Phonetica
Editor: K. Kohler, Kiel
Publishers: S. Karger, Basel
Separatum (Printed in Switzerland)

Phonetica *37:* 327–343 (1980)

Sonority and Syllabicity: Acoustic Correlates of Perception

P.J. Price

Haskins Laboratories, New Haven, Conn., USA

Abstract. Evidence is accumulating that native speaker-hearers are not as consistent, confident, or in agreement, on counting the number of syllables in natural utterances as is commonly assumed. There are, however, instances where speaker-hearers give clear, consistent syllable counts. It is the position of this paper that the unclear cases as well as the clear cases are phonetically classifiable in terms of sonority. The experiments presented here are intended to delimit what is meant by sonority in acoustic terms.

Introduction

The syllable is more often appealed to than defined [see surveys in Bell, 1978; Price, 1978]. The problems arising from attempted definitions are sometimes 'explained away' by positing the syllable as a 'natural perceptual unit' [see, e.g., Hooper, 1976]. In this view, native speakers have strong intuitions about syllables, but definitions cannot be developed from these intuitions due to complex interactions of morphology, phonology, orthography and phonetics. However, evidence is accumulating that even this weak claim for the syllable may not hold. Bell [1975] tried a variety of methods in attempting to elicit natural units. Lebrun [1966] asked for syllable counts of short sentences repeated as often as subjects wished. Price [1978] asked for syllable counts of very short utterances with dialect background strictly controlled. In all these studies, where the assumption about the intuitive status of the syllable was tested, the results converged: native speaker-hearers were not extremely consistent, confident or in agreement on syllable counts. Moreover, automatic segmenting algorithms tend to fail in areas phonetically similar to those where native speakers are inconsistent or disagree: neighboring segments of roughly the same degree of sonority. Mermelstein [1975] mentions cases of syllabic versus nonsyllabic /r/ or /n/ ('hɵrizon' as /hrajzən/ or 'apparently' as /əppɛrnli/ [Mermelstein's notation]) and contiguous vowels as in 'so I'. The inconsistency of listeners, the lack of agreement across listeners, and the failure of algorithms to match dictionary syllabications do not necessarily imply that the syllable does not exist or is useless. Clear cases exist, and, further, the unclear cases may be taken as evidence that a more flexible definition of the syllable is necessary, i.e., a definition that accounts for both the clear and the unclear cases. Such a definition in terms of sonority will be outlined here. The experiments reported here investigate the acoustic correlates of the perceptual term 'sonority'.

The terms 'prominence' or 'sonority' have been applied to various aspects of speech: as an overall feature of voice quality [see, e.g., Wedin et al., 1978], as a feature of stress or accent carried by certain syllables [see, among many others, Gaitenby and Mermelstein, 1977], and as a feature of segments forming the internal structure of syllables. Only the latter usage of the terms will be dealt with in the present study. Acoustic correlates for these terms that have been investigated, however, show a good deal of overlap. Most of the studies consider fundamental frequency (absolute value or extent of excursion), intensity, and duration. The relative roles of these acoustic attributes are under debate, perhaps largely because of methodological differences. I know of no studies that have investigated experimentally the acoustic features of sonority for segments smaller than the syllable, although there has been a fair amount of theorizing.

Since Sievers [1893] the internal structure of the syllable has been discussed in terms of sonority, strength or prominence [see, e.g., Bloomfield, 1933, p. 120; Venneman, 1972; Hooper, 1976]. Bell and Hooper [1978] provide a good survey of cross-linguistic evidence for such hierarchies. The basic notion is that syllabic peaks are peaks of sonority, and that segments increase in sonority before the peak and decrease in sonority after it. This implies that English /l/ is more sonorant than English /p/ because instances of /#plV_/ and /_Vlp #/ occur but not /#lpV_/ or /_Vpl #/. The two latter examples may in fact be realized in English, but only if the /l/ is sonorant enough to be a syllable peak itself, as in 'I'll put it away' (/ƚpʊᴅɪᴅəwei̯/), or 'people' (/pipƚ/). In this notation /l/ and /ƚ/ are treated as phonemically distinct; some linguists would write /ƚ/ as /əl/ and lodge the distinction in another segment.

