop + /r/ stimuli were within a few percentage points. Fricative + /l/ and fricative + /r/ fusion rates were also comparable. The /l/-for-/r/ substitutions were frequent for stop/liquid pairs, but not for fricative/liquid pairs. In fact, /r/ substitutions occurred on 64% of all trials in which fricative + /l/ stimuli fused, while corresponding /l/ substitutions were rare. A second type of substitution also occurred. Fricative/liquid stimuli did not always yield fricative + liquid responses; for example, FED/RED → BRED. In fact, about 70% of all fricative/liquid pair fusions were actually stop + liquid responses. Stop-for-fricative substitutions were not the result of poor fricative stimuli, since subjects identified them in isolation on a diotic test with a high degree of accuracy. Instead, these substitutions appear to be an extension of the differences in fusibility between the stops and fricatives.

Experiment 10: Phoneme Level: Liquids and Semivowels

The present experiment examined the second consonant in the to-be-fused cluster. Stimuli beginning with semivowels (i.e., /w/ and /y/) were generated to see whether they would fuse as readily as the liquids, and to see whether the /l/-substitution effect would be extended to /w/ and /y/.

Method

Two sets of stimuli were used: the KICK set (KICK, LICK, RICK used previously, plus WICK) and the COO set (COO, LIEU, RUE, YOU) generated on the Haskins' synthesizer. The only stop consonant that clusters with all liquids and semivowels in English is /k/; yet /ky/ occurs only before the vowel /u/, while /kw/ does not occur before /u/. Thus, it was necessary to synthesize two sets of stimuli, one for /l, r, w/ and the other for /l, r, y/. (Note that LIEU could also be represented as LOU.) Liquid and semivowel stimuli within the same set were identical in all respects except for the direction and slope of the second- and third-formant transitions, as shown in Figure 2. These are the cues that distinguish all liquids and semivowels (Lisker, 1957; O'Connor et al., 1957).

All stimuli and possible fusion responses for both sets were common English words or names: KICK/LICK → CLICK, KICK/RICK → CRICK, KICK/WICK → QUICK, COO/LIEU → CLUE, COO/RUE → CREW, and COO/YOU → CUE. A tape was prepared with 108 dichotic items: (2 sets of stimuli) × (3 liquid and semivowel stimuli per set) × (3 lead times) × (2 channel arrangements) × (3 observations per pair). Twelve subjects listened to two passes through the tape, reversing headphones after the first pass.

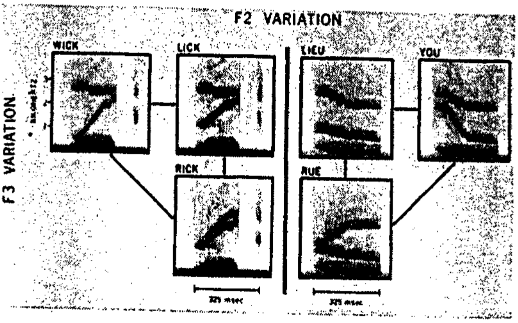

FIGURE 2. Spectrograms of liquid and semivowel stimuli used in Experiment 10.

Results and Discussion

KICK/LICK, KICK/RICK, and KICK/WICK pairs all fused at an average rate of 70%, while COO/LIEU, COO/RUE, and COO/YOU all fused at an average rate of 42%. There were no significant differences within each set. However, regardless of which stimuli were presented, most responses were stop + /l/; /l/ was substituted for /r/, as in previous studies, and it was also substituted for /w/ and /y/. CLICK and CLUE responses occurred in 89% of all trials in which fusions occurred. Again, word frequency of possible fusions cannot account for the substitutions. For example, QUICK is much more common than CLICK, and CREW is more common than CLUE (Carroll et al., 1971). Nevertheless, the /l/ substitutions for both sets of stimuli yielded relatively common English words. The data of Day (1968) suggest that when /l/ substitutions do not yield acceptable words, they occur considerably less often.

Difficulties with /r/ versus /l/ occur in a wide variety of other situations besides phonological fusion. Children, for example, have more difficulty in pronouncing /r/ than /l/ and sometimes pronounce both phonemes as /l/ (Morley, 1957; Murray, 1962; Powers, 1957); the deaf have more trouble processing /r/ than /l/, and often hear /l/ in both cases (Rosen, 1962). The articulation pattern of /r/ is less stable than /l/ (Bronstein, 1960; Delattre, Note 3), and /r/ is more difficult to pronounce correctly under conditions of delayed auditory feedback than /l/ (Applegate, 1968). The dichotic fusion results are complementary to these studies: Perhaps /r/ is less stable than /l/ in perception and production, and the fusion results of this study are a manifestation of /r/ instability.

