Journal of Phonetics (1975) 3, 99–113

The perception of stop-liquid clusters in phonological fusion

James E. Cutting
Wesleyan University, Middletown, Conn., and Haskins Laboratories, New Haven, Conn. 06510, *U.S.A.*
and Ruth S. Day
Yale University and Haskins Laboratories, New Haven, Conn., U.S.A.

(*Received 1st March* 1975)

Abstract: Phonological fusion occurs when items such as PAY and LAY are presented separately to each ear and listeners report hearing PLAY. Input items that begin with a stop consonant (e.g. /p/) and a liquid (/l/ or /r/) fuse especially well. The present studies examined the effect of various factors on the frequency of fusion responses. Allophonic variation in the liquids ("trilled" versus plain) had no effect on fusion frequency. Phonemic similarity also had no effect; that is, when the input items differed in all phonemes (PAY and LED) they still fused. However, the phonemic order and location of clusters within a syllable did have a large effect: initial stop-liquid clusters fused readily (PAY/LAY yielded PLAY) while final liquid-stop clusters rarely fused (PEEL/PEED rarely yielded PEELED). Various fusion phenomena remained the same when tested in both identification and discrimination paradigms. Finally, fusion scores were not normally distributed over subjects; that is, some subjects fused on most or all trials, while others fused less frequently.

When a different message is presented to each ear in the dichotic listening procedure, fusion often occurs (Day, 1968). For example, the dichotic pair BANKET/LANKET→ (yields) BLANKET. This type of fusion is called phonological because the process appears to take place at the linguistic level where phonemes are combined into sequences. Fusion responses obey phonological rules concerning permissible sequences in the language. For example, in English, clusters consisting of a stop consonant followed by a liquid can occur in initial position, but liquid-stop clusters cannot. Thus BANKET/LANKET→BLANKET, not LBANKET. However, when the target clusters are reversible, fusion occurs in both directions: TASS/TACK→TASK and TACKS (Day, 1970*b*). Phonological fusions do not occur when potential clusters are prohibited, e.g., BAD/DAD does not yield BDAD or DBAD.

Several studies have been performed to determine the extent to which the fusion phenomenon can be affected by "higher level" linguistic and "lower level" nonlinguistic factors. In the linguistic domain, the status of the inputs and fusions as words or nonwords has been examined (Day, 1968, Experiment II). Fusions are especially frequent when words are formed from nonwords (BANKET/LANKET→BLANKET). However, fusions also occur when the inputs are words (BACK/LACK→BLACK). Nonword fusions also occur (GORIGIN/LORIGIN→GLORIGIN; GAB/LAB→GLAB). At the sentence

level, a given item fuses more readily when it is embedded in a sentence than when it appears in isolation: THE TRUMPETER PAYS FOR US/THE TRUMPETER LAYS FOR US versus PAY/LAY (Cutting, 1975). At the phonological level, some clusters fuse more readily than others (Day, 1968). For example, stop-liquid pairs (GO/LOW→GLOW) fuse more readily than fricative-liquid pairs (FOE/LOW→FLOW) (Cutting, 1975). In summary, favorable word, sentence, and cluster conditions increase fusion frequency, although a substantial number of fusions still occur even when these conditions are absent.

Nonlinguistic conditions have very little effect on phonological fusion. When the input items are varied over a wide range of values in terms of fundamental frequency (as much as 20 Hz difference between the items), intensity (15 db difference), or relative onset time (150 ms difference), there is virtually no change in fusion scores (Cutting, 1975).

The present studies were designed to study the effects of other factors on the frequency of fusion responses. Experiment I establishes baseline data for items to be used in the other four experiments. Experiment II examines the effect of allophonic variation: different acoustic forms of the "same" phonemes are used. Experiment III examines the effect of phoneme order and cluster position: initial stop-liquid clusters are compared with final liquid-stop clusters. Experiment IV examines phonemic similarity: pairs that differ in all phonemes (PAY/LED) are compared with those that differ only in their initial phonemes (PAY/LAY). Experiment V compares the perception of fusible dichotic items in identification and discrimination paradigms. A final section considers findings common to all the experiments, namely those concerned with individual differences.

Experiment I: baseline data[1]

Since the basic purpose of the studies reported here was to examine the effect of various factors on the frequency of fusion responses, it was necessary to obtain baseline data before introducing variation in these factors. Experiment I was designed to obtain such baseline data. In addition it presents a procedural standard on which the other experiments are based, and therefore its methods are described in considerable detail while those of later experiments are presented in abbreviated form.

Stimuli and tapes

Four sets of fusible stimuli were generated on the Haskins Laboratories' parallel-resonance synthesizer: the BED set (BED, LED, RED), the PAY set (PAY, LAY, RAY), the CAM set (CAM, LAMB, RAM), and the GO set (GO, LOW, ROW). The items are given in orthographic rather than phonetic form in order to highlight the fact that all were permissible one-syllable words, as were the possible fusions (BLED, BREAD, PLAY, PRAY, CLAM, CRAM, GLOW, GROW). The items in each set differed only in their initial phonemes; one item began with a stop consonant (/b/, /p/, /k/, or /g/) while the others each began with a liquid consonant (/l/ or /r/). Only one stop-initial item (called simply "stop" for convenience) was synthesized for each set. However, five different acoustic forms were prepared for the liquid-initial items (called "liquids"). They differed only in starting frequency, direction, and extent of the third-formant (F3) transition, and hence varied along an acoustic continuum known to distinguish /l/ from /r/ (O'Connor, Gerstman, Liberman, Delattre & Cooper, 1957). These variations are labelled A through E according to their ordinal position in the F3 transition arrays as shown in Fig. 1. The

[1] A preliminary version of Experiment I was presented at the 83rd meeting of the Acoustical Society of America, Buffalo, New York, April, 1972, entitled "Dichotic fusion along an acoustic continuum".

acoustic form of each A-item is characteristic of /r/ while that of the E-items is characteristic of /l/. Each stop item was identical to the liquid items in its set except for the initial 150 ms which consisted of formant transitions appropriate for the perception of a stop consonant.