Some cross-linguistic evidence [see, e.g., Bell, 1978] may be taken to mean that sonority hierarchies are language specific: Russians seem to feel that / #rtV/ structures are one syllable, although in English a monosyllable of this structure is impossible. If utterance-initial Russian /rtV/ structures are monosyllabic structures, then, in terms of sonority, we are forced to say that Russian /t/ is more sonorant than Russian /r/, and we cannot say the same for English. However, this does not necessarily mean that Russian and English /r/s are of the same sonority, or that sonority has no explanatory value. If an acoustic definition of sonority is developed, the relative sonority of linguistic units from different languages can be compared without reference to the language-specific phonotactics of the two linguistic systems.

A fundamental generalization of the sonority theory is that vowels are more sonorant than consonants. A crucial aspect of the theory, however, is that it is useful to divide the set of phonetic segments into a richer classification system than one involving only consonants and vowels. Even linguist who make no specific mention of sonority hierarchies may define the syllable in terms of a vocalic nucleus surrounded by consonantal margins (onset and/or coda) Hjelmslev, 1938; Trager and Bloch, 1941; Trager and Smith, 1951; Hockett, 1958; Greenberg, 1962; Delattre, 1965; Chomsky and Halle, 1968; Studdert-Kennedy, 1976]. Peaks of sonority are, in general, vowels; the troughs are generally consonants. Defining syllables in terms of alternations of consonants and vowels works insofar as the classes of consonants and vowels are clear. By examining the cases of clear and unclear vowels, we can outline the classes of clear and unclear sonority peaks, and, hence, predict listener inconsistency, disagreement, or possible problems for automatic segmentation algorithms and the intuitions of a native speaker-hearer.

'Clear', 'good' or 'prototypical' vowels correspond to prototypical syllabic peaks and are fairly easy to describe in articulatory or in acoustic terms. They are characterized by an open vocal tract, vibrating vocal folds, and relatively long duration. Good consonants or syllabic margins are characterized by the opposite: a constricted or closed vocal tract, interrupted voicing, and relatively short duration (a 'transient' as opposed to 'steady-state' character of the more audible portions of the articulation, see Delattre, [1965]). These three factors – degree of opening of the vocal tract (Opening), glottal or other source characteristics (Source), degree of transience (Rate of Change) – are all involved in sonority. DELATTRE [1940, 1944/1966] discusses syllabic structure in terms of articulatory features that partially overlap with those just outlined: aperture (along with articulatory force, articulatory direction and articulatory distance) is discussed in DELATTRE [1940] and expanded upon in DELATTRE [1944/1966]; and change is discussed in DELATTRE [1965] in both articulatory and acoustic terms. Sonority is never explicitly mentioned, although DELATTRE [1944/1966] describes a scale from more open (more vocalic) to more closed (more consonantal), and DELATTRE [1965] discusses syllabic margins and nuclei, as well as subclasses of the class of margins. The role of source characteristics is dismissed in DELATTRE [1965], however, since 'whispered speech is perfectly intelligible and therefore contains all the acoustic cues essential to speech perception' (p. 15). I would doubt, however, that whispered speech is as intelligible as nonwhispered speech. Further, I would suspect that the articulatory and acoustic differences of these two speech modes are large enough that an equal distribution and function of 'cues' is not to be expected.

All the experiments to be reported in this paper bear on the meaning of sonority and the role it plays in syllabicity judgments. In the acoustic domain, these three factors may correspond to the presence versus absence of a clear formant structure, voice versus hiss (or no) excitation source, and steady-state versus transient formant patterns. The Rate of Change characteristic may apply to parameters other than formant structure (fundamental frequency or amplitude, e.g.), but other aspects will not be specifically investigated here. In the idealized situation, then, chains of syllables are series of vocal tract openings and closings with the open parts (syllabic peaks) corresponding to vowels and the closed parts corresponding to consonants. In the clearest cases this is true, with the exceptions that: (1) though the closed 'hold' portion is identified with the consonant articulatorily, acoustically (auditorily) this portion is not always very salient, hence, the transitions between the closed hold and the open hold have come to be identified acoustically with the consonant; transitions, by definition, manifest a faster rate of change than 'holds'; thus, Rate of Change is also involved in sonority; (2) vocal tract openings and closings cannot be heard unless they are excited via glottal buzz, and/or friction noise at the glottis or above; and (3) there is a tendency to think of these opening and closing gestures as organized into discrete consonant-plus-vowel units, which may imply that opening transitions are more sonorant than closing transitions; this matter, however, will not be discussed further here.

