Reprinted from LANGUAGE
Vol. 33, No. 1, January–March, 1957
Printed in U.S.A.

CLOSURE DURATION AND THE INTERVOCALIC VOICED–VOICELESS DISTINCTION IN ENGLISH

LEIGH LISKER

University of Pennsylvania and Haskins Laboratories

Thanks to the means available for speech analysis and synthesis, there is at present considerable knowledge of the acoustic differences whereby we distinguish among the various sounds of English. We know, for example, what the more important acoustic cues are by which listeners decide whether a stop sound is velar, alveolar, or bilabial.[1] In the case of the stops, moreover, we know by and large what the acoustic differences are between the members of each of the homorganic pairs *p–b*, *t–d* and *k–g*[2] (in other words, between the so-called voiced and voiceless stops[3]). But we are as yet unable to state with any exactness the extent to which each of these acoustic differences separately affects our ability to discriminate between the two kinds of stops. Moreover, where progress has been made in this direction it is certainly in some measure due to the fact that we have dealt chiefly with the stops in those positions where the voiced–voiceless contrast is phonetically maximal, that is, in initial and final positions. In the first context, aspiration is a prominent feature of the voiceless as against the voiced stops; in the second, the duration of a preceding vowel varies, depending upon whether the stop is of the one kind or the other. For these two positions it is difficult to separate features which contribute to what the phonetician would call voicing as distinct from aspiration and vowel duration, for there is some evidence suggesting that the acoustic correlates of these latter features may suffice as cues to the voicing distinction in the absence of any other differentia.[4] There is, however, at least one position in English, namely the post-stressed intervocalic, where the voiced–voiceless distinction is observed even though,

[1] Delattre, Liberman, and Cooper, Acoustic loci and transitional cues for consonants, *Jour. Acoust. Soc. Am.* 27.769–73 (1955).

The experiments reported here were all conducted at the Haskins Laboratories in New York, where I benefited from discussions with more of my colleagues there than I have space to list here. The work was supported in part by the Carnegie Corporation of New York and in part by the Department of Defense in connection with Contract DA49-170-sc-1642.

[2] See, for example, Cooper, Delattre, Liberman, Borst, and Gerstman, Some experiments on the perception of synthetic speech sounds, *Jour. Acoust. Soc. Am.* 24. 597–606 (1952).

[3] The nonlinguist is often confused, and legitimately so, by the two different uses to which the term *voice* and its derivatives are put. In one usage the expression *voiced–voiceless distinction* is the name of each of the phonemic contrasts /b/–/p/, /d/–/t/, /j/–/c/, /v/–/f/, /z/–/s/, etc., where the first of each pair of opposed phonemes is called VOICED and the second VOICELESS. In another usage, a speech sound is called VOICED only if glottal tone is involved in its production; in the absence of audible glottal tone a sound is VOICELESS. That these two usages are not synonymous is shown, for example, by experiments whereby one and the same acoustic segment may be perceived as voiced or voiceless depending on the context. See P. Denes, Effect of duration on the perception of voicing, *Jour. Acoust. Soc. Am.* 27.761–764 (1955).

[4] See reference cited in fn. 3.

according to the phonetic descriptions of these stops, neither aspiration nor vowel duration operates to a significant extent.[5] From the accepted phonetic descriptions we have every right to expect that in post-stressed intervocalic position the voiced–voiceless distinction is phonetically minimal. Thus this position poses both a special challenge and an opportunity to acoustic phonetics.

I chose for study the pair *p–b*, assuming that any findings would bear on the *k–g* distinction as well. (The *t–d* contrast, in those American dialects where it exists in post-stressed intervocalic position, is phonetically not altogether comparable with the other two.) The examination of the *p-b* contrast began with the inspection of spectrograms (Kay Sonagraph), and then proceeded to a series of tape-splicing experiments on recorded samples of speech. In the first stage of this work my attention was drawn to what seemed to be a consistent difference in CLOSURE DURATION—that is, the time interval between termination of the vowel-formant transition preceding the stop and onset of the transition to the following vowel. While it was recognized at the outset that this difference could not possibly turn out to be of major cue value in more than a restricted number of environments,[6] it appeared none the less worth while to try to determine the extent to which a variation in closure duration might be phonetically relevant.