All items were digitized, stored on disc file, and then recorded on audio tape using the Pulse Code Modulation system at Haskins (Cooper & Mattingly, 1969). Two types of tapes were prepared. Single-item identification tapes were prepared in order to determine how well the individual stimuli could be identified when presented alone. Thus the same item was recorded on both channels of the audio tape such that a single item was perceived "in the center of the head" when presented over stereo earphones. One identification tape was prepared for stops with 40 items: (4 stop items, one/set) × (10 observations/item);

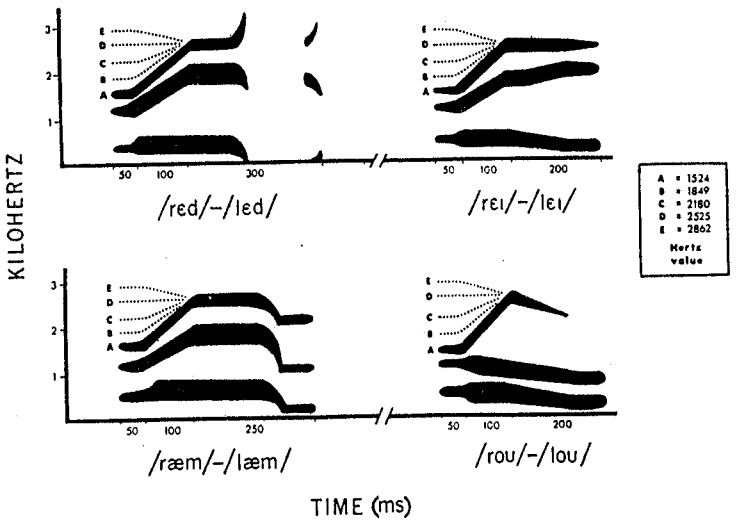

Figure 1 Schematic spectrograms of the liquid stimuli. The letters A through E identify the F3 transitions used for different stimuli in each set. Start frequencies for each F3 are shown in the box (Experiment I).

and another tape for the liquids with 200 items: (5 liquids/set) × (4 sets of stimuli) × (10 observations/item). The inter-item interval for both tapes was 2·0 s. The presentation order of items was random in these and all other tapes.

A dichotic tape was prepared with the stop member of a given set recorded on one channel of the audio tape and one of its liquid members on the other channel. There were three relative onset times for each pair: the stop led by 50 ms, the liquid led by 50 ms, or both items began at the same time. These same "lead times" were used in all subsequent dichotic tapes. The Experiment I tape had 120 dichotic pairs: (5 stop-liquid pairs/set) × (4 sets) × (3 lead times) × (2 channel arrangements/pair)[2], with 4·0 s between pairs.

Subjects

Subjects were selected according to the same criteria in all experiments. All were Yale University undergraduates who were right-handed, native American English speakers,

[2]That is, given the assignment of stop and liquid items to the two channels of the audio tape in one arrangement, the items were reversed in the other arrangement.

with no history of hearing trouble. They received course credit for participating. In Experiment I, 16 subjects were tested; different subjects participated in the others.

Procedure

Since the dichotic test was of primary interest, it was given first, so that subjects would have no prior experience with the items. Subjects were asked to write down "what they heard", no matter whether they heard "one word or two", "real words or nonsense", following the procedure of Day (1968). Practice trials were given and procedural questions were answered prior to beginning the main test. Halfway through the main test the headphones were reversed.

In the single-item identification tests, subjects were asked to write down a letter to indicate the first "sound" (phoneme) of each item, *B*, *P*, *K* or *G* for the stop tape, and *L* or *R* for the liquid tape. Before the liquid identification test was given, subjects listened to pairs of stimuli from each set, namely the most "/r/-like" and "/l/-like" items (stimuli A and E, respectively). As they listened to these endpoint stimuli, subjects saw the written version of each item listed in chronological order on a sheet of paper.

Subjects were tested in groups of four in a sound-treated room. The tapes were played on an Ampex AG500 tape recorder and the output signals were sent through a listening station to matched Telephonics earphones (model TDH39) at 80 db re 20 $\mu N/m^2$.

Results and discussion

In the single-item identification tasks, stops were identified with near-perfect accuracy: 98% of all responses were correct. All four stimulus sets showed comparable results and therefore the data were pooled. Items in each array with low-initial and fast-rising F3 transitions (stimuli A and B) were perceived as beginning with /r/ on 95% of trials while those at the opposite end of the continuum (D and E) were perceived as beginning with /l/ on 94% of trials. The middle stimulus (C) was perceived as /r/ on roughly half the trials and as /l/ on the other half. This sharp shift in perception from /r/ to /l/ is similar to the results obtained for simple liquid-vowel syllables varied along a comparable F3 continuum (O'Connor *et al.*, 1957).

In the dichotic task, fusions occurred on 57% of all trials. When subjects did not fuse, they usually reported the stop item presented on that trial. Fusion scores for each set were: 40% for the BED set, 75% for the PAY set, 64% for the CAM set, and 46% for the GO set. These figures are comparable to results obtained previously for the same stops paired with A and E liquids (Cutting, 1975). Fusions were more frequent[3] for stop-leading pairs (60%) than for simultaneous (54%) or liquid-leading pairs (54%). This effect occurred in some of the other conditions reported in the rest of the paper. However the overall trend is not very convincing since in many cases the statistical tests were of borderline significance ($P = 0{\cdot}05$), and in several conditions no lead time effects occurred. The failure of lead time to produce reliable changes in fusion scores was observed in previous studies (Day, 1970*a*, 1970*b*; Cutting, 1975). Other factors that yielded no significant differences in this or subsequent experiments included assignment of the stop and liquid to each tape channel and/or ear.