It is important to notice that the characteristics Opening, Source, and Rate of Change are all relative rather than absolute terms. Furthermore, there are many cases where only one or two of these qualities may appear. For example, insofar as openness of the vocal tract indicates degree of vocalicness, open vowels (say, [a]) are more vowel-like than close vowels (say, [i] or [u]). A number of linguists [e.g., Hockett, 1942; Pike, 1943, pp. 110–111; Jones, 1950, p. 15] treat [j] and [w] as nonsyllabic counterparts of [i] and [u]. The orthography chosen may highlight this view: they are often written identically, sometimes with a diacritic added to distinguish them. Close vowels are not only similar to glides (sometimes called 'semi-vowels' or 'semi-consonants'), but they also risk confusion with segments that are not 'semi' but 'real' consonants: slight deviations in control of air supply for constricted vowels can produce friction noise, causing a similarity to fricatives (as in, e.g., American English 'heed your' [hidi̯ɚ] – [hidžɚ]). Some vowels are more vowel-like than others with respect to Opening, Source, or Rate of Change. All three characteristics are a matter of degree. Voicing, as a source characteristic, is a matter of degree both in its relative onset [see, e.g., Lisker and Abramson, 1964] and in the amount of accompanying friction noise (as in, e.g., voiced versus murmured versus whispered vowels).

Furthermore, these three characteristics are relatively independent. That is, they may differ as indices of how vocalic (sonorant) a particular segment is. For example, in voiceless vowels, the mouth can be very open, steady-state portions may be clearly present, but voicing is absent. Glides represent a case in which the vocal tract is relatively open and voicing is present, but there is a rapid rate of change. There are also cases in which a voiced steady-state period occurs when the vocal tract is obstructed, as for nasals, voiced obstruents, and liquids. Voiceless fricatives may have a long steady-state period, but rank low on the scales of Source and Opening. In fact, the set of clear vowels or clear consonants is probably smaller than the set of unclear cases.

When one considers the combinatorial properties of these elements, the problem of syllables becomes more complex. In terms of the characteristics of prototypical vowels (Opening, Source, and Rate of Change), prototypic syllables can be defined as alternations of prototypic vowels and prototypic consonants. This predicts that listeners will agree more on the number of syllables in utterances that consist of alternations of prototypic consonants and vowels than they will on alternations of the less clear cases. Support for this hypothesis is found in PRICE [1978].

The present study considers liquids (English /r/s and /l/s) in /C_V/ position. The degree of openness of the vocal tract cannot be systematically varied for most sounds, since we tend to define classes of phones largely with respect to this aspect. It is possible to vary relative and absolute duration, amplitude, and voice onset time (VOT). The present study investigates these aspects of sonority in the case of the pairs *plight–polite* and *prayed–parade*. There are many such pairs where the sonority or prominence of the /l/ or /r/ may be all that is needed to keep lexical items distinct: *bray–beret*, *round–around*, *long–along*, *set lit–settle it*. There are also more ambiguous pairs where it is not clear that there is a distinction at all: *hire–higher*, *aisle–I'll*, etc. Assuming that syllabic peaks are peaks of sonority, then increasing the sonority of certain segments of variable sonority should lead to an increase in the number of syllables perceived. Experiment 1 tests the roles of duration and amplitude for their contribution to the sonority of /r/ in natural productions of *prayed* and *parade*. Experiment 2 tests VOT and the relative roles of voicing, hiss, and silence in a synthetic *plight–polite* continuum. Experiment 3 tests the roles of relative versus absolute /l/ durations in the same synthetic *plight–polite* continuum.

Relative intensity and relative duration together led to the best prediction of perceived syllable stress in GAITENBY and MERMELSTEIN's [1977] study, with the value of intensity outweighing that of duration and of fundamental frequency. One might expect some overlap in perception of prominence within and across syllables, but the two are not necessarily identical.