EVIDENCE FROM SPECTROGRAMS. A number of sentences including trochees in the position of primary stress, as well as trochees spoken in isolation, were recorded on magnetic tape, and from these recordings spectrograms were made. All trochees thus analyzed contained either a *p* or a *b* in intervocalic position; some of the words used formed minimal pairs with respect to the *p–b* contrast: *ruby–rupee*[7], *rabid–rapid*, *stable–staple*. On examination of the spectrograms these gross acoustic differences were noted as consistently present in the patterns:

(1) The first, second, and third formant transitions following closure begin at lower frequencies and move less abruptly for *b* than for *p*.

(2) The closure duration is greater for *p* than for *b*. Thus, for a list of thirty-four isolated words spoken at moderate conversational speed,[8] closure durations for *p* fall in the 90–140 msec range, with an average value of about 120 msecs, while values for *b* vary from 65 to 90 msecs, with an average of 75 msecs.[9]

[5] Phonetic descriptions of the allophones of /p/ and /b/ proper to this position agree that these sounds differ in the matter of voicing. Trager and Smith, *An outline of English structure* 32, describe the *p* in this position as both voiceless and FORTIS relative to the *b*; other phoneticians, however, are not consistent in reporting a difference of the kind labeled by the terms *fortis* and *lenis*. Many do not apply these terms at all; and Heffner, *General phonetics* 132, reports without comment that Jespersen, *Lehrbuch* 107, described this *p* as lenis.

[6] By its very definition, closure duration can be measured only when the stop is medial (and final, if audibly released). In a series of preliminary measurements, overlapping values of closure duration were obtained for the two stops (1) preceding main stress (*abúse, repúte*; *rebél, repél*); (2) in gemination (*grab bag, cap pistol*); (3) between unstressed vowels (*sýllabus, éntropy*).

[7] In my dialect *rupee* has weak stress on the second syllable.

[8] Each word has a duration of about 500 msecs.

[9] Closure durations for the same words when included in longer utterances are generally shorter, but the same relation holds between *p* and *b*.