Fusion occurred with comparable frequency for all forms along the liquid continuum. However, the particular liquid reported in the stop-liquid cluster did not always correspond to the identity of the presented liquid as determined by the single-item identification results. The items clearly identified as /l/ (D and E) did yield fusion of stop + /l/, however

[3]All differences reported in the paper are significant at the $P < 0{\cdot}01$ level by a two-tailed sign test, unless indicated otherwise.

those identified as /r/ (A and B) were nevertheless reported as stop + /l/ in 76% of all fusion responses. This substitution of /l/ for /r/ occurred in all stimulus sets: it accounted for 59% of the total fusions given for BED/RED, 79% for PAY/RAY, 82% for CAM/RAM, and 92% for GO/ROW. Thus, although the F3 transition cue was crucial in the perception of single liquid items, it was almost irrelevant in perception of the present dichotic stop-liquid pairs.

This /l/-substitution effect was first reported by Day (1968) who found that GOCERY/ROCERY occasionally yielded GLOCERY. However, these substitutions were very infrequent in the Day studies (e.g., less than 4% of all fusions in Day, 1968, Experiment I). The /l/ substitution effect is minimal when the items are multisyllabic, when the inputs of a given pair differ considerably from each other in terms of overall duration, vowel quality, and general intonation contour (as in most natural speech productions), and when the possible stop +/l/ fusion is not a permissible word. The effect is maximal for single syllable items that are highly similar (in the present case identical) in duration, vowel quality, and intonation contours, and when the stop +/l/ fusion is a word. When natural speech items are prepared to be highly similar, they also yield many /l/-substitution fusions (Cutting, 1975). This curious phenomenon will be discussed in more detail in Experiment V.

Experiment II: allophonic variation

The acoustic structure of a given phoneme may vary widely over different productions and in different contexts and yet still be perceived as the "same" phoneme. Such allophonic variation is not linguistic in that it does not lead to perception of different phonemes or larger linguistic units. Yet it is crucial to language perception: concatenation of "prototype" phonemes without allophonic variation appropriate to the given context produces a hopeless jumble of unintelligible sounds. Thus allophonic variation lies in some sense at the "borderline" of the linguistic and nonlinguistic domains. Since phonological fusion is affected by factors in the linguistic domain but is relatively unaffected by factors in the nonlinguistic domain, it is of considerable interest to examine the effect of borderline factors such as those involved in allophonic variation. Experiment II introduced allophonic variation into the liquid stimuli by adding trills.

Method

Two sets of stimuli were used, the PAY set (PAY, LAY, RAY) and the KICK set (KICK, LICK, RICK). Only two liquids were used for each set. They were prepared as "best exemplars" for /r/ and /l/ and hence used pattern types A and E along liquid continua as described in Experiment I. Each /r/ and /l/ item was prepared in two allophonic forms, plain and trilled. The plain form corresponded roughly to American English pronunciation, and was the same as used in Experiment I. The trilled /r/ was similar to that found in Eastern European languages. It was prepared in the following manner. The first 50 ms of initial steady-state information was removed. The formant values in this segment were then used to make a shorter 25 ms segment. Finally three such 25 ms bursts were concatenated on to the remainder of the original item, each followed by 25 ms of silence. The "trilled /l/" was synthesized in the same manner, even though the articulatory gymnastics required to produce it may well be impossible.[4] All four liquid stimuli used in the PAY set are shown in Fig. 2.

[4] A trilled /r/ is discontinuant and not technically a liquid phoneme. However, the term "liquid" is used here as a generic term for /l/, /r/, and all of their allophones, including trills.