Overview of Experiments 9 and 10

When the first consonant in the to-be-fused cluster is a stop, fusion rate is high, but when it is a fricative, fusion occurs considerably less often. The role of the second consonant in the to-be-fused cluster is less clear: Fusion occurred equally well for all stop/liquid and stop/semivowel pairs, yet all pairs tended to yield a stop + /l/ response.

Experiment 11: Acoustic Level: Liquid Transitions

Linguistic cues at the sentence and phoneme levels affect fusion rate. Perhaps linguistic cues at the level of acoustic structures are also important. Since the liquid is perceptually interpolated between the stop and the vowel in the fusion response, one key to fusion may lie in the acoustic structure of the liquid. Experiments 11–13 examine aspects of the acoustic structure of liquids.

Experiment 10 showed that for the present sets of stimuli, liquids (/r, l/) and semi-

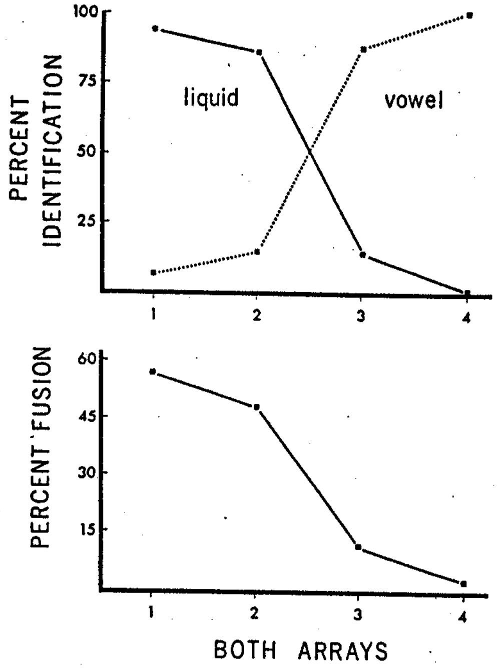

FIGURE 3. Results of identification and fusion tasks involving liquid-to-vowel stimulus arrays in Experiment 11.

vowels (/w, y/) tend to yield stop + /l/ fusions. It would appear that any combination of rising and falling second- and third-formant transitions is sufficient for stop + /l/ fusions to occur. If any combination is sufficient for fusion, perhaps no transitions are needed at all. For example, PAY/AY, a pair without any liquid transitions, might also yield stop + /l/ fusions. The present study varied the slope of the liquid transitions to determine the extent of formant transitions necessary for fusion to occur and to gain additional support for the view that fusion is not a response bias.

Method

The PAY set and the KICK set were altered to include five stimuli, one stop stimulus and four stimuli that formed a liquid-to-vowel continuum. At one end of the continuum, Stimulus 1 had full liquid transitions in all formants, as found in LAY and LICK. At the other end of the continuum, Stimulus 4 had the same duration but began with a steady state vowel, AY and ICK. Between the extremes were Stimuli 2 and 3 with intermediate amounts of formant transitions. Equal increments of acoustic change occurred between successive stimuli. Sixteen Yale undergraduates participated in two tasks, dichotic fusion and then diotic identification of the liquid-to-vowel stimuli in isolation. Since the results of the identification task were highly relevant to the fusion task, those data are discussed first.

A diotic identification tape of 80 randomized liquid-to-vowel items was prepared: (2 sets of stimuli) × (4 stimuli per array) × (10 observations per stimulus). There was a 3-sec interval between each item. Subjects wrote down the single item that they heard presented on each trial. Dichotic items were constructed by pairing the stop stimuli with all items in the liquid-to-vowel arrays. The tape consisted of 96 pairs: (2 sets of stimuli) × (4 stimuli per array) × (3 lead times) × (2 channel arrangements) × (2 observations per pair). Subjects listened to two passes through the tape, reversing headphones after the first pass. As usual they wrote their responses, indicating what they heard.

Results and Discussion

Stimuli 1 and 2 were identified as beginning with /l/ on 88% of all trials, as shown in the top panel of Figure 3. Stimuli 3 and 4 were identified as beginning with /l/ on only 8% of all trials. All subjects showed this quantal trend. There was no difference between the LAY-to-AY and LICK-to-ICK stimulus arrays.