Experiment 1: /r/ Duration versus /r/ Amplitude in Natural Productions of prayed *and* parade

In this experiment, amplitude and duration of the /r/ portions (defined as the portions where F_2 and F_3 are close to each other, an acoustic indication of retroflexion) in natural productions of *prayed* and *parade* were manipulated by computer editing and presented to naïve listeners for labeling.

Measurement Data

Ten productions each of *prayed* and *parade* by each of two talkers, one male and one female, were measured. VOT was measured from waveforms, /r/ duration from spectral displays. Amplitude of aspiration and of the /r/ were measured in decibel down from the peak by computer analysis.

While the duration of aspiration (VOT) was, on the average, about 10 ms longer for *prayed* than for *parade*, the range of these durations for *parade* (40–60 ms) was wholly included in the range for *prayed* (40–80 ms). Thus, it was decided not to manipulate this parameter in the present experiment. The amplitudes of aspiration did not differ significantly either in range or in mean value. /r/ durations did vary significantly: the mean for *prayed* was found to be 80 ms with a 55- to 110-ms range, and the mean for *parade*, 145 ms with a range of 110 to 170 ms. While some tokens of *parade* were apparently pronounced by the male talker as /pəreɪ̯d/, as evidenced by spectral displays, amplitude envelopes (two humps in the display), and by listening, all productions by the female talker (and most productions by the male talker) were pronounced /pr̩eɪ̯d/ – with syllabic /r/. The amplitude levels for the /r/s were measured at the amplitude peak for the /r/ in decibel down from the peak amplitude for the entire token, where such a peak occurred: for some tokens, amplitude increased constantly throughout the /r/. When such was the case, amplitude levels were averaged over 12.6-ms intervals throughout the /r/, and a mean /r/ amplitude was calculated. This ad hoc procedure may not have resulted in a meaningful measurement. In fact, average amplitude levels by this measure differed only by 1 dB.

Stimuli

In this experiment, /r/ duration and amplitude were altered independently. Based on the measurement data, an 'average' token for each of the source words was selected from among the productions by the male talker, since his formants were easier to track and measure, and the longer pitch periods made it easier to extend the /r/ by pitch pulse iteration without disturbing the naturalness of the tokens. From the *parade* chosen, /r/ duration was shortened by deleting pitch pulses after onset of voicing. Deleting 30 ms resulted in a derived stimulus with a 115-ms /r/. Similarly, deleting 60 ms resulted in a derived stimulus with a 85-ms /r/. The most drastically shortened stimulus, then, had an /r/ duration roughly equal to the mean for the set of *prayed* tokens. Amplitude of the /r/ portion was decreased by 6 dB for the original and for the portion of the /r/ remaining in the shortened versions. For the *prayed* chosen, /r/ duration was increased by iteration of the first pitch pulse. Two stimuli were derived in this fashion, one with /r/ duration increased by 30 ms (110 ms total /r/), and one with /r/ duration increased by 60 ms (140 ms total /r/). Adding 60 ms to the original token created a stimulus whose /r/ duration (140 ms) was roughly equal to the mean value for tokens of *parade* (145 ms). For each of these three /r/ durations two amplitude levels were used: the original, and 6 dB up from the original. Four tokens of each stimulus appeared in a randomized test sequence on separate tapes for each source word.

Subjects

The subjects were 12 paid volunteers (all Yale undergraduates), who were asked to listen to the tapes twice, in counterbalanced order, once to count syllables and once to identify words by circling either 'prayed' or 'parade' on prepared answer sheets.

Results and Discussion

The responses resulting from these two tasks (syllable counting and word identification) did not differ significantly: 'one-syllable' responses correspond to 'prayed' responses to within 6% for every source stimulus. The two tasks combined yield 96 responses to each of the stimuli (4 tokens × 12 subjects × 2 tasks). Figure 1 plots percent one-syllable or 'prayed' responses versus /r/ duration for both source words, and for both amplitude levels. It is clear from this figure that /r/ duration has a decisive effect on listener judgments for both source words. It appears that the effects of amplitude are negligible for short and for long /r/ durations. The case of the 'medium' durations is less simple: for source *prayed*, amplitude affected judgments significantly, but this was not the case for source *parade*. This asymmetry may indicate a general difference between productions of *prayed* and productions of *parade*, at least for this talker, or it may be due to token-specific differences. Although an 'average' token of each source word was chosen, these were natural productions and, hence, differ along many uncontrolled dimensions. In any case, the set of 'medium'-duration stimuli support the conclusion that duration is a more effective cue than amplitude: when the duration used resulted in judgments split between the two