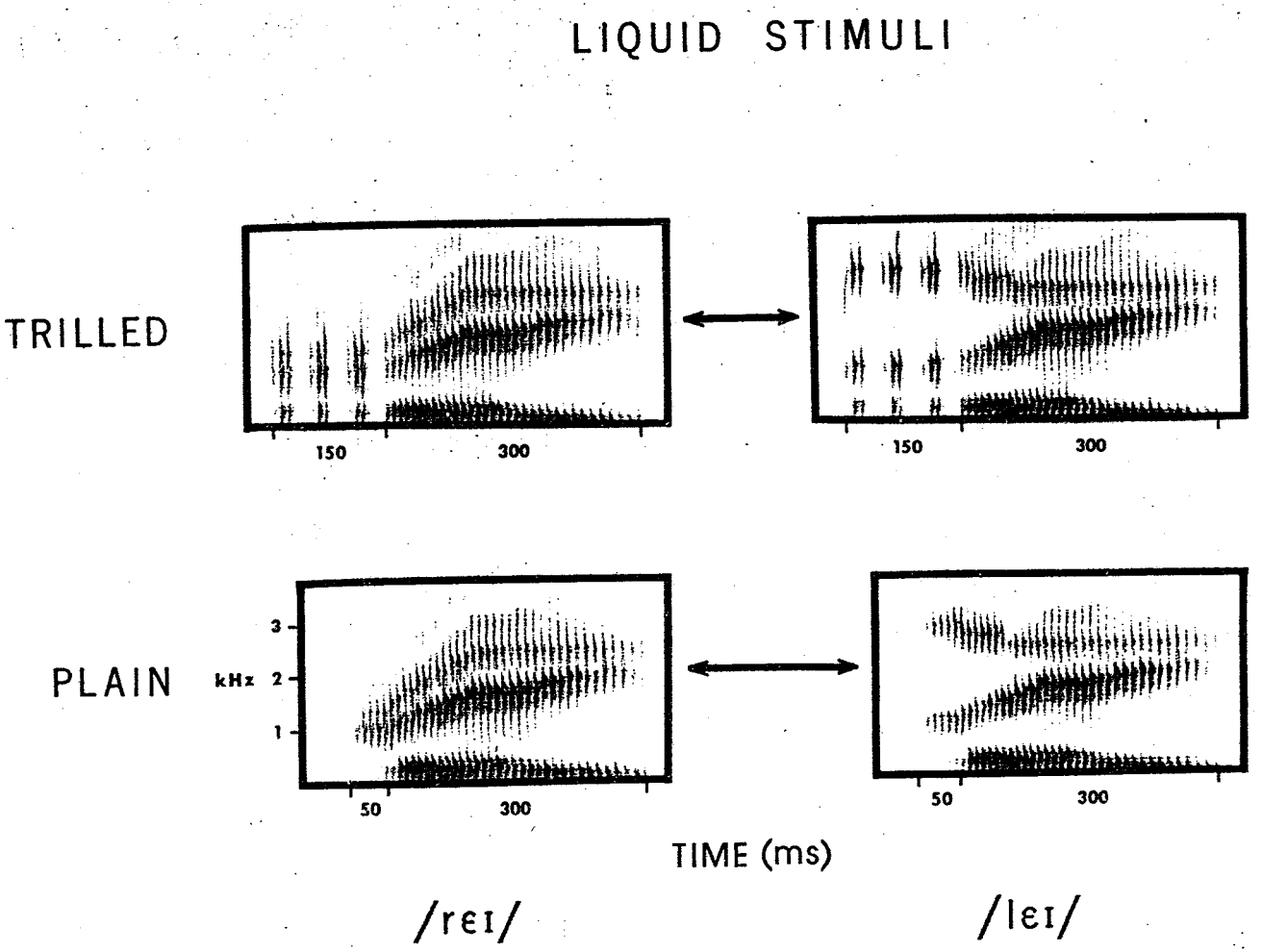

Figure 2 Wide-band spectrograms of two allophonic forms of RAY and LAY: trilled and plain (Experiment II).

A single-item liquid identification tape was prepared consisting of 80 items: (2 sets of stimuli) × (2 liquids/set) × (2 allophones/liquid) × (10 observations/stimulus), with 3·0 s between items. The 12 subjects wrote down *L* or *R* to indicate the first phoneme of every item. A dichotic tape was prepared consisting of 96 items: (2 sets of stimuli) × (4 stop-liquid pairs/set) × (3 lead times) × (2 channel arrangements) × (2 observations/pair), with 4·0 s between pairs. Listeners wrote down "what they heard".

Results and discussion

Identification functions for the plain liquid stimuli followed the same pattern described in Experiment I, as shown in Table I(a). However, most trilled liquids were identified as /r/. The trilled /l/ was identified as /r/ on 79% of the trials. Thus, the presence of trills greatly suppressed the effectiveness of the F3 transition cue observed to be important in previous liquid identification results.

Table I Identification scores for trilled and plain liquids (a) and fusion scores for fusible dichotic pairs containing the same liquids (b) (Experiment II)

(a) Identification of liquids

	Stimulus:	/l/		/r/	
	Response:	/l/	/r/	/l/	/r/
Trilled liquids		21%	79%	5%	95%
Plain liquids		88	12	7	93

(b) Fusion of stop-liquid dichotic pairs

			Breakdown of overall fusions*			
		Stimulus:	stop + /l/		stop + /r/	
	Overall Fusions	Response:	stop + /l/	stop + /r/	stop + /l/	stop + /r/
Trilled	50%		25%	75%	17%	82%
Plain	43		78	18	25	73

*Where fusions do not sum to 100%, the balance consists of stop + /w/ responses.

In the dichotic task, fusions occurred on 47% of all trials. They occurred both when plain and trilled liquids were paired with stops (43 and 50%, respectively). Thus phonological fusion occurred for liquid allophones which do not naturally occur in American English. However, listeners spontaneously distinguished between the two types of trials by writing in extra /r/s for many trilled-liquid pairs (e.g. PAY/RAY→PRRAY, KICK/RICK→KRRICK) or even multiple clusters (e.g. PAY/RAY→PRDRAY, KICK/RICK→KRTRICK). These responses were scored as fusions because they combined information presented to different ears into the same percept.

Internal analyses of the fusion data are shown in the bottom portion of Table I(b). As usual, most fusions on stop + plain /l/ trials contained stop + /l/ clusters. However, most fusion of stop + trilled /l/ trials contained stop + /r/ clusters; thus /r/ was substituted for the hard-to-pronounce stop + trilled /l/ clusters. In contrast to Experiment I fusion on stop + /r/ trials usually contained stop + /r/ clusters. The presence of a trill in the /r/

reduced the /l/-substitution effect to 17% of all fusions. Presence of trills in other test items on the tape reduced the /l/-substitution effect for plain /r/s to 25%. Table I shows that the relative proportion of stop + /l/ versus stop + /r/ fusions in the dichotic task corresponds to the relative proportion of /l/ versus /r/ responses in the liquid identification task. This cross-task agreement did not occur in Experiment I since there were many /l/-substitution fusions. Thus the presence of trills overrides the /l/-substitution effect.