Fusion occurred at a rate of 52% for pairs containing Stimulus 1 or Stimulus 2, the stimuli which had been identified as beginning with a liquid. Other pairs yielded only 6% fusions (stop + liquid responses), as shown in the lower panel of Figure 3. Thus, liquid-like transitions are necessary for phonological fusion: Fusion occurs in direct proportion to the extent that liquid-like items are indeed perceived as liquids in isolation.

Experiment 12: Acoustic Level: Degraded Liquids

Experiment 11 showed that transitions in the liquid stimulus were necessary for fusion to occur. The present experiment was designed to determine which formant transition (or combination of transitions) in the liquid is necessary for fusion.

Method

The PAY set and the KICK set were again used. Liquid stimuli appeared in many forms. Some were degraded in that certain formants were omitted from their acoustic structure. Figure 4 shows the component parts of liquid stimuli. All possible combinations of all formants were used. Eleven liquid-like stimuli resulted in each set: 2 three-formant stimuli identical to those used in previous studies, 5 two-formant stimuli, and 4 one-formant stimuli, as listed in Table 2.

Each of the 11 liquid-like stimuli was paired with its appropriate stop stimulus. In addition, two control pairs were constructed per set: One pair was a stop/stop pair (PAY/PAY and KICK/KICK like those used by Day, 1968), and the other was a stop presented to one ear and nothing to the other (PAY/——— and KICK/———). No fusion responses should occur for control pairs if fusion occurs only for pairs containing liquid-like stimuli. Twelve subjects listened to a dichotic tape of 156 items: (2 sets of stimuli) × (13 pairs per set) × (3 lead times) × (2 channel arrangements per pair).

Results and Discussion

There were two general fusion rates: Most pairs fused at a rate of about 55%, but a few pairs fused at a considerably lower rate, as shown in Table 2. Pairs that rarely fused contained liquid-like stimuli with only the first formant or the third formant of /l/. Figure 4 shows that these stimuli lacked formant transitions in the mid-frequency range (1,000–2,000 Hz), while all other liquid-like stimuli had transitions in this region (either the second formant or the third formant of /r/).

TABLE 2

Percent Fusion Responses for All Pairs in Experiment 12

Dichotic pairs	Percent fusion
Stop + three-formant liquids	
1, 2, 3/l/	57
1, 2, 3/r/	51
Stop + two-formant liquids	
1, 2	57
1, 3/l/	23
1, 3/r/	51
2, 3/l/	55
2, 3/r/	54
Stop + one-formant liquids	
1	19
2	60
3/l/	16
3/r/	64
Control items	
Stop + stop	6
Stop + blank	2

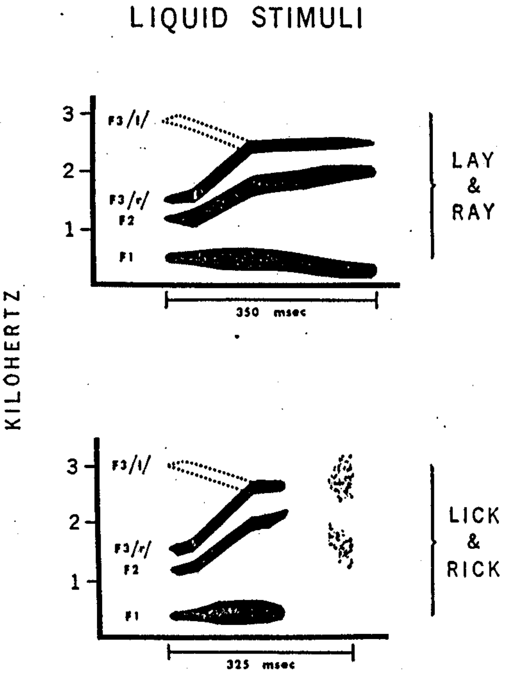

Figure 4. Schematic spectrograms of liquid stimuli and their component parts used in Experiment 12. (F1 = first formant, F2 = second formant, F3 = third formant.)

Stop + three-formant liquid stimuli fused at rates comparable to previous studies: 54%. Stop + two-formant liquid pairs fused at a rate of 52%, with the exception of the pairs with only first and third formants of /l/, where fusion level was only 23%. All subjects showed this drop in fusion rate ($z = 3.18$, $p < .001$). Stop + one-formant liquid pairs also showed high and low fusion rates, following a pattern similar to that of pairs with two-formant liquids. Again, all subjects showed these quantal differences in fusion rate. As in previous studies, stop + /l/ responses occurred on more fused trials than did stop + /r/ responses. Stop/stop pairs yielded few stop + /l/ responses. Such responses would be "false fusions," since the subject would be reporting a liquid when, in fact, none was presented (Day, 1968).