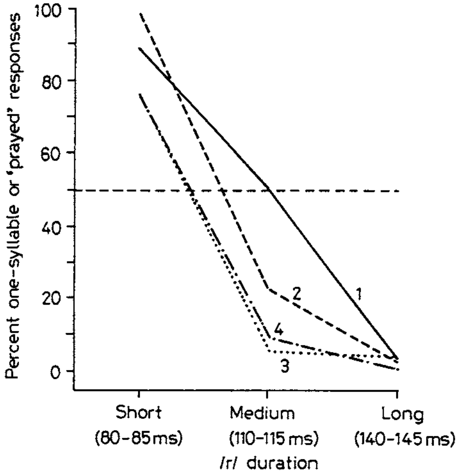

Fig. 1. /r/ duration and amplitude (n = 96). 'Short', 'medium', and 'long' refer to /r/ durations. The 'short' condition corresponds to *prayed* with original /r/ duration and to *parade* with 60 ms of the /r/ deleted. The 'medium' condition corresponds to /r/ lengthened or shortened by 30 ms for sources *prayed* and *parade*, respectively. The 'long' condition refers to *parade* of original /r/ duration and to *prayed* with /r/ duration increased by 60 ms. Represented in plots 1 and 2 are 'prayed' responses to stimuli derived from source *prayed*. Plots 3 and 4 correspond to source *parade*. Plots 1 and 3 correspond to original /r/ amplitude levels, plots 2 and 4 to manipulated /r/ amplitude levels (plot 2: /r/ + 6 dB; plot 4: /r/ —6 dB). Note that /r/ duration has a decisive effect on labelings. Amplitude may have some effect for the 'medium' condition.

words (curve 1), amplitude had a significant effect (curve 2); when, however, the duration used resulted in judgments strongly in favor of one word or the other ('parade', in this case, curve 3), amplitude had little effect (curve 4). The measurement data indicate that absolute /r/ durations may well be ambiguous as indices of *prayed* versus *parade* in natural productions: the longest /r/ of *prayed* was of the same duration as the shortest /r/ for *parade*. When the words are embedded in sentences, it is likely that the ranges of the /r/ durations for the two source words will overlap. In sum, duration of /r/ for these words seems to be a sufficient cue to their distinction. Amplitude may play a role where this cue is neutralized. While more open vowels are generally louder (of higher amplitude level) than less open vowels, differences in formant frequencies, or vowel color, are also generally involved. If sonority is considered in articulatory terms, then the rather small effect of amplitude is reasonable, given that spectral information was unchanged. The independent testing of amplitude and spectral information as they relate to the openness of the vocal tract is left for future research.

Experiment 2: VOT in Synthetic Stimuli

The measurement data for *prayed* and *parade* revealed longer mean values of VOT for *prayed* than for *parade* for both talkers. Although the ranges of these values overlap heavily for both talkers, the mean difference is 10 ms for the male talker, 20 ms for the female. Further, the longer VOTs are correlated with shorter /r/ durations. The situation is not entirely parallel to that of voiced versus voiceless stops in initial position: (1) the duration of the segment following the initial stop (as well as VOT value) serves to distinguish 'prayed' – 'parade', 'plight'–'polite', etc., but not initial /bdg/ versus /ptk/; (2) the differences in VOT correspond to differences not in the voicing of the initial stop but in the syllabicity of the following segment, and (3) increases in VOT are not necessarily correlated with increases in formant frequency onset values, since the liquid may be steady state throughout a wide range of VOT values. However, both situations involve coordination of vocal tract opening and the onset of voicing. In other words, sonority is not merely a matter of opening and closing the vocal tract, but of vocal tract dynamics and their interaction with laryngeal control. Thus, a continuum that switches judgments from one to two syllables based on VOT alone is evidence that the perceptual significance of the relative timing of vocal tract gesture and laryngeal pulsing, as evidenced by VOT, is generalizable beyond the class of initial stop consonants.

Stimuli and Subjects

The stimuli for this experiment were prepared on the OVE-3 synthesizer at Haskins Laboratories. Stimuli perceived as 'plight' and 'polite' were created. Naturally produced /p/ bursts were added word-initially by waveform editing. All stimuli included 35 ms of transitions between this initial burst and a 91-ms steady-state /l/. Spectrograms of endpoint stimuli are shown in figure 2. In order to avoid the intrusion of initial /b/ percepts, the shortest VOT used was 35 ms. As VOT was increased from 35 to 112 ms in 7-ms steps, the buzz-excited steady-state /l/ duration was thereby decreased from 91 to 14 ms, while the hiss-excited steady-state /l/ duration increased from 0 to 77 ms. A set

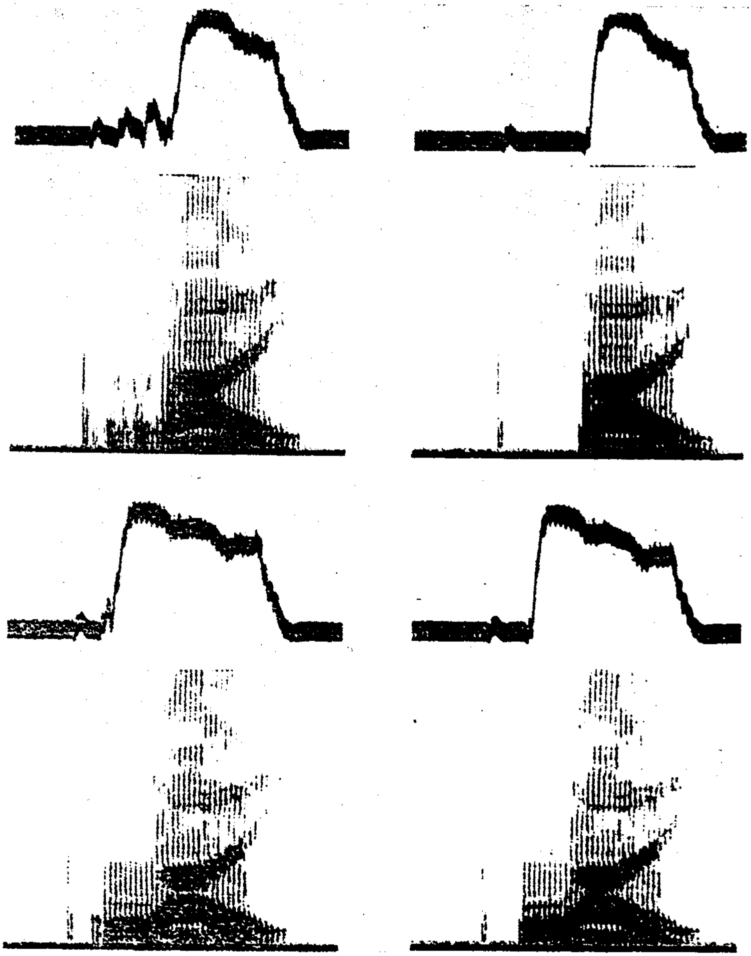

Fig. 2. Spectrograms of synthetic stimuli. At the left are the endpoints of the hiss condition of the synthetic 'plight'–'polite' continuum. At the right are the endpoints for the silence condition of this continuum. The two displays at the top represent the stimuli with longest VOT values, and, hence, shortest voiced /l/ duration. The displays at the bottom represent the shortest VOT values used which correspond to the longest duration of the voiced steady-state /l/.

similar to these 12 stimuli was created in which silence replaced the hiss between initial burst and voicing onset. These stimuli are probably less representative of actual articulations than the first set, but they do permit the investigation of the perceptual effect of hiss versus silence. It is reasonable to use these stimuli a priori since aspiration hiss may not al-

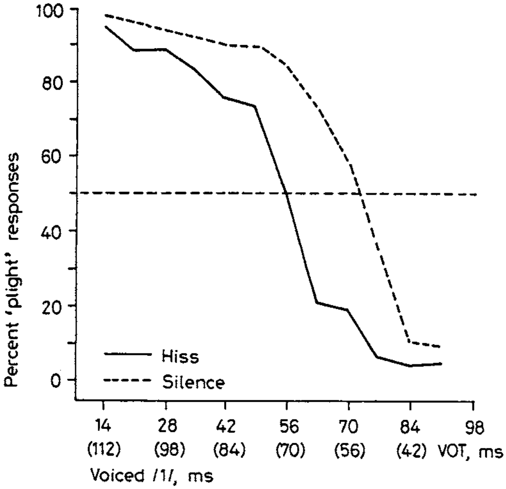

Fig. 3. VOT or duration of voiced /l/ (n = 48). The solid line indicates percent 'plight' responses to stimuli in which the interval between burst and voicing onset was hiss-filled. The dotted line represents the condition in which silence filled this interval. The abscissa is labeled with duration of voiced /l/. Underneath the /l/ duration figures, the corresponding VOT values appear in parentheses. All stimuli have the same steady-state /l/ duration (91 ms); they differ in the point at which formant excitation is switched from hiss to buzz. Note that the longer the voicing of the /l/ (i.e., the shorter the VOT), the more 'polite' responses elicited (i.e., fewer 'plight' responses). Further, the duration of voicing of the /l/ at the 'plight'–'polite' cross-over point is about 16 ms longer when silence rather than hiss is present in the interval between burst and voicing onset. This suggests that, with respect to duration, voicing is more effective than hiss in cueing 'polite' rather than 'plight', and that hiss, in turn, is more effective than silence.

ways be audible in speech contexts, and a posteriori because they in fact result in a convincing 'plight'–'polite' continuum. Four randomizations of these 24 stimuli were presented to 12 paid volunteers (Yale undergraduates) for labeling as 'polite' or 'plight'. The graphs in figure 3 thus represent 48 responses to each stimulus (12 subjects × 4 tokens).

Results and Discussion

In figure 3 it is seen that VOT is an effective cue to the *plight–polite* distinction. Further, it appears to make a difference whether the period between burst and onset of voicing is noise-filled or silent: subjects, on the average, need about 16 ms longer voiced steady-state /l/ to hear *polite* versus *plight* when silence replaces hiss in this interval. The steady-state portion of the /l/ is crucial to hearing *polite* versus *plight*, but the voiced part is more critical than the voiceless part. That is, the total steady-state /l/ is not the critical factor here: all stimuli have the same duration in this respect. What appears to be critical is the overall sonority of the /l/. As is shown here, duration of voicing of the /l/ effectively switches judgments from 'plight' to 'polite' in both the hiss and the silent conditions. If, however, hiss is present between burst and voicing onset, the cross-over is realized with a shorter voiced /l/ duration (about 16 ms shorter) than if this interval is silent. This suggests a sonority hierarchy of voicing over hiss over silence.

Experiment 3: Relative versus Absolute Duration in Synthetic Stimuli

Stimuli and Subjects

The third and final experiment to be reported here involves the issue of rate, or absolute versus relative durations. In this experiment two factors are pitted against each other: the absolute duration of the steady-state /l/ and the overall rates of the stimuli. Stimuli similar to those used in the hiss condition of experiment 2, but with larger step sizes, were used in this experiment under four conditions:

(1) Original: stimuli of experiment 2 (hiss condition, /l/ duration = 91 ms) with VOT increased from 42 to 126 ms in 14-ms steps as voicing of the /l/ decreased from 84 to 0 ms;

(2) Extended: /l/ duration of the stimuli augmented by 35 ms (/l/ duration = 126 ms) and VOT increased from 42 to 154 ms in 14-ms steps as voicing of /l/ decreased from 119 to 7 ms;

(3) Fast: stimuli of condition (1) synthesized at a 40% faster rate (/l/ duration = 65 ms), i.e. VOT increased from 30 to 90 ms in 10-ms steps as voicing of /l/ decreased from 60 to 0 ms;

(4) Extended fast: stimuli of condition (2) synthesized at a 40% faster rate (/l/ duration = 90/ms), i.e. VOT increased from 30 to 110 ms in 10-ms steps as voicing of /l/ decreased from 85 to 5 ms.