Experiment III: initial and final clusters

Most of the studies of phonological fusion to date have involved initial clusters. Day (1968) reported some pilot work in which fusion occurred at the end of words (GI*F*/GI*T*→GI*FT*) and in the middle (DEMON*S*ATION/DEMON*T*RATION→DEMON*ST*RATION) as well as at the beginning (*B*ACK/*L*ACK→*BL*ACK). Substantial fusion scores have been obtained for final clusters that are reversible: TASS/TACK→TASK and TACKS (Day, 1970*b*). At present no statement can be made concerning the relative effectiveness of fusion locus—beginning versus end of a syllable—since different types of clusters have been examined in each position (e.g. stop-liquid clusters in initial position and fricative-stop clusters in final position). Experiment IV examined the fusibility of clusters containing stops and liquids in both positions. However, phoneme order necessarily covaried with initial versus final syllable locus, since English permits only initial stop-liquid clusters (e.g. *PL*EAD) and final liquid-stop clusters (e.g. PEA*LED*).

Method

Two sets of stimuli appropriate for final liquid-stop fusions were synthesized. The PEED set contained PEED, PEEL, and PEER, and the possible fusions were PEELED and PEERED. The PAID set contained PAID, PAIL, and PAIR, and the possible fusions were PALED and PAIRED. The orthographic notation which shows items ending in -ED should not obscure the fact that all fusions end in a liquid-stop cluster (e.g. /pild/). Again all items and fusions were acceptable words. Within a set, all stimuli were identical except for the last 175 ms; /l/ and /r/ items differed only in their F3 transitions and accompanying F3 offglide.

A single-item liquid identification tape of 40 items was prepared: (2 stimulus sets) × (2 liquids/set) × (10 observations/item), with 3·0 s between trials. Subjects wrote down *L* or *R* to indicate the last phoneme of each item. A dichotic tape of 96 items was also prepared: (2 stimulus sets) × (2 liquids/set) × (3 lead times) × (2 channel arrangements) × (4 observations/pair), with 4·0 s between trials. Subjects wrote down "what they heard".

Tapes were also prepared to study initial clusters. Items were drawn from the stimuli used in Experiment I. Each stimulus set was represented by only three items: the stop and each of the endpoint liquids (A and E). A liquid identification tape was prepared consisting of 48 liquids presented one at a time: (4 sets of stimuli) × (2 liquid items/set) × (6 observations/item), with a 3·0 s interval between items. Subjects wrote down *L* or *R* to indicate the first sound of each item. A dichotic identification tape was prepared with 96 items: (4 sets of stimuli) × (2 stop-liquid pairs/set) × (3 lead times) × (2 channel arrangements) × (2 observations/pair), with 4·0 s between fusible pairs. Subjects wrote down "what they heard", as described in Experiment I.

All 12 subjects were tested on both types of clusters. Half began with the initial cluster tapes while the other half began with the final cluster tapes. Both dichotic fusion tasks were given, then both liquid identification tasks.

Results and discussion

Position had no effect on the identifiability of the liquid phonemes. Per cent correct scores were 89% for initial liquids and 90% for final liquids. However, position had a sizeable effect on the fusibility of dichotic items. Per cent fusion scores were 51% for initial stop-liquid clusters but only 13% for final liquid-stop pairs. Often the items designed to yield final liquid-stop fusions were reported as initial stop-liquid fusions: PEED/PEEL→*PL*EAD and PAID/PAIL→*PL*AYED. These position reversals were scored as fusions and comprised almost half of the fusions on final cluster trials. Thus only 8% of all trials yielded responses with clusters in final position.

One reason why the final cluster items did not yield sizeable fusion scores is based on the notion of context-conditioned variation. For example, although all items in the PEED set were prepared to have identical acoustic information in their first two phonemes (/pi/), the vowels in each are ordinarily produced in somewhat different ways and are perceived as having somewhat different vowel quality depending on which phoneme follows it, in this case /d/, /l/, or /r/. Furthermore, the vowels in natural productions of PEELED and PEERED differ considerably from the /i/ synthesized here. Final cluster items involved more allophonic variation in vowels from the inputs items to the possible fusion responses than did the initial cluster items. Initial cluster items certainly involved allophonic variation in the initial consonants, but their vowels are highly similar to those that occur in a fused response.

It is difficult to derive a set of principles which govern the fusibility of various cluster types partly because English phonology is unco-operative. That is, if one wants to determine the relative importance of phoneme order (e.g. /sk/ and /ks/) versus syllable position (beginning or end) in facilitating fusion for a given pair of phonemes, all four possible cases do not occur. For example, items such as TASK, TACKS, and STACK are permissible, but not TSACK. The final clusters in Experiment III also represent an incomplete domain in terms of phoneme order and syllable position. For example, in the PEED set, /d/ and /l/ can occur only in one of the four possible cases: /-ld/ as in PEELED. They cannot occur initially in either order /dl-/ or /ld-/, and the only way to get a final /-dl/ is to make several changes in other phonemes in order to get utterances such as PEDAL or BEETLE. Thus a complete description of the rules governing fusion of cluster types cannot be obtained directly given the phonological constraints of English. However, study of languages that allow other phonological patterns, such as the initial stop-fricative cluster in Slavic languages, will be helpful in arriving at such a description. There may be some basic principles of fusibility which are universal across languages and which may or may not be modified as a function of a speaker's language background.

Experiment IV: phonemically similar and dissimilar pairs

In most of the fusion studies reported to date, the input items were identical phoneme strings except for their initial phonemes. However, Day (1968) reported that items with few phonemes in common also yielded responses that combined information from the two inputs. For example, CATCH/LIFT sometimes yielded CLATCH or CLIFT. These blends were especially frequent when an initial stop-liquid cluster was possible, as in CATCH/LIFT. Blends were also frequent when the two items had an overlapping phoneme as in MAN/BAG (which yielded both MAG and BAN), or several overlapping phonemes as in TURTLE/SQUIRREL (which sometimes yielded SQUIRTLE). Experiment IV studied fusion of dissimilar pairs in a systematic fashion by comparing fusion scores for items such as PAY/LED and BED/LAY with those for BED/LED and PAY/LAY.