A specific acoustic cue for phonological fusion of stop/liquid stimuli appears to be the second-formant transition or the third-formant /r/ transition of the liquid stimulus. A single rising formant transition in the range 1,000–2,000 Hz is necessary for fusion to occur.

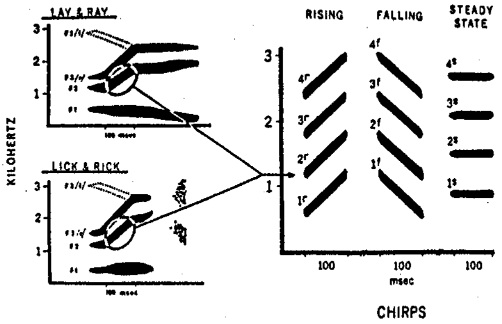

FIGURE 5. Schematic spectrograms of liquid stimuli and liquid "chirps" used in Experiment 13. (The original second-formant transition from the liquid stimuli is the rising chirp stimulus designated 2^r.)

Experiment 13: Acoustic Level: Liquid "Chirps"

The previous study found that a mid-range formant transition was necessary for fusion to occur. The present study was designed to determine whether that transition or any other is sufficient for fusion when paired with a stop stimulus. When the second-formant transition is removed from a liquid stimulus and synthesized by itself, it sounds similar to a bird's twitter, hence the name "chirp." Mattingly, Liberman, Syrdal, and Halwes (1971) and Wood (1975) have demonstrated that these chirps are not processed as speech. Chirp stimuli have two general features, relative frequency range and direction (rising vs. falling).

Method

Two stop stimuli were synthesized, PAY and KICK. Liquid stimuli were degraded so that only the second-formant transition remained, a 100-msec chirp rising rapidly from a value of 1,200 Hz to 1,800 Hz. This and other chirps were synthesized at the same amplitude as the second-formant transition in the full liquid stimulus. Twelve chirp stimuli were used; there were four frequency values for rising, falling, and steady state chirps, as shown in Figure 5. Specific stimuli are numbered from low to high according to ordinal position on the frequency scale, and each is given a superscript to designate the class to which it belongs. Endpoints for rising and falling chirps were 600, 1,200, 1,800, 2,400, and 3,000 Hz, while steady state chirps had frequencies of 900, 1,500, 2,100, and 2,700 Hz. The original second-formant transition from the liquid stimuli is the rising chirp stimulus designated 2^r.

In addition to the stop/chirp pairs, there were control pairs. Ordinary fusible items such as PAY/LAY and PAY/RAY were employed to obtain baseline fusion rates. Stop/stop pairs were also included to set a lower boundary on fusion rate since Experiment 12 found few stop + liquid responses to such trials. Thus the control pairs provided boundary conditions within which to compare the fusion rates for stop/chirp items. Three lead times were used such that stop/chirp stimulus pairs had the same temporal relationships as the stop and the second-formant transition of the full liquid in previous experiments. Twelve subjects listened to a dichotic tape of 180 items: (2 sets of stimuli) × (15 pairs per set, 12 stop/chirp pairs plus such control pairs as PAY/LAY, PAY/RAY, and PAY/PAY) × (3 lead times) × (2 channel arrangements per pair).

Results

Fusions, or stop + liquid responses, occurred at a substantially reduced rate for all stop/chirp pairs. While fusion rate for stop/liquid control pairs was 47%, a rate comparable to previous studies, fusion rates for stop/chirp pairs averaged only 8%. This difference was highly significant ($z = 3.18$, $p < .001$).

Fusion did occur, however, for selected stop/chirp pairs as shown in Table 3. Pairs with rising chirps yielded an average fusion rate of 14%, higher than all other stop/chirp pairs combined ($z = 2.6$, $p < .005$). Within this category of rising chirps, fusion occurred most readily for pairs involving chirp number 2^r. Eight of 12 subjects fused at a higher rate for these pairs than for any other stop/chirp pair ($z = 4.0$, $p < .0001$), but even here fusions occurred significantly less often than for the stop/liquid control pairs ($z = 3.18$, $p < .001$). Thus, even the chirp stimulus most appropriate to the full liquid

TABLE 3

PERCENT FUSION RESPONSES FOR STOP/CHIRP PAIRS IN EXPERIMENT 13

Stimulus number and frequency domain	Class of chirp stimulus			
	Rising	Falling	Steady state	*M*
1 (600–1,200 Hz)	14	5	5	8
2 (1,200–1,800 Hz)	24	6	7	12
3 (1,800–2,400 Hz)	12	5	5	7
4 (2,400–3,000 Hz)	7	8	3	6
M	14	6	5	

stimulus is not entirely sufficient for fusion to occur at an unreduced rate.