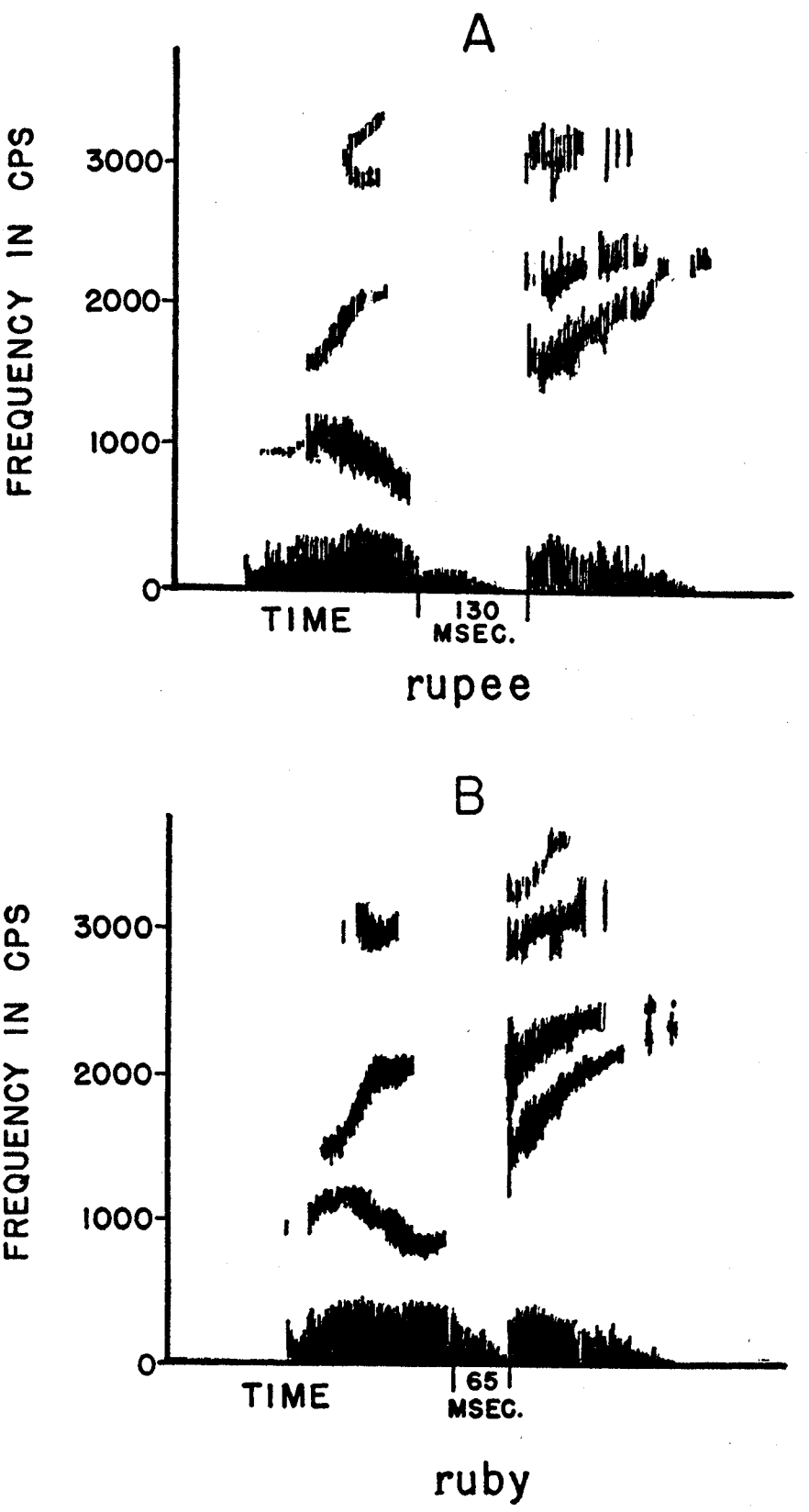
A
FREQUENCY IN CPS
3000
2000
1000
0
TIME
130
MSEC.
rupee
B
FREQUENCY IN CPS
3000
2000
1000
0
TIME
65
MSEC.
ruby

FIG. 1

Other differences were observed in the spectrograms of words containing *p* as against those containing *b*, but they were not found invariably. Such differences were the following:

(1) The lowest harmonics of the vocal cord tone continue throughout the entire duration of closure for most *b*'s, while the *p*'s often show no trace of acoustic energy throughout the duration of closure. In some cases, for example in the spectrograms of Fig. 1, *b* closures do not show these harmonics continuing without interruption, while on the other hand *p*'s sometimes show these low harmonics extending over a considerable portion of the closure interval.

(2) The vowels preceding *p* tend to be shorter in duration than those preceding *b*.[10]

(3) Transitions of vowels preceding *p* shift frequency position more rapidly than transitions before *b*.

(4) The overall intensity at onset of phonation following closure is greater, or increases more rapidly, for *p* than for *b*.

(5) In the case of *p* there is sometimes a very brief interlude of noise between release of the stop and onset of the vowel.

These differences are not seriously examined in the present discussion, although the fact that they do not consistently appear in the spectrograms studied by no means implies that they are of negligible cue value.