Method

The BED, PAY, and GO stimulus sets were used, as described in Experiment I. However, stops were paired with liquids both from their own sets and those from other sets. Thus there were 18 pairs in all, six similar within-set pairs (BED/LED, BED/RED, PAY/LAY, PAY/RAY, GO/LOW, GO/ROW) and 12 dissimilar between-set pairs (BED/LAY, BED/RAY, BED/LOW, BED/ROW, PAY/LED, PAY/RED, PAY/LOW, PAY/ROW, GO/LED, GO/RED, GO/LAY, GO/RAY). Since the individual items had achieved high identification scores in previous experiments no identification data were collected. Instead, only a dichotic tape of 108 pairs was prepared: (18 pairs) × (3 lead times) × (2 channel arrangements), with 5·0 s between trials. Sixteen subjects wrote down "what they heard".

Results and discussion

Fusions were less frequent than in Experiment I (43% overall). However, fusion occurred for all pairs, regardless of their status as within-set or between-set (40% versus 45%, respectively). Internal analyses are shown in Table II. Within-set items yielded more fusions

Table II Fusion scores for within-set pairs (e.g. PAY/LAY) and between-set pairs (PAY/LED) (Experiment IV)

Stimuli		Overall fusions	Breakdown of overall fusions: Stop + /l/ responses	Breakdown of overall fusions: Stop + /r/ responses
Within-set	Stop + /l/	47%	70%	30%
	Stop + /r/	33	38	62
Between-set	Stop + /l/	44	87	13
	Stop + /r/	50	35	65

for stop + /l/ items than stop + /r/ items (47 and 33%, respectively), although no significant differences were found between these two types of items in any of the other experiments.

Between-set fusions took interesting forms. Often, subjects wrote down two responses, a fusion and one of the single inputs, e.g. PAY/LED→PLAY and LED, or even PAY and PLED. The migration of the liquid phoneme into the stop item (PLAY) and the migration of the stop phoneme on to the beginning of the liquid item (PLED) occurred equally often, even though many were nonword fusions (PLED, BLAY, GRED, etc.). As long as a cluster occurred in a response, that trial was scored as a fusion.

Two responses per trial are frequent only when the input items are different either in terms of intonation, duration, or vowel quality as in natural speech productions (Day, 1968, 1970*a*), or in terms of contrasting phonemes as in Experiment IV. When the input items are identical acoustically except for their initial phonemes, one response is the norm. The average number of responses per trial was 1·14 in Experiment I and 1·50 here.

Experiment IV extends the notion that fusion is not heavily controlled by cues below the phonological level. The data suggest that fusion is insensitive to phonemic similarity; that is, items with no phonemes in common (e.g. PAY/LOW) fuse as readily as those that differ only in their initial phoneme (PAY/LAY). Apparently the input items need only be compatible at the phonological level; that is, the initial phonemes of each item must yield a permissible cluster when combined.

Experiment V: identification and discrimination

Given the items used in the present studies, when stop + /r/ items were presented, fusion responses often contained stop + /l/ clusters. These /l/ substitutions sometimes occurred as a large subset of fusion responses: 76% for Experiment I and 74% for initial clusters in Experiment III. However, they were reduced to 25% for plain and 17% for trilled items in Experiment II and to 38% for both within- and between-set items in Experiment IV. When less highly stylized stimuli are used, /l/ substitutions are virtually nonexistent (e.g. less than 4% in Day, 1968, Experiment I). Nevertheless, since they often accounted for such a high proportion of fusions in some of the present experiments, Experiment V was designed to study their origin.

Since stop + /l/ clusters accounted for almost 100% of fusions on trials of the general form PAY/LAY, but were reduced to 76% on PAY/RAY trials (Experiment I), it is possible that subjects could discriminate between the two types of trials, but happened to find PLAY the best label for both. Experiment V was designed to determine whether subjects can discriminate items such as PAY/RAY from PAY/LAY. Identification and discrimination data were obtained both for single liquid items and for fusible stop-liquid pairs.

Method

The same liquid identification tape as described in Experiment III (initial clusters) was used here. A liquid discrimination tape was also prepared consisting of triads such as LAY—RAY—LAY. The interval between members of a triad was 1·0 s while that between triads was 4·0 s. All four possible arrangements of liquids within triads were used: LRL, RLL, LRL, and LRR. There were 96 triads: (4 sets of stimuli) × (4 order arrangements/set) (6 observations/triad). A typical ABX procedure was used (Liberman, Harris, Hoffman & Griffith, 1957). The last item of the ABX triad was always identical to one of the first two items. Subjects wrote down *A* or *B* to indicate the identity of the last item.

The same dichotic identification tape as described in Experiment III (initial clusters) was used here. A dichotic discrimination tape was also prepared using a modified ABX design. Each triad was composed of three dichotic pairs such as PAY/LAY—PAY/RAY—PAY/LAY. The time pattern within and between triads was the same as that for the liquid discrimination tape. The stop in each of the three pairs was identical, as were the channel assignments and the lead time configurations. Only the liquid stimuli were varied within the triad, with the final liquid identical to one of the first two liquids. The tape consisted of 96 trials: (4 sets of stimuli) × (4 ABX arrangements/set) × (3 lead time configurations) × (2 channel arrangements/ABX triad). Subjects wrote down *A* or *B* to indicate the identity of the last pair.

All 12 listeners performed the tasks in the following order: dichotic identification, dichotic discrimination, liquid identification, liquid discrimination.