The /l/-for-/r/ substitutions occurred for stop/liquid control pairs at rates comparable to previous studies. Fusion of stop/chirp pairs, however, were not dominated by stop + /l/ responses. In fact, only stop/2^r pairs yielded more stop + /l/ fusions than stop + /r/, /w/, or /y/ fusions. Fusions for the lowest frequency chirps paired with a stop stimulus were dominated by stop + /w/ responses, whereas those for highest frequency chirps were dominated by stop + /y/ and stop + /l/ responses. False fusion responses for stop/stop pairs occurred only 2% of the time.

Overview of Experiments 8–13

Three levels of linguistic cues were explored (the semantic level, the phoneme level, and the acoustic level), and the effect of cues at each level on fusion rate was observed. The results suggest that the cognitive processes involved in phonological fusion are influenced by cues at *all* three linguistic levels. Fusion rate was enhanced when fusible pairs were imbedded in sentence contexts; fusion occurred best for certain classes of phonemes; and specific acoustic cues were found to be important for fusion. Thus, linguistic cues within and outside of the consonant/liquid pairs have a distinct effect on fusion rate. Given the appropriate experimental situation, cues from all three levels appear to work in concert. Consider one of the sentence pairs from Experiment 8: THE TREES ARE GOING AGAIN presented to one ear and THE TREES ARE ROWING AGAIN to the other. Semantic cues at the sentence level increased fusion rate for GO/ROW pairs beyond the rate they yielded when presented in isolation. Cues at the phoneme level were also influential: Experiment 9 demonstrated that GO/ROW pairs fused at a higher rate than FOE/ROW pairs. Experiment 12 found that the second-formant transition in the liquid was a specific cue for fusion. Moreover, this cue appears to be pertinent to the /l/-for-/r/ substitutions: Fusion rates and /l/ substitutions were approximately equal for stop/second-formant and stop/liquid pairs. Thus, the high rate of THE TREES ARE GLOWING AGAIN responses appears to be the result of the synergy of cues from three very different levels of language. For further discussion of linguistic influences on phonological fusion, see Cutting and Day (1975).

Additional Comments

Day (1973; Note 1) found marked individual differences in phonological fusion using natural speech disyllabic pairs. In the present series of fusion experiments, using synthetic speech monosyllabic fusible items, such individual differences did occur, but to an extent less marked than that found by Day. Consider fusion rates for the 140 individuals in Experiments 1, 2, and 4–10 (the other 48 individuals are eliminated because fusion broke down in their tasks), and consider the number of subjects whose fusion rates fell in the five quintiles between 0% and 100% fusion. Only 7 subjects' rates fell between 0% and 20%, whereas 39 fell between 21% and 40%, and 54 between 41% and 60%. Only 14 subjects' fusion rates were between 61% and 80%, whereas 26 were between 81% and 100%. This overall bimodal trend is indicative of subjects within each of the nine experiments in question. It is similar to that found by Day, but the lower fusion mode has been shifted "upward" in frequency through the use of higher fusing synthetic pairs.

Phonological fusion has a direct analog in vision. In fact, Ruth Day discovered this phenomenon in search of an auditory parallel to the work of Rommetveit and his coworkers (Rommetveit, Berkeley, & Brøgger, 1968; Rommetveit & Kleven, 1968; Rommetveit, Toch, & Svendsen, 1968a, 1968b). They found that when letters such as SHAR were presented to one eye and SHAP to the other eye, many subjects reported seeing SHARP.

Conclusion

Phonological fusion occurs in dichotic listening in a very systematic fashion. Within a relatively wide range of variation, temporal alignment and physical attributes of the signals appear to matter very little. Linguistic aspects of the stimuli, however, matter quite a lot. Stop consonants and liquids fuse most readily, and fuse in the order in which they are phonologically reasonable in English. Given a stop stimulus and a liquid stimulus of the same general pattern, linguistic attributes within and outside of the dichotic pairs contribute to the percept.