EVIDENCE FROM TAPE-SPLICING EXPERIMENTS. The primary purpose of the experiments to be described was to determine whether the difference in closure duration observed in the spectrograms has differential cue value or not. The technique of tape cutting and splicing makes it possible to artificially vary the duration of closure while holding constant all other features of a recorded speech sample. This is accomplished simply by cutting the tape recording of a stimulus word into three parts, taking care that the middle section of tape contains all the signal between the transition preceding and the transition following closure. This middle portion of tape is then replaced by blank tape of any length desired, thus effectively permitting variation of closure duration without altering anything in the stimulus word either before or after the closure. For each cut and splice a spectrogram is made to insure that the operation has been accomplished as intended. Cuts are made at an angle of 10° to a line normal to the longitudinal tape axis; this angle is sufficient to prevent introduction of any audible 'cut-noise', and at the same time introduces less than a three millisecond smear in the time dimension.

EXPERIMENT 1.

(a) The closure durations for *p* in one production of *rupee* (Fig. 1a) and for *b* in a production of *ruby* (Fig. 1b) were determined to be 130 and 65 msecs respectively. The closure duration of the *p* was effectively reduced to 65 msecs by removing 0.35 in. from a properly selected portion of the tape.[11] Listeners reported hearing not *rupee* but *ruby*.

[10] For ten minimal and near-minimal pairs the average difference was about 25 msecs. But despite the small size of the sample there were two cases where values clearly overlapped.

[11] To eliminate any possibility that the low-frequency energy occurring within the *p*

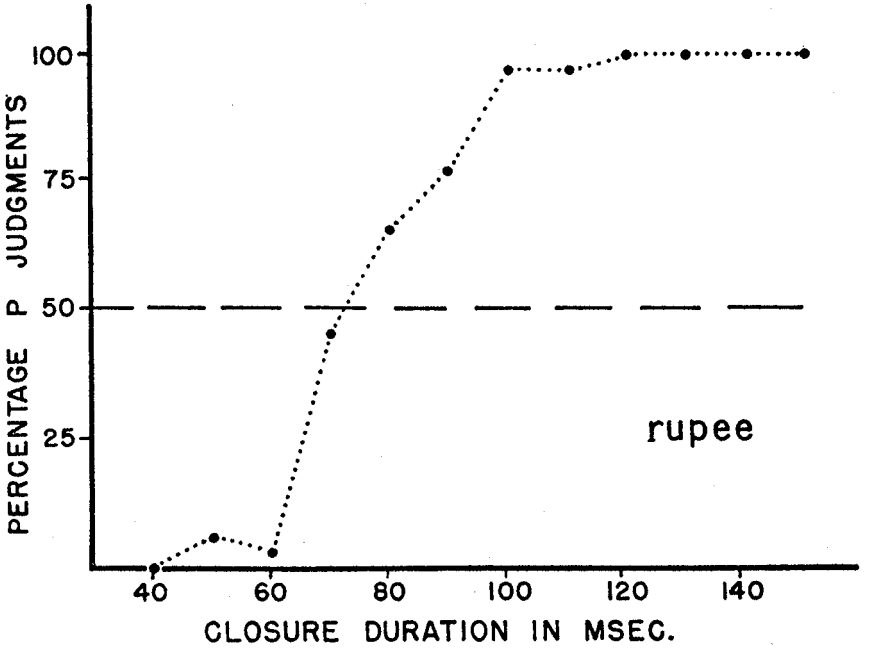

FIG. 2

(b) In order to learn more precisely the effect of reducing the closure time of *p*, this same production of *rupee* was subjected to a succession of cuts and splices in such a way that the closure duration of the included stop was varied in 10 msec steps over the 40–150 msec. range. The thirteen stimuli thus generated, acoustically identical except for this one feature of closure duration, were then presented in random order to a group of American subjects, who were asked to identify each stimulus as either *ruby* or *rupee*. Thirty-five independent judgments were secured for each stimulus. Fig. 2 shows the relation between closure duration and the percentage of *p* (really *rupee*) judgments. The closure duration for which judgments were divided evenly between the two words lies between 70 and 80 msecs.

EXPERIMENT 2.

(a) The section of tape corresponding to the *b* closure of *ruby* was replaced with blank tape of the same length (0.975 in. = 65 msecs). Although all low-frequency voicing information was thus removed from the closure interval, listeners still heard *ruby*.

(b) The artificially 'unvoiced' *b* closure of *ruby* was varied in duration over the 40–150 msec range in 10 msec steps, and a test was administered similar to the one in Experiment 1b. The percentage *p* judgments for a group of nine listeners are shown in Fig. 3. Judgments are divided evenly between the two categories at about 105 msecs closure duration.

---

closure interval (Fig. 1a) might be a factor in any shift in the distribution of judgments, the portion of tape corresponding to this interval was first replaced by blank tape.

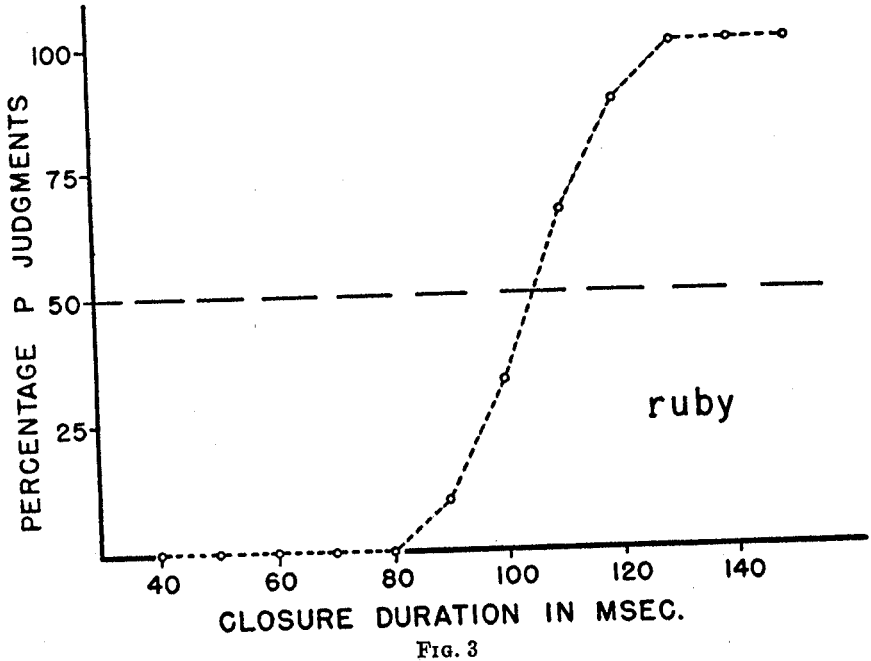

FIG. 3

DISCUSSION. These experiments point to closure duration as a major cue to the voiced-voiceless distinction in intervocalic stops, at least when the interval of closure is acoustically void. Comparison of the curves in Figs. 2 and 3 permits us to form a rough idea of the relative importance of this measure as against the sum of whatever other differentiating features have been retained in the test stimuli—presumably all but that contained in the glottal tone. The two curves are remarkably similar in shape, the chief difference being one of about 30 msecs in their general locations on the abscissa, so that we might say that the sum of all other cues bearing on the *p–b* contrast balances the effect of a 30 msec difference in closure duration. However, for durations of 40–60 msecs and 130–150 msecs these other cues appear to contribute nothing to the listeners' ability to differentiate between the two stops. Neatly enough, although not at all unexpectedly, the closure duration for which there is maximum difference between the two distributions of judgments (i.e. maximum contribution of the other cues) is 90 msecs, which is precisely the duration marking the boundary between the *p*'s and *b*'s measured in our spectrograms.

EXPERIMENT 3. The cues other than closure duration which serve to differentiate *p* from *b* in the tests already described cannot easily be isolated by the cut-and-splice technique. Since, however, they obviously are contained somewhere in the acoustic signal, two additional tests were made to discover whether they could be said to occur predominantly either preceding or following the closure.