Results

The liquid items were easy to perceive when presented singly. In the liquid identification task, per cent correct scores for /r/-initial and /l/-initial items were 88 and 91%, respectively. In the liquid discrimination task, the overall per cent correct was 80%.

The dichotic pairs fused readily in the identification task, with 64% of all trials yielding fusion responses. The /l/ substitution fusions were also frequent, accounting for 88% of fusions on stop + /r/ trials. The dichotic discrimination task was quite difficult. Overall per cent correct was only 54%. Only two of the 12 subjects scored significantly above

chance. All subjects achieved lower scores on the dichotic discrimination task than on the single-item liquid discrimination task; in all cases the decrement was at least 20 percentage points. Furthermore, the distribution of scores for these two tasks did not overlap.

Discussion

The results of the two discrimination tasks show that subjects can discriminate liquid items presented singly, but they cannot discriminate the same items when they appear in fusible dichotic pairs. The F3 transition is the most potent cue for distinguishing /l/ from /r/, yet it is functionally irrelevant in the dichotic task when highly similar synthetic stimuli are used. Perhaps another cue, which is suggested below, overrides the F3 effectiveness.

The /l/-substitution effect for the present items appears to be based on comparable perceptions for items such as PAY/RAY and PAY/LAY since subjects cannot tell these items apart in the ABX discrimination task. Thus, the possibility that somewhat different percepts are given the same label is less likely. However, other discrimination procedures should be used to lend further support to this view, since different procedures vary in the extent to which they encourage labelling phenomena (Pisoni, 1971).

In an attempt to determine what circumstances are responsible for the /l/-substitution effect, natural speech utterances of PLAY and PRAY were spoken by one of the authors (J.E.C.) and converted into wide-band spectrogram form, as shown in Fig. 3. There are several interesting differences between the two items. PLAY has about 40 ms of low-amplitude steady-state resonance structure after /p/, just before the release of /l/. A comparable segment does not occur in PRAY: instead the /r/ formant transitions rise directly after the aspiration of /p/. Another aspect present in PLAY but not in PRAY is a discontinuity in both formant structure and intensity, which occurs between the 40 ms steady-state segment and the rise of the formants for the liquid. This break in the acoustic pattern may be an important cue for distinguishing initial stop + /l/ versus stop + /r/ clusters in English. Results of early tape-splicing experiments support this view.[5] When the stop consonant of a consonant–vowel syllable is spliced out of its original syllable and spliced on to another vowel, listeners sometimes perceive a stop + /l/ cluster before the vowel, but never a stop + /r/ cluster. The tape splicing procedure introduces discontinuities similar to those shown in Fig. 3.

The dichotic presentation mode could well introduce discontinuities which are responsible for the /l/-substitution effect. Since the beginning transition of the stop and liquid items are different, they remain spatially distinct; however, since the subsequent portions of the two items are identical in all aspects in the synthetic stimuli used here they are localized as a single item in the middle of the auditory field. This discontinuity from spatially distinct and different information to spatially fused and identical information may mimic the discontinuities that occur in natural stop + /l/ clusters.

The discontinuity interpretation is supported by results from another study (Cutting, 1975, Experiment II). Fusible dichotic pairs were presented in two modes. In the dichotic mode, items such as PAY and LAY were presented to opposite ears, while in the binaural mode, the two items were electrically mixed, then both presented to both ears. The binaural presentation procedure reduced fusion scores from 45% in the dichotic mode to 15% in the binaural mode. It also reduced /l/-substitution scores; per cent of overall fusions which were /l/ substitutions fell from 77 to 43%. Thus the spatial discontinuities introduced by the dichotic presentation technique may facilitate both fusions and /l/ substitutions.

[5] Leigh Lisker, personal communication.