REFERENCE NOTES

1. Day, R. S. *Temporal-order judgments in speech: Are individuals language-bound or stimulus-bound.* Paper presented at the meeting of the Psychonomic Society, St. Louis, Missouri, November 1969. (Also in Haskins Laboratories *Status Report on Speech Research,* 1969, *SR-21/22,* 71–87.)
2. Day, R. S., & Cutting, J. E. *Levels of processing in speech perception.* Paper presented at the meeting of the Psychonomic Society, San Antonio, Texas, November 1970.
3. Delattre, P. C. A dialect study of American /r/ by x-ray motion picture. *The general phonetic characteristics of languages* (Final Report). University of California at Santa Barbara, 1967.

REFERENCES

Applegate, R. B. Segmental analysis of articulatory errors under delayed auditory feedback. *Project on Linguistic Analysis* (University of California at Berkeley), 1968, *8,* 1–27.

Bronstein, A. M. *The pronunciation of American English.* New York: Harper & Row, 1960.

Carroll, J. B., Davies, P., & Richman, B. (Eds.). *Word frequency book.* New York: Houghton & Mifflin, 1971.

Cooper, F. S., & Mattingly, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America,* 1969, *46,* 115. (Abstract)

Cutting, J. E., & Day, R. S. The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics,* 1975, *3,* 9–23.

Day, R. S. Fusion in dichotic listening (Doctoral dissertation, Stanford University, 1968). *Dissertation Abstracts International,* 1969, *29,* 2649B. (University Microfilms No. 69-211)

Day, R. S. Temporal-order perception of a reversible phoneme cluster. *Journal of the Acoustical Society of America,* 1970, *48,* 95. (Abstract)

Day, R. S. Digit span memory in language-bound and stimulus-bound subjects. *Journal of the Acoustical Society of America,* 1973, *54,* 287. (Abstract)

Denes, P. B. On the statistics of spoken English. *Journal of the Acoustical Society of America,* 1965, *35,* 892–904.

Hultzén, L., Allen, J., & Miron, M. *Tables of transitional frequencies of English phonemes.* Urbana: University of Illinois Press, 1964.

Lisker, L. Minimal cues for separating /w, r, l, y/ in intervocalic position. *Word,* 1957, *13,* 257–267.

Mattingly, I. G., Liberman, A. M., Syrdal, A. K., & Halwes, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology,* 1971, *2,* 131–157.

Morley, M. E. *Development of speech disorders in childhood.* London: Livingstone, 1957.

Murray, R. S. The development of /r/ in the speech of preschool children (Doctoral dissertation, Stanford University, 1962). *Dissertation Abstracts International,* 1963, *23,* 4462. (University Microfilms No. 63-2734)

O'Connor, J. D., Gerstman, L. S., Liberman, A. M., Delattre, P. C., & Cooper, F. S. Acoustic cues for the perception of initial /w, y, r, l/ in English. *Word,* 1957, *13,* 25–43.

Pisoni, D. B. Perceptual processing time for consonants and vowels. *Journal of the Acoustical Society of America,* 1973, *53,* 369. (Abstract)

Powers, M. H. Functional disorders of articulation: Symptomatology and etiology. In L. E. Travis (Ed.), *Handbook of speech pathology.* New York: Appleton-Century-Crofts, 1957.

Rommetveit, R., Berkeley, M., & Brøgger, J. Generation of words from tachistoscopically presented nonword strings of letters. *Scandinavian Journal of Psychology,* 1968, *9,* 150–156.

Rommetveit, R., & Kleven, J. Word generation: A replication. *Scandinavian Journal of Psychology,* 1968, *9,* 277–281.

Rommetveit, R., Toch, H., & Svendsen, D. Effects of contingency and contrast on the cognition of words. *Scandinavian Journal of Psychology,* 1968, *9,* 138–144. (a)

Rommetveit, R., Toch, H., & Svendsen, D. Semantic, syntactic, and associative context effects in stereoscopic rivalry situation. *Scandinavian Journal of Psychology,* 1968, *9,* 145–149. (b)

Rosen, J. Phoneme identification in sensorineural deafness (Doctoral dissertation, Stanford University, 1962). *Dissertation Abstracts International,* 1962, *23,* 2253. (University Microfilms No. 62-5510)

Studdert-Kennedy, M., Shankweiler, D. P., & 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.

Thorndike, E. L., & Lorge, I. *The teacher's word-book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

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.

Wood, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *104,* 3–20.

(Received May 1, 1974)