(a) The portion of *rupee* preceding the *p* closure was combined with the portion of *ruby* following the *b* closure. For this combination a test was run in which a

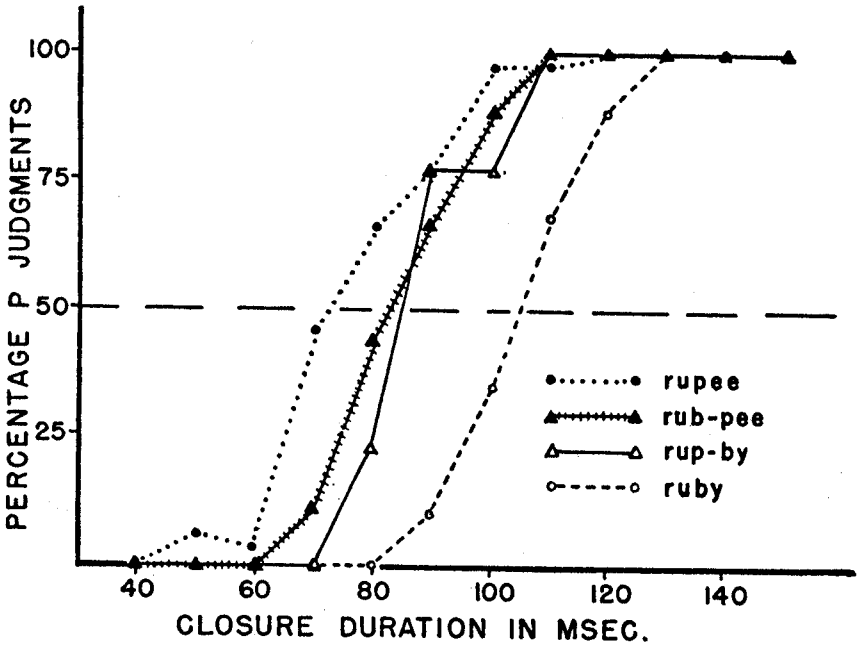

FIG. 4

silent interval between the half-words was varied in duration over the 40–150 msec range, in 10 msec steps. The percentage *p* judgments are shown by the curve marked *rup-by* in Fig. 4.

(b) The first part of *ruby* and the second part of *rupee* were combined and tested with the several durations of medial silence previously used. Percentage *p* judgments are given by the *rub-pee* curve of Fig. 4.

DISCUSSION. Comparison of the *rup-by* and *rub-pee* curves with those for *ruby* and *rupee* indicates that cues to the voiced–voiceless distinction are to be found in the acoustic signal both before and after closure. The *rub-pee* curve lies closer to the *rupee* than to the *ruby* curve, and perhaps the *rup-by* curve tends very slightly toward the *ruby* position; but there is not a large enough difference between the two to support any statement that would impute greater importance to cues in the post-closure position.[12] The effect of the cues in either position is roughly the same: to shift slightly the value of the crossover duration between *p* and *b*.

CONCLUSION. The experimental results in sum support the view that closure-durational differences play a major role in the voiced–voiceless stop distinction in the type of context studied. Subject to the stricture that the cue value of

[12] There is some basis for the expectation that the post-closure cues will outweigh the others. For example, in a tape-splicing experiment in which the words *ragged* and *rabid* were cut and recombined as *rag-bid* and *rab-ged*, the intervocalic stops reported by listeners were *b* and *g* respectively.

glottal tone has not been adequately determined,[13] it appears that there exists a closure duration, perhaps about 75 msecs for trochees in isolation, such that all stops of briefer duration are heard as voiced by American subjects; and that there is another closure duration, 130 msecs for isolated trochees, such that stops of this duration or longer are assigned to the voiceless category.

[13] Glottal tone was introduced into the closure interval with generally negative effect, but certain problems encountered in this operation have as yet not been adequately enough resolved to permit full-scale testing.