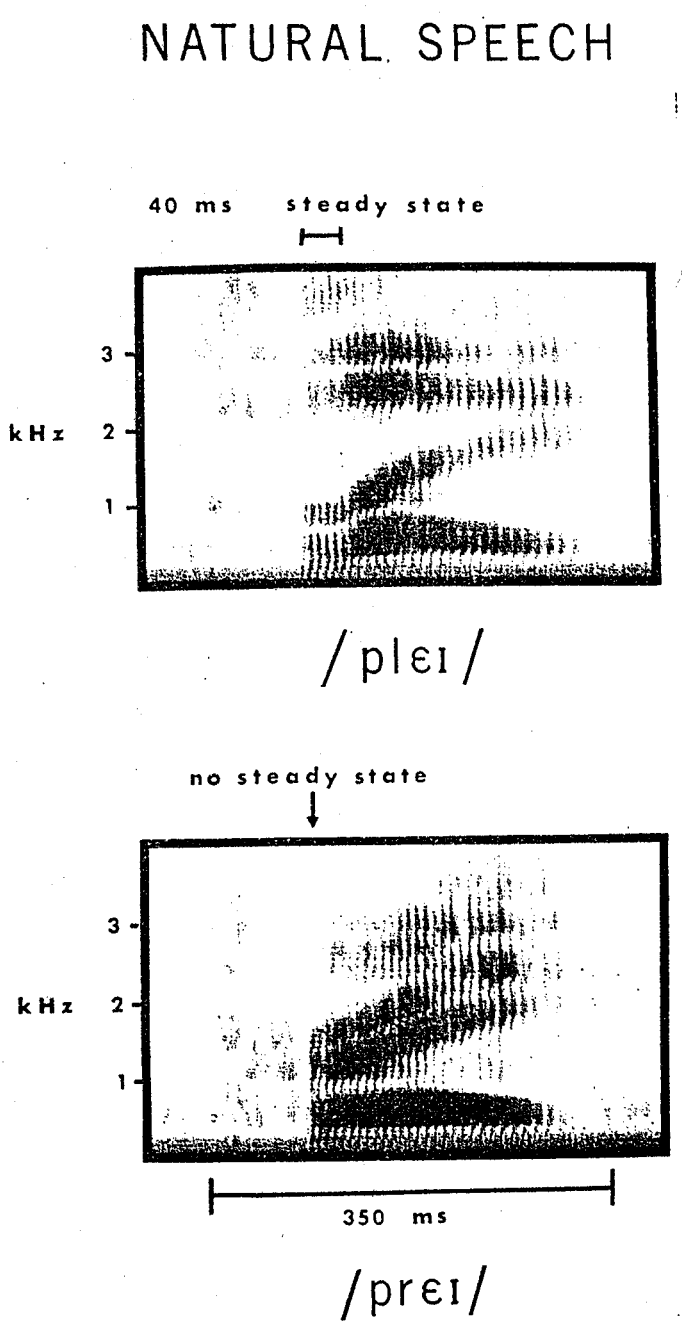

Figure 3 Wide-band spectrograms of the natural speech utterances PLAY and PRAY (discussed in Experiment V).

In any event, it should be remembered that the /l/-substitution effect occurs only for highly stylized items, and virtually disappears when less stylized items are used.

Individual differences

A substantial number of fusions occurred in all experiments. The only marked decrease in fusions occurred for final liquid-stop clusters. There were no significant differences in fusion scores for plain versus trilled liquids, nor for phonemically similar versus dissimilar pairs. However, it should be noted that overall fusion scores were not constant for the baseline items common to all experiments. Baseline dichotic pairs for Experiment I–V, respectively, were those that involved endpoint liquids, plain liquids, initial clusters, within-set items, and all dichotic identification items. Fusion scores for these items, averaged over the subjects in each experiment were: 60, 43, 51, 40 and 64%. One possible explanation for this variability in fusion scores is that presence of an experimental manipulation (such as trilled versus plain liquids) can decrease overall fusion scores even though there are no differences between the experimental conditions. A more plausible explanation rests on the concept and data of individual differences.

The analyses presented throughout the paper have been concerned primarily with items and tasks, collapsed over individual subjects. A common assumption is that individual differences in task scores are normally distributed, so that individuals differ in degree, not kind. In the present paper, the 68 individuals tested in all experiments were compared in terms of their fusion scores for baseline items. The resulting distribution revealed two groups: those whose scores clustered around 39% (with a range from 2 to 60%) and those who clustered around 88% (with a range from 70 to 100%).

This bimodal distribution of fusion scores was first observed by Day (1970*a*). Day found that these two groups of individuals show markedly different patterns of performance on another task using the same fusible dichotic items. When asked to determine which phoneme led on each trial, the high fusers reported hearing the stop consonant first even when the liquid led by a considerable interval. These individuals are termed "language-bound" (Day, 1970*a*) since they are heavily influenced by language rules. Even though the temporal-order judgment task requires a decision only concerning which phoneme led, the language-bound individuals still obey constraints set at the phonological level. The low fusers had no difficulty with the temporal-order judgment task: they gave correct responses both when the stop or liquid led. These individuals are "language-optional" since they can use language rules or set them aside according to task demands. Differences between language-bound and language-optional individuals are evident in a wide variety of perception and memory tasks (Day, 1973).

For the purposes of the present paper, it was important to determine whether the two types of individuals were represented in approximately similar proportions in all experiments. If there were more subjects of one type in a given experiment, overall fusions scores could be based on the individuals sampled rather than the types of items or tasks used. The per cent of high fusers (baseline scores clustered around 88%) for Experiments I–V, respectively, were: 31, 8, 25, 13 and 42%. These figures should be positively related to the overall fusion scores for the baseline items, since the presence of many high fusers would boost overall scores (and *vice versa*). However, the correlation was virtually perfect ($r = 0{\cdot}96$, $P < 0{\cdot}01$). Thus individual differences were totally responsible for the variation in fusion scores for the baseline items.

These individual differences results present a methodological lesson: language perception experiments should use within-subject comparisons. In the present studies, if

separate groups of subjects had been used for each condition, and the proportion of language-bound to language-optional subjects had varied widely across conditions without our knowledge, then we would be unable to make meaningful statements about the effects of the experimental variables.

This work was supported in part by NICHD Grant HD-01994 to the Haskins Laboratories. The first author was at Yale University and Haskins when the research was conducted.

Special thanks go to T. Tucker for his continued interest in seeing this work completed.

References

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

Cutting, J. E. (1975). Aspects of Phonological fusion. *Journal of Experimental Psychology: Human Perception & Performance* **104**, 95–112.

Day, R. S. (1968). Fusion in dichotic listening. Doctoral dissertation, Stanford University (Psychology). *Dissertation Abstracts* (1969), **29**, 2649B, (University Microfilms No. 69–211, Ann Arbor, Michigan.)

Day, R. S. (1970*a*). Temporal-order judgments in speech: are individuals language-bound or stimulus-bound? *Haskins Laboratories Status Report on Speech Research*, **SR-21/22**, 71–87.

Day, R. S. (1970*b*). Temporal-order perception of a reversible phoneme cluster. *Journal of the Acoustical Society of America*, **48**, 95(A). Also in *Haskins Laboratories Status Report on Speech Research*, **SR-24**, 47–56.

Day, R. S. (1973). Individual differences in cognition. Paper presented at the 13th meeting of the Psychonomic Society, St. Louis, November, 1973.

Liberman, A. M., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme categories. *Journal of Experimental Psychology* **54**, 358–68.

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

Pisoni, D. B. (1971). On the nature of categorical perception of speech sounds. Doctoral dissertation, University of Michigan (Psycholinguistics). (Issued as Supplement to *Haskins Laboratories Status Report on Speech Research.*)