The Lag Effect in Dichotic Speech Perception

Emily F. Kirstein

INTRODUCTION

An important factor in dichotic competition is the temporal alignment of syllable onsets at the two ears. It is well known that if different syllables are presented simultaneously to opposite ears, syllables at the right ear are more accurately identified than those at the left (Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970). Recently, it was discovered that the size of the right-ear effect can be increased by delaying syllable onsets at the right ear 5 to 120 msec behind the left and, conversely, that the ear advantage can be reduced or even reversed (giving a left-ear superiority) by causing the left-ear syllable to be delayed behind the right (Lowe, Cullen, Thompson, Berlin, Kirkpatrick, and Ryan, 1970; Studdert-Kennedy, Shankweiler, and Schulman, 1970). That is, in general, lagging syllables at either ear have an advantage over leading syllables; this "lag effect" is seen superimposed on the right-ear effect in dichotic experiments.

In control experiments syllables differing in time of arrival by 5 to 120 msec were mixed electronically and the resulting signal was delivered to one ear (monotic) or to both ears (diotic). For these conditions leading syllables were more accurately identified than lagging syllables (Lowe et al., 1970; Studdert-Kennedy et al., 1970; Kirstein, 1971; Porter, 1971a). Studdert-Kennedy et al. (1970) attributed the diotic and monotonic lead effect to peripheral masking. They considered the dichotic lag effect to be a higher-level phenomenon involving competition for perceptual processing. They proposed that the lagging syllable is more intelligible than the leading syllable because it interrupts the phonetic analysis of the leading syllable and captures the speech processor.

+University of Michigan, Dearborn Campus, Department of Psychology.

Acknowledgment: This paper reports a portion of the research conducted at Haskins Laboratories for my doctoral dissertation in psychology at the University of Connecticut (Kirstein, 1971), under the supervision of Dr. Alvin Liberman. I wrote this paper while a postdoctoral trainee at the MIT Speech Communication Group, supported by a National Institutes of Health training grant to the MIT Research Laboratory of Electronics. I would particularly like to thank Dr. Ignatius Mattingly, Dr. Michael Studdert-Kennedy, and Dr. Kenneth Stevens for their help and encouragement.

All the experiments referred to above used as stimuli the stop consonant-vowel (CV) syllables /ba/, /da/, /ga/, /pa/, /ta/, /ka/, where consonants differed but vowels were shared. Subsequently, other stimuli have been used to determine precisely the conditions required to elicit the lag effect. Apparently, as long as there are stop consonants in CV syllables contrasting between ears, the lag effect will persist, despite substantial variations in the acoustics of interaural competition. For example, the lag effect has been demonstrated where vowels as well as consonants vary between ears, as in /ba-/ /ge/ (Kirstein, 1971), where the fundamental frequency of the syllables varies between ears (Halwes, 1969), and where the duration of the competing syllables has been shortened from the usual 300 msec to only 75 msec (Porter, 1971a). The minor influence of such acoustic variations strengthens the view that a critical condition for producing the lag effect is the "perceptual class" of the stimuli.

It is not certain whether the lag effect is peculiar to speech or whether it is a more general phenomenon of auditory perception. Darwin (1971) asserted that the lag effect was related to the perception of rapidly changing acoustic signals (transitions) whether in speech or nonspeech tasks. He supported this claim by demonstrating a lag effect for perception of pitch changes (rising, falling, level) in the initial 50 msec of a 150 msec steady-state vowel. However, Porter (1971b) found no evidence of a lag advantage in perception of formant transitions isolated from the speech signal itself. Also, in Darwin's study, the use of a vowel as a carrier of the pitch transition makes it difficult to classify the sounds unambiguously as nonspeech. Thus, at present there is no strong evidence to refute the hypothesis of Studdert-Kennedy et al. (1970) that the lag effect is a speech perception phenomenon.

Not all classes of speech sounds are equally effective for eliciting the lag effect. Porter (1971b) compared stop consonants (/b/, /d/, /g/) with sonorant consonants (/l/, /w/, /y/). He found that some subjects had a lag effect for both stops and sonorants while other subjects had the lag effect for stops only. Porter, Shankweiler, and Liberman (1969) presented steady-state vowels dichotically with delays between ears, and they found a slight advantage for leading over lagging vowels. However, Kirstein (1971) found a preference for lagging vowels if the vowels were embedded in CV syllables. She also found a lag effect for isolated steady-state vowels among subjects who had previously taken a dichotic test involving stop consonants. The finding that the lag effect is an extremely robust effect for stops, less robust for sonorants, and marginal for vowels supports the view that the effect is related to special decoding processes in speech perception. The term "encoding" has been used by Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967) to refer to the fact that the acoustic cues for perception of a particular phoneme may be greatly affected by the nature of the adjacent phonemes. Liberman et al. (1969) proposed that highly encoded phonemes like stop consonants require special decoding to arrive at phonetic identification, while unencoded phonemes like the isolated steady-state vowels could optionally be identified through purely auditory perception modes. (Sonorant consonants are more highly encoded than vowels, but less encoded than stops.) The finding that the lag effect occurs for vowels under some circumstances but not under others suggests that the lag effect is related to the perceptual mode (speech or nonspeech) rather than to some acoustic feature of the stimulus.

The present research is concerned with the methodology of lag effect experiments, and specifically with the role of attention. In dichotic experiments
listeners are generally required to attend simultaneously to both ears. The
question examined here is whether leading syllables might be as accurately iden-
tified as lagging syllables are if attention were concentrated on leads or lags
only, rather than divided between them. It had been claimed by Inglis (1965) and
by Treisman and Geffen (1968) that the ear advantage in dichotic tasks can be
attributed to systematic biases on the part of subjects in their order of reporting
the two ears or in the distribution of attention between ears. To control for
these factors in the study of ear effects Kirstein and Shankweiler (1969)
introduced the procedure of having listeners concentrate on one ear at a time,
reporting only the syllables at the "attended" ear. They found that for dichoti-
cally presented stop consonants report was more accurate under right-ear atten-
tion than under left-ear attention. They concluded that neither response bias
nor the distribution of attention could explain the ear asymmetry. In dichotic
experiments where syllables are temporally offset between ears, the distribution
of attention and response ordering might also affect the pattern of identifica-
tion errors. It seemed desirable, therefore, to determine whether the lag advant-
age would occur with attention directed selectively toward lagging or leading
syllables.

METHOD

A dichotic tape consisting of pairs of CV syllables was constructed. Sylla-
bles within a pair always contrasted in the initial consonant (/b/, /d/, or /g/)
but shared the same vowel. Within a pair, one syllable was always delayed rela-
tive to the other by 10, 30, 50, 70, or 90 msec. This tape was presented to sub-
jects under three different task conditions. In the two-response task the listen-
ers were instructed to report both consonants on each trial and to indicate which
was the clearer. This is essentially the method of Studdert-Kennedy et al.
(1970). In the ear monitoring task subjects were instructed to concentrate their
attention on a particular ear and to report only the consonants at the "attended"
ear. In the temporal order task listeners were instructed to attend to the order
of arrival of the consonants within a pair and report only the leading stops or
only the lagging stops according to instructions.

Stimuli. Nine consonant-vowel syllables were synthesized on the Haskins
Laboratories'parallel resonance speech synthesizer. These were /ba/, /da/, /ga/
/ba/, /da/, /ge/, /bo/, /do/, /go/. The duration of each syllable was 350 msec.
Syllables beginning with /d/ or /g/ started with a 10 msec noise burst, followed
by appropriate formant transitions. No burst was needed with /b/. In all cases
the formant transitions were completed and the steady-state vowel parameters
reached within 70 msec.

The intelligibility of the syllables was assessed by asking four listeners
unfamiliar with synthetic speech to identify the consonants in the 180-trial
randomization. Ninety-five percent correct identifications were obtained, an
adequate level of intelligibility for the dichotic tests.

Waveforms of the stimuli were stored on a computer disc file using the Pulse
Code Modulation System (Cooper and Mattingly, 1969). This computer system also
controlled the alignment of syllable onsets as the syllables were recorded in
pairs in a specified order onto the two-channel dichotic test tape.

In the design of the dichotic tape, care was taken to counterbalance to pre-
vent confounding of ear effects, lag effects, and tape channel imbalance. Each

83
ear received the same number of lag and lead trials, and the same permutations of the syllables. From the nine CV syllables there are 18 possible permutations of two syllables where only the consonants differ. A 180-trial randomization was assembled in which each of the 18 permutations occurred twice at each of the five delay intervals (10, 30, 50, 70, and 90 msec), once with channel-1 delay and once with channel-2 delay. All conditions of offset were randomly ordered on the tape. There was a 6-sec pause between pairs and a 10-sec pause after ten pairs.

Procedure. The dichotic tape was played from a General Radio stereo tape deck into a special amplifier built by D. Zeichner of Haskins Laboratories for group dichotic experiments. As many as six subjects could be tested at one time. The subjects listened to the tape over Grason-Stadler stereo headphones. The tape was presented at a comfortable listening level, and the output intensity from the two channels was equated to within 1 db with the aid of calibration signals on the tape.

As an added control for channel effects, the dichotic tape was always presented twice within a test session with the headphone orientation physically reversed on the second presentation.

Two-response task. The subjects were told that they would receive two different syllables on each trial, one syllable to each ear, and that the syllables would differ between ears in the consonant (/b/, /d/, or /g/), never in the vowel. The instructions were to report both of the consonants, giving two different responses and guessing if necessary. An added aspect of this task was the clarity judgment. The responses were to be ordered on the answer sheet so that the clearer consonant was written in the first column for each pair and the less clear consonant in the second column. If one of the two responses was a guess, it was to be written in the second column.

Each subject had 360 trials in a single 1-hour test session.

Ear monitoring task. The subjects were told that they would receive two different syllables on each trial, one syllable at each ear, and that the syllables would differ between ears in the consonant, (/b/, /d/, or /g/), never in the vowel. The instructions were to attend to a particular ear designated by the experimenter and to write down on each trial the consonant arriving at the "attended" ear. The subjects were required to respond on every trial even if the response was a guess. Each subject had 360 trials under left-ear attention and 360 trials under right-ear attention.

Each subject had two 1-hour (360-trial) test sessions, with each session subdivided into four 90-trial blocks. At the start of a block the subject was told whether to report the right ear or the left, and this instruction was in effect for the entire block. Within a session the order of the blocks was Right-Left-Left-Right or Left-Right-Right-Left; a subject was randomly assigned to one of these orders for his first session and was automatically assigned to the other order for the second session.

Temporal order task. The subjects were told that they would receive two different syllables on each trial, one at each ear; that the syllables would differ between ears in the consonant (/b/, /d/, or /g/); and that the syllables would also differ slightly in onset time, with the leading syllable randomly at
the right or left ear. The instructions were to attend to the order of arrival of the two syllables and to report either the lagging or leading syllable, as specified by the experimenter. The subjects were required to respond on every trial even if the response was a guess. Each subject had 360 trials reporting the leading syllables and 360 trials reporting the lagging syllables.

Each subject had two 1-hour (360-trial) test sessions, with each session subdivided into four 90-trial blocks. At the start of a block the subject was told whether to report the leading syllables or the lagging ones, and this instruction was in effect for the entire block. Within a session the blocks were arranged Lags-Leads-Leads-Lags or Leads-Lags-Lags-Leads; a subject was randomly assigned to one of these orders for his first session and was automatically assigned to the other order for the second session.

**Subjects.** The ear monitoring and temporal order tasks were originally studied together as part of a single experiment. A group of 12 subjects took both tasks, half taking the temporal order first, and half the ear monitoring task. Later, an additional group of 10 subjects was run on the ear monitoring task only, making a total of 22 ear monitoring subjects. (The data analysis revealed no systematic differences between the original group of 12 and the added group on the ear monitoring task. For purposes of analysis sometimes only the original 12 subjects are considered while in other cases data are presented for the entire group of 22.)

The two-response task was performed by 24 subjects, none of whom had previously participated in dichotic experiments.

The subjects, all Introductory Psychology students at the University of Connecticut, received course credit for their participation. They were native speakers of English, were self-classified "right-handers," and had (to their knowledge) normal hearing in both ears.

**RESULTS**

**Two-response task.** The percent correct responses on the two-response task is shown in Figure 1, presented according to the various stimulus conditions: lag vs. lead, right vs. left ear, and length of interaural delay. The data replicate the earlier findings: lagging syllables were more accurately identified than leading syllables, and the right ear was more accurate than the left ear. The steps that were lagging and at the right ear were most often correctly identified, next in accuracy were left-ear lags, then right-ear leads, and finally left-ear leads. The only exception to this ordering was at 10 msec, where right-ear leads were slightly better than left-ear lags. An analysis of variance on the two response data is summarized in Table 1. Both the right-ear effect and lag effect were highly significant. Also, averaging over lags and leads and over the left and right ears, there was a statistically significant rise in the number of correct identifications for longer interaural delay intervals. Figure 1 shows that both the lag advantage and ear advantage vary in magnitude depending on the length of interaural delay; however, in the analysis of variance only the interaction of lag effect with delay was significant.

The clarity-judgment instructions were apparently being followed, since errors occurred primarily on second responses: 5 percent of responses in the
Figure 1: Percent correct responses on two-response task (N=24). Each point is based on 864 trials, 36 trials per subject.
TABLE 1: Analysis of variance summary based on the number of correct responses in each task.

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Two-Responses (N=24)</th>
<th>Ear Monitoring (N=22)</th>
<th>Temporal Order (N=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F</td>
<td>df</td>
<td>p</td>
</tr>
<tr>
<td>Right Ear Effect</td>
<td>21.8</td>
<td>1,23</td>
<td>*</td>
</tr>
<tr>
<td>Lag Effect</td>
<td>28.6</td>
<td>1,23</td>
<td>*</td>
</tr>
<tr>
<td>Length of Delay</td>
<td>6.78</td>
<td>4,92</td>
<td>*</td>
</tr>
<tr>
<td>Ear Effect x Delay</td>
<td>1.66</td>
<td>4,92</td>
<td>NS</td>
</tr>
<tr>
<td>Lag Effect x Delay</td>
<td>6.90</td>
<td>4,92</td>
<td>*</td>
</tr>
<tr>
<td>Ear x Lag Effect</td>
<td>2.01</td>
<td>1,23</td>
<td>NS</td>
</tr>
<tr>
<td>Ear x Lag x Delay</td>
<td>.57</td>
<td>4,92</td>
<td>NS</td>
</tr>
</tbody>
</table>

* indicates effect is significant with p<.001 or better.
first column were errors, but 28 percent of second responses were errors. Errors on second responses decreased with longer interaural delay intervals, but the first-response error rate was independent of delay interval. An analysis of first responses showed that lagging stops were judged to be clearer than leading and the right ear was judged clearer than the left. (The first-response results are shown in Figure 7, which is discussed later. Of the 24 subjects, 23 favored lags over leads in first responses, and 18 favored the right ear over the left. (Only the direction of preference was tabulated, not the magnitude of the effect.)

**Ear monitoring and temporal order tasks.** The results of the ear monitoring task are shown in Figure 2 and the temporal order results in Figure 3. Both tasks gave essentially the same results. On the ear monitoring task, listeners were more accurate on right-ear attention than on left-ear attention, and they were more accurate when the attended ear received a lagging syllable than when it received a leading syllable. On the temporal order task, listeners made fewer errors under "report lags" instructions than under "report leads" instructions, and report of either lags or leads was more accurate from the right ear than from the left. The analysis of variance for the selective listening tasks is summarized in Table 1. For both tasks the lag effect and right-ear effect were significant, and both the ear effect and the lag effect showed significant variations in magnitude with the length of interaural delay. Accuracy of report by ear or by temporal order improved significantly with longer delay intervals.

Since 12 subjects took both the temporal order and ear monitoring tasks, it was possible to make a comparison of the consistency of individual ear effects and lag effects across the two tasks. A measure of each subject's lag effect was obtained for the temporal order task by subtracting the number of correct responses under "report leads" from the number of correct responses under "report lags." For the ear monitoring tasks, a measure of the lag effect was obtained by summing the number of correct right-ear and left-ear responses when the attended ear received lagging syllables and subtracting the number of correct responses when the attended ear received leading syllables. Similarly, a measure of the ear effect was determined for each subject on each task. The reliability of individual differences across tasks was assessed by calculating the Pearson product-moment correlation coefficient. The individual lag effect measures gave a correlation coefficient of .85 across tasks, and the ear effect measures gave a correlation coefficient of .95, both of which are highly significant, with p < .001. These results indicate that individual lag effect scores and individual ear effect scores are reliable, even when the measures are obtained on different tasks.

The types of errors made on the ear monitoring and temporal order tasks fell into two categories. The response could be identical to the unattended syllable, in which case it is termed an "intrusion" error; or the response could differ from both the attended and unattended syllables, in which case it is termed a "nonintrusion" error. Figure 4 gives a breakdown of all responses as correct responses, intrusion errors, or nonintrusion errors. Intrusions were the primary source of errors under both attention conditions. That is, listeners had difficulty in discriminating between the attended and the unattended syllables. The ability to select the attended syllable improved with longer delay, and Figure 4 shows that this improvement is entirely due to a reduction in the number of intrusions of unattended syllables. Somewhat surprisingly, nonintrusion errors increased slightly with longer interaural delay, whereas they might have been
Figure 2: Percent correct responses on ear monitoring task (N=22). Each point is based on 792 trials, 36 trials per subject.
Figure 3: Percent correct responses on temporal order task (N=12). Each point is based on 432 trials, 36 trials per subject.
Figure 4: Percent correct responses, intrusion errors, and nonintrusion errors on selective listening tasks (N=12).
expected to decline. For the ear monitoring task, the increase in nonintrusion errors with longer interaural delay proved to be statistically significant [Friedman two-way analysis of variance \( \chi^2 = 14.8 \) p < .01 (Siegel, 1956)]. Overall, reporting a particular ear was more accurate than report by order of arrival (\( F = 17.8 \) df = 1,11 p < .005).

**Lag effect and ear effect as functions of delay interval.** Thus far it has been demonstrated that there is an advantage for lagging syllables and right-ear syllables in all three tasks. The present section examines how the magnitude of lag effect and ear effect varied with length of delay.

The lag effect was treated independently of the right-ear effect by computing the mean percent correct on right- and left-ear lags and subtracting from this the mean percent correct on right- and left-ear leads. Figure 5 displays these lag effect scores at each delay for the three tasks. For the two-response task separate plots are shown for the lag effect in clarity judgments (first responses) and for the lag effect in intelligibility (both responses). The same trend was observed in all tasks: the advantage for lags over leads progressively increased with longer interaural delay up to 50 msec; with still longer delays the lag effect began to diminish. The lag effect showed a maximum at 50 msec for all three tasks, so this location must be considered a reliable finding, at least for this particular set of stimuli.

The finding of a lag effect peak in the 50-70 msec delay range is also in agreement with other observations (Berlin, Lowe-Bell, Cullen, Thompson, and Loovis, 1973; Studdert-Kennedy et al., 1970; Porter, 1971b).

A right-ear effect score at each delay was computed similarly by subtracting the mean percent correct on left-ear leads and lags from the mean percent correct on right-ear leads and lags. These scores are plotted in Figure 6. The right-ear advantage was greatest at short interaural delay intervals and declined with longer delays. Again all three tasks showed the same trend, although for the two-response task the change in ear effect with delay was not statistically significant. From these results it can be inferred that the right-ear advantage would have been maximal with simultaneous onsets.

While the pattern of results was basically the same in all tasks, the magnitude of ear advantage and lag advantage did vary considerably among the tasks, as can be seen in Figures 5 and 6. Statistical factors were considered first in trying to account for these magnitude differences. It can be shown that the maximum possible lag effect or ear effect which can be obtained in the two-response task is only 50 percent whereas in the selective listening and clarity judgment tasks a 100 percent lag effect or ear effect could be obtained.\(^1\) A correction

\(^1\)The 50 percent ceiling on the lag effect in the two-response task derives from the fact that the subject must give two different responses on each trial from a set of only three possible responses. Suppose that the listener always heard correctly the lagging consonants but not the leading consonants. His guess for the second response would nevertheless be correct for the leading consonant for half the trials. Thus, the lag effect would have been 100 percent considering only first responses but is automatically reduced by half when both responses are considered. The same argument applies to the ear effect. Using a larger response set would have given a higher ceiling on the magnitude of the effects obtained in the two-response task.
Figure 5: Magnitude of lag effect as a function of interaural delay interval. Comparison of clarity judgments, ear monitoring, temporal order, and two-response tasks.
Figure 6: Magnitude of right-ear effect as a function of interaural delay interval. Comparison of clarity judgments, ear monitoring, temporal order, and two-response tasks.
must be applied to the two-response data (multiplying all ear effect and lag effect scores by 2) before comparing the two-response with the other tasks. However, even after this correction has been applied, the magnitude of the effects obtained in clarity judgments is still greater than in any of the other tasks. A possible explanation for this result is that clarity judgments are more sensitive than the other tasks to the effects being studied. The clarity judgment task required listeners to compare the competing syllables qualitatively, while the two-response and selective listening tasks asked subjects to identify the syllables.

The idea that clarity judgments are more sensitive than identification tasks receives support from a more detailed comparison of clarity judgments with responses on the ear monitoring task. Figure 7 compares these two tasks directly, plotting on the same chart the percent "clearer" judgments for the various temporal offset conditions against the percent of trials on which these same syllables were correctly identified under selective listening. If the curves for the two tasks were to coincide, this would indicate that the subjects could correctly identify in ear monitoring only those stops which were independently judged to be the relatively clearer within the pair. Divergence between the ear monitoring and clarity judgment curves gives the percentage of trials on which the "less clear" consonant could be identified when attention was concentrated on that sound. It is evident that very frequently the "less clear" sound could in fact be identified. It is interesting also that at short delays the responses given were not greatly affected by the specific task instructions, while longer delays produced greater divergence between tasks. The greater sensitivity of the clarity judgment over the identification task can be seen in the fact that as we move from the most favored condition (right-ear lags) to the least favored conditions (left-ear leads) the change between conditions is greater for clarity judgments than for identification.

**DISCUSSION**

All three tasks gave essentially the same patterns of identification errors. The effects observed in all tasks were the lag effect, the right-ear effect, the variation in the magnitude of lag effect and ear effect with delay, the improved performance with longer delay, and the susceptibility to intrusion errors in the selective listening tasks. The consistency in error pattern across tasks is convincing evidence that there are genuine variations in intelligibility of dichotic syllables depending on ear of arrival, order of arrival, and temporal offset. These effects are clearly not under the listeners' voluntary control, and they cannot be explained in terms of attentional strategies or response order.

It is often said that dichotic presentation causes errors because of perceptual competition between stimuli, that is to say, because the perceptual system is unable to process both ears simultaneously. The term "perceptual competition" is usually interpreted to mean that verbal stimuli presented simultaneously to opposite ears must compete for entry to language processing areas. The assumption that only one ear at a time can have access to language areas of the brain underlies much theorizing about the ear effect and lag effect. For example, Kimura (1961, 1967) attributed the right-ear advantage for dichotically presented verbal stimuli to competition between inputs for access to language processing areas of the left hemisphere; the right ear was thought to win the competition because of the greater strength of its neuroanatomical connections to the left hemisphere. The explanation of the lag effect offered by Studdert-Kennedy et al.
Figure 7: Clarity versus identifiability of dichotic stop consonants. Clarity judgments and ear monitoring scores are compared for all delay conditions.
(1970) also assumes that only one ear at a time can be admitted to the speech processor. They proposed that the lag effect occurs because the leading syllable is ejected from the processor when the lagging syllable arrives on the opposite channel. It will be argued in the discussion that follows that recent data do not support the idea that the ears must compete for entry to the speech processor. An attempt will be made to redefine the notion of dichotic competition and to account for the ear effect and lag effect in light of that definition.

The view that dichotic stimuli compete for perceptual processing was elaborated by Broadbent (1958) in his "filter theory" of selective attention. According to that theory, the flow of sensory data to central processing areas is regulated by a filtering mechanism which blocks peripherally all irrelevant sensory channels. Broadbent treated the two ears as separate sensory channels. He considered that in dichotic listening only one ear could be processed at a time and that the system would have to "switch channels" in order to process both ears. Peripheral blocking of one ear and channel switching were assumed to be under a subject's voluntary control.

A motivation for the filter theory was to explain the finding of Cherry (1953) that when different continuous messages were presented to opposite ears listeners could easily repeat back (shadow) the message at one ear and ignore the other. Moreover, they showed no retention of the unattended message. Broadbent attributed the shadowing results to the filtering out of the unattended message so that only the attended message could have access to linguistic processing. However, in subsequent research using the shadowing technique the unattended ear was not fully suppressed. For example, Treisman (1960) found that words in the unattended messages would occasionally be repeated by the shadower if those words were semantically probable within the context of the attended message. Also, if the experimenter switched messages between ears, subjects would sometimes also switch ears unconsciously, maintaining the continuity of the shadowed message. Such intrusions from the unattended ear relating to the semantics of the messages indicate that the unattended message must have been analyzed linguistically, and subjects' failure to retain the content of the unattended message in these tasks must be attributed to memory rather than perception.

Ear monitoring experiments with dichotic syllables prove even more strikingly that listeners cannot voluntarily select a particular ear for perceptual analysis and filter out the other. When stop-vowel syllables are presented dichotically with simultaneous onsets, subjects frequently confuse ears when instructed to report a particular ear (Kirstein and Shankweiler, 1969; Halwes, 1969) and when instructed to report both syllables by ear (Gerber and Goldman, 1971). The present results expand these earlier findings by showing that confusions between ears persist even when syllable onsets are not precisely simultaneous. Accuracy of monitoring is, however, related to interaural delay interval, and the fact that intrusion frequency depends on the timing relation between ears is strong evidence that listeners cannot voluntarily exclude a particular ear from processing. The results suggest, rather, that selection from a particular ear depends on the physical relation between the dichotic stimuli. The ear monitoring results contradict two assumptions of the filter theory: first, that dichotic inputs are always strongly "tagged" by ear of origin, and second, that listeners can voluntarily turn off an ear in dichotic tasks.

If the unattended ear cannot be voluntarily inhibited peripherally, then how are we to account for the ease of selectively shadowing a particular ear for
continuous dichotic messages? Moray (1969) considered this matter and concluded that the mechanism of selection of a particular ear is the same in dichotic listening as in ordinary binaural situations where many messages are arriving simultaneously to both ears—the so-called "cocktail party" effect. A particular message can be selected for attention providing there are sufficient physical cues to establish a distinct spatial origin for the message. In everyday listening situations the two ears do not behave as independent channels. Ordinarily the same signals arrive at both ears, perhaps with slight differences in time, phase, or intensity between ears; these interaural differences provide the physical cues for spatial localization of the sound. Thus, normally inputs at the two ears are integrated to yield a unitary percept and are compared to locate the sound source (Cherry, 1961). Moray proposed that dichotic inputs are handled in essentially the same manner.

If we assume, following Moray, that in dichotic listening the auditory system compares stimuli from the two ears and locates the source based on differences between ears, we can then readily understand why reporting stop consonants from a particular ear is difficult, while shadowing continuous messages is much easier. For ongoing messages the acoustic signals at opposite ears would be generally quite distinct at any point in time, so that by comparing the two inputs the auditory system could establish two distinct sound sources and a subject could shadow the message emanating from a particular location. In stop-vowel syllables the acoustic information distinguishing one stop from another is contained within the first 70 msec or less of any syllable, and the vowels are acoustically identical at the two ears. Thus, if selection between ears presupposes a clear acoustic distinction between simultaneous dichotic signals, it is understandable that selection would be faulty for CV syllables. The role of acoustic similarity between ears in selective listening for dichotic stop consonants was demonstrated convincingly in an experiment by Halwes (1969). He presented CV syllables dichotically with syllable onsets precisely aligned for simultaneity and compared accuracy of selective report for syllable pairs which shared fundamental frequency at the two ears or which differed in fundamental frequency. When fundamental frequency was shared, listeners were unable to distinguish the attended from the unattended ear; performance was significantly improved when fundamental frequencies varied. The present study shows that another physical dimension, interaural asynchrony of onsets, is also important in selection.

Confusion between ears for dichotic stops was attributed by Halwes to an acoustic fusion effect. Halwes claimed that the perceived localization of a syllable presented to one ear was shifted toward the midline when an acoustically similar syllable was delivered at the opposite ear. Often the listeners heard a single syllable localized at the midline or diffusely rather than at a particular ear. Earlier, a similar phenomenon was described by Broadbent and Ladehoff (1957). They presented the first formant of a synthetic vowel to one ear and the second formant to the opposite ear; these fused perceptually into a single vowel.

While listening to the tapes for the present experiments, I observed frequent fusion for short delay (10–30 msec) trials. For these, generally only one of the two stops could be heard, and it often could not be definitely assigned to either ear. For longer delays, two stimuli could usually be detected although often the identity of one was still unclear. These observations accord with the experimental data. On the two-response task, for example, the subjects could identify one stop from each pair, regardless of delay, but correct identification
of both stops was facilitated by longer offsets. For ear monitoring, longer delays reduced intrusions; this effect can be explained by the assumption that selection can occur only under conditions where both syllables can be detected.

A somewhat surprising result in the ear monitoring task was that while intrusion errors decreased with longer interaural delay, nonintrusion errors increased. The increase in nonintrusion errors with longer delays may result from the fact that at the shortest delays one nearly always hears one of the two syllables clearly, although that might be the unattended syllable. At longer delays listeners would be better at discriminating the attended from the unattended syllable, but they would often be unable to identify the attended syllable. That is, the increase in nonintrusion errors with longer delay can be related to the temporal alignment condition where the attended syllable can be selected but not identified.

Dichotic fusion phenomena are of interest because they support the hypothesis that there is perceptual integration of dichotic stimuli. Cutting (1972) proposed that dichotic fusion can occur at various levels of perceptual processing. He considered the effects described by Halves (1969) and by Broadbent and Ladefoged (1957) to be low-level or auditory fusions because both effects depend on a purely acoustic property of the stimuli: the fusion is disrupted if fundamental frequency is varied between ears. Other types of fusion have been discovered in recent experiments where the integration apparently arises at higher perceptual levels and does not depend so critically on acoustic parameters. An example of a higher-level fusion is the phenomenon of "feature blending" (Studdert-Kennedy and Shankweiler, 1970). If stop consonants contrast between ears in two distinctive features, place of articulation and voicing, many responses are "blends" where the voicing feature at one ear is combined with the place feature at the other. Studdert-Kennedy, Shankweiler, and Pisoni (1972) found no change in frequency of blend responses whether vowels are shared between ears, as in /pi/-/di/, or whether vowels vary, as in /pi/-/du/; based on this result they argued that left- and right-ear syllables are not blended acoustically, but that abstract phonetic features become mixed between ears in the course of phonemic identification. Switching of stimulus elements between ears was also reported by Treisman (1970) for dichotic consonant-vowel-consonant syllables, differing between ears in all three phonemes; responses often combined phonemes from opposite ears, for example, "taz" + "geb" → "teb." A final type of dichotic fusion to be considered is the combination of simultaneously presented dichotic consonants to yield a perceived sequence of consonants (Day, 1968). Day presented a word beginning with a stop consonant to one ear and a word beginning with a liquid to the opposite. For example, one ear received "lack" and the other "back." Many subjects (about half) reported hearing a single word beginning with a stop-liquid cluster, "black:" the remaining subjects heard one or both of the actual stimulus items. This effect has been termed "phonological fusion" because the structure of the response is apparently determined by a rule of English phonology which permits stop + liquid clusters syllable initially but prohibits liquid + stop. Phonological fusion occurs with CV syllables as well as with more complex words (Cutting, 1973).

The existence of perceptual fusions such as feature blending, phoneme switching between ears, and phonological fusion argues strongly against the hypothesis that only one ear at a time has access to the speech processor. For fusion to occur, stimuli to both ears must enter the processor and undergo
phonetic analysis within a single "time frame." Moreover, it is interesting that feature blending was observed originally in a task giving a highly reliable right-ear effect. Apparently, the same stimulus material can give rise to different types of perceptual effects—either fusion effects (where the response contains parts of both stimuli) or suppression effects (where only one of the two stimuli is correctly reported). If there is a mechanism for preventing an overload of sensory data into the speech processor, it is unlikely that such a device would work so sporadically, producing sometimes suppression and sometimes fusion. It is argued here that the right-ear effect, lag effect, and fusion effects all represent outcomes of normal speech perception strategies applied to stimuli arriving at the speech processor from both ears.

A significant observation in connection with the lag effect is that fusion responses for stop-liquid pairs occur frequently even when the stop and liquid are temporally offset between ears by 50–100 msec (Day, 1970; Cutting, 1973). Day found a constant fusion rate whether stop or liquid led and regardless of length of offset. In Cutting's work, stop + liquid cluster responses were more common when the stop led, but still occurred frequently when the liquid led. The occurrence of fusion for stop-liquid pairs contradicts the notion that the arrival of a delayed syllable automatically causes the leading syllable to be ejected from the speech processor. The fusion results suggest that if the lagging syllable causes interruption of the processing of the leading syllable, it does so only after the phonetic class of the lagging syllable (stop or liquid) has been determined. We may infer that both lagging and leading sounds undergo at least rudimentary phonetic analysis.

Why would the lag effect and ear effect occur regularly for stop-stop and liquid-liquid pairs but not for liquid-stop or stop-liquid? Two explanations seem reasonable. Perhaps the notion "perceptual processor" has been too broadly defined. If there are separate processors or "feature detectors" for stops and for liquids, then a stop-liquid pairing may not constitute a condition of perceptual competition. The second approach is to view the situation linguistically. In English phonology stop-liquid consonant clusters are permitted syllable initially, while liquid-stop, stop-stop, and liquid-liquid are prohibited sequences. If the speech processor were attempting to integrate the two inputs into a single syllable, a "correct" response would be available only for stop-liquid pairs. Liquid-stop pairs could be construed as stop-liquid, but for liquid-liquid or stop-stop there is no possible response which could integrate the two ears.

Thus, even if inputs from both ears are admitted to the speech processor, there are still at least two levels of processing at which perceptual competition could rise. First, there is the level of "feature extraction," where distinctive phonetic features like voicing, place of articulation, nasality, etc. are identified. Second, there is a higher level where the percept, consisting of a sequence of segmental phonemes, is formed. The phonetic features extracted from the speech signal provide only part of the relevant information for the decisions made at this second stage. Here the phonological rules of the language play a role. For example, nasality occurring during a vowel would be assigned by the phonological rules of English to the following consonant, but for French, nasality would be assigned to the vowel itself. Also, in running speech, not all words are clearly articulated, and context facilitates understanding. The influence of context on speech perception may operate at this stage.
It is proposed here that the right-ear effect arises at the feature extraction level of processing, but that the lag effect and perceptual fusions both arise at the phonemic decision stage. For this model to work, we must assume that the feature extraction stage involves an independent analysis of acoustic cues for both values of a binary feature. That is, rather than being either voiced or voiceless, the output of voicing analysis for English stops might be both voiced and voiceless, providing there were sufficient cues (aspiration, VOT, f0 contour, etc.) to support both analyses. The output from feature extraction would include a weighting or probability based on the number of acoustic cues consistent with each feature value. For example, on a dichotic trial with a voicing contrast between ears (+) voicing could have a .80 weighting and (-) voicing a .50 weighting. Under binaural conditions (i.e., without competition) the analysis would favor a particular value more strongly. This approach is consistent with the finding in dichotic studies that listeners often do identify stops at both ears correctly, and that listeners rarely misperceive feature values which are shared between ears (Kirstein and Shankweiler, unpublished data). This view is consistent with the importance of multiple acoustic cues in perception of many distinctive features and with the context sensitive nature of the cues (e.g., VOT for voiceless stops varies with place of articulation).

In their model of the ear effect, Studdert-Kennedy and Shankweiler (1970) attributed the right-ear advantage to degradation in the auditory representation of left-ear syllables due to their more circuitous neural connections to the left hemisphere. They considered that ipsilateral connections were inhibited in dichotic presentation so that left-ear stimuli could reach speech processing areas in the left cerebral hemisphere only by first traveling to the right hemisphere and then crossing to the left via the corpus callosum. If the right ear provides superior sensory data to the speech processor, this would, in the present model, produce higher weightings for features extracted from right-ear syllables. However, there might conceivably be reasons for weighting the right ear more highly than the left besides the quality of their auditory representation. The model claims simply that both ears are admitted to the speech processor but that the syllables originally presented at the right ear emerge from the feature extraction stage of processing with higher weighting than left-ear syllables. This idea is consistent with the finding that stop consonants contrasting between ears in only voicing or place of articulation give smaller ear effects than consonants contrasting in both features (Shankweiler and Studdert-Kennedy, 1967; Studdert-Kennedy and Shankweiler, 1970).

The output from the feature extraction level is also the input to a higher level of processing where the identity and order of segmental phonemes are decided. Because the input to this stage may contain incompatible feature values (e.g., voiced and voiceless, labial and velar), there must be perceptual strategies available to resolve conflicting analyses. It is proposed that the lag effect reflects one such strategy and that feature blending and phonological fusion reflect others. The blending in the response of place and voicing from opposite ears could be a function of weights assigned during the feature analysis. The highest voicing and place values may have been originally from opposite ears. This account agrees with Studdert-Kennedy and Shankweiler (1970) in considering the locus of feature blending to be at a stage of processing subsequent to feature extraction where features are combined to yield segmental phonemes.

How many distinct syllables the listener can hear for a dichotic pair probably depends on acoustic fusion as well as on the phonological analysis. The
auditory system might recognize two sound sources or only one, and this conclu-
sion could in turn affect the number of phonemic solutions produced by the pho-
nological analysis. A striking result in dichotic experiments is the wide vari-
tion among individuals in frequency of fusion responses for stop-liquid pairs
(Day, 1970) and in their accuracy of selecting a particular ear in ear monitoring
tasks. Such individual differences might arise in the auditory analysis of
spatial localization for stimuli which are physically similar but not identical.

During the phonological analysis a number of possible phonemic solutions
might be rejected; the listeners would "hear" only those that are finally
accepted. Solutions involving low-weighted features would probably be rejected.
It is suggested that the lag effect occurs because as a perceptual strategy leads
tend to be rejected more than lags. That is, if the feature extraction produces
first one feature value and then another, incompatible with the first, the pho-
nological analysis may strongly favor the second result. The second feature
value might be treated as if it were a revision of the first. The claim is being
made that both the lagging and leading syllables undergo phonetic analysis but
that the leading syllable is subsequently rejected. This would occur for stop-
stop pairs or for liquid-liquid pairs. For stop-liquid pairs the leading syllable
would not be rejected because stop-liquid is an acceptable phonemic sequence.

It is argued that the lag effect and right-ear effect both involve speech
perception, but that within the course of speech processing the two effects are
independent. Moreover, it is claimed that the ear effect arises at a level of
processing prior to the lag effect. Is there evidence in the present data to
support these hypotheses? First, there is the fact that the lag effect and ear
effect have completely different temporal parameters: the ear effect is greatest
with the shortest offsets while the lag effect increases with longer delays up to
60 msec. Beyond the different time functions for the two effects, there are
other aspects of the data which suggest that the effects arise at different
levels.

At any delay, the size of the ear advantage was the same whether measured on
lags—as percent correct right ear lags—left ear lags—or measured on leads—as
percent correct right ear leads—left ear leads.¹ It is puzzling that the size of
the ear effect should be the same on lags and leads because the conditions of
competition between ears are quite different in the two cases. The onset of a
lagging syllable always coincides with some portion of an ongoing stimulus at the
other ear but the onset of a leading syllable coincides with silence in the other
ear. Since it is clear from the data that the ear effect is enhanced by simulta-
aneous competition, we would expect the ear difference in lags to be much
greater than in leads. That is, left-ear leads should suffer only slightly in
comparison with right-ear leads, but for lags the left ear should be much poorer

¹For the two-response and temporal order tasks there was no interaction between
size of ear effect and lag vs. lead. In ear monitoring there was a signifi-
cantly smaller ear effect in leads than in lags. However, the analysis pre-
sented here was based on correct responses only. In intrusion responses from
the unattended ear, the right-ear effect was greater in lags than leads so if
all ear monitoring responses are considered, there is no interaction between ear
effect and lag effect.

102
than the right ear. Interestingly, the model just proposed provides a way of accounting for the data. Suppose that the ear effect arises at the feature extraction level of processing and that at this level the ear difference is only on lags: features extracted from left-ear lags have received lower weights than those extracted from right-ear lags. At the phonological decision stage there is a tendency to favor lags over leads. The asymmetry between ears arising initially in lags would then be mirrored in leads as well. Right-ear leads are always paired with left-ear lags; these would be rejected less frequently as a phonemic solution than left-ear leads, which are paired with right-ear lags.

The relation between the right-ear effect and lag effect has always seemed paradoxical because all models of the ear effect imply a temporal advantage for the right ear. This is supported by Springer (1971), who reported a 50 msec right-ear advantage in reaction time for correctly identified dichotic stops. The procedure of temporally offsetting stops at the two ears was proposed originally as a possible method of measuring the right-ear effect (Studdert-Kennedy et al., 1970; Berlin et al., 1973). It was thought that giving the left ear a certain lead time would make left- and right-ear stops equally intelligible. The finding that left-ear lags wash out the ear advantage was, thus, incompatible with our understanding of the ear effect. The present view in which the lag effect and ear effect arise at different levels of perceptual processing offers a solution to this paradox. In this model only left-ear lags suffer from the laterality effect, not left-ear leads. Thus, the use of a left-ear lead to compensate for the right-ear advantage would be effective at the lower level of processing but would be wiped out at the higher level. By assigning the lag effect to a higher level of perceptual analysis than the ear effect, we leave open the possibility that the lag effect may not be specific to speech perception. The lag effect may reflect a general strategy in various sense modalities for handling incompatible inputs in pattern perception.

REFERENCES