Conditions (1) and (4) thus represent stimuli with the same absolute /l/ duration (90–91 ms), but synthesized at different rates. Conditions (1) and (3), on the other hand, have the same /l/ durations relative to the duration of the entire stimulus. Likewise, stimuli in conditions (2) and (4) have /l/ durations of the same percentage of overall duration, though the two sets of stimuli are synthesized at different rates. Figure 4 shows spectrograms of the stimuli used for the shortest VOT value

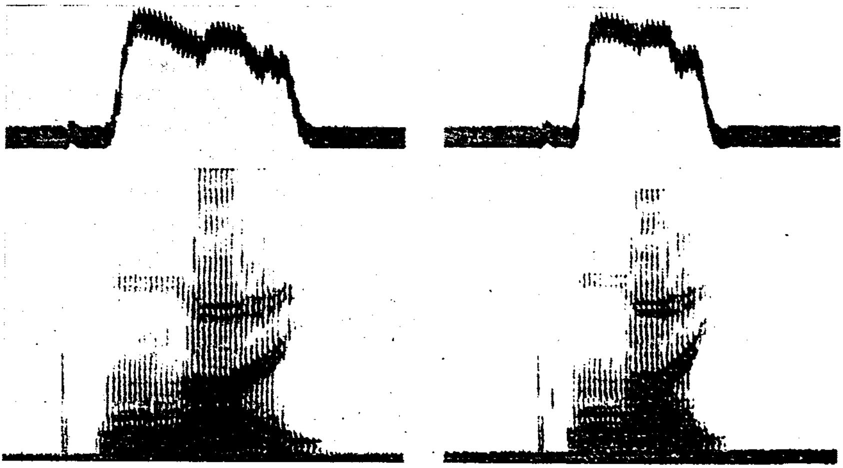

Fig. 4. Spectrograms of synthetic stimuli used in experiment 3. These displays represent the shortest VOT (longest voiced /l/) stimuli used for the extended, and the extended fast conditions.

in the extended and extended fast conditions. The extended conditions included 9 stimuli each; the unextended conditions 7 stimuli each. Three tokens of each of these 32 stimuli were presented to 10 paid volunteers (Yale undergraduades): n = 30 responses per stimulus.

Results and Discussion

Figures 5a, b show 'plight' responses as a function of duration of voiced /l/ expressed as a percentage of overall duration. As in experiment 2, VOT and voiced /l/ duration are inversely correlated: longer VOT values correspond to shorter voiced /l/ durations. Note that for both the original (fig. 5a) and the extended (fig. 5b) stimuli, an increase in overall rate elicits more 'plight' responses. That is, when relative durations are equated, an increase in rate does affect listener judgments. The effect of absolute duration is shown in figure 6. In this figure 'plight' responses are plotted as a function of the absolute duration of the voicing of /l/. It thus appears that, other things being equal, the absolute duration of the voiced portion of the /l/ has a greater effect on listener judgments than its relative duration, at least for the rates and the durations used here. Further research may reveal

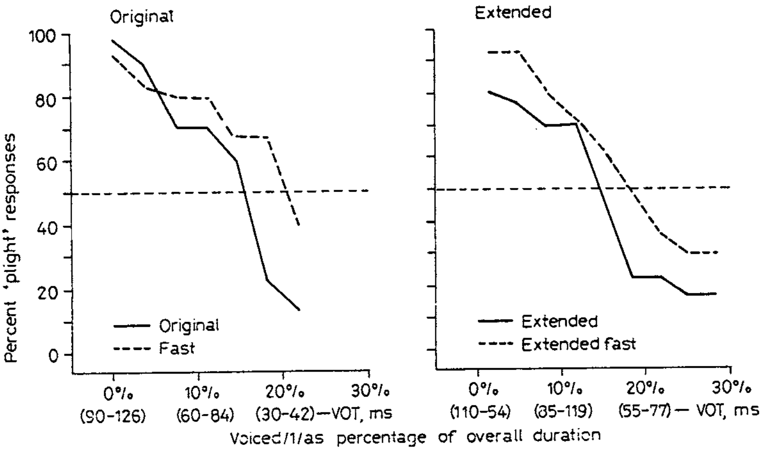

Fig. 5. Relative duration of voiced /l/ (n = 30). Plotted here are 'plight' responses as a function of voiced /l/ duration. (VOT values are in parentheses.) /l/ durations are expressed here in percentages of overall stimulus duration. Note that for the same relative durations (either for original or extended stimuli), the faster overall rates elicit more 'plight' responses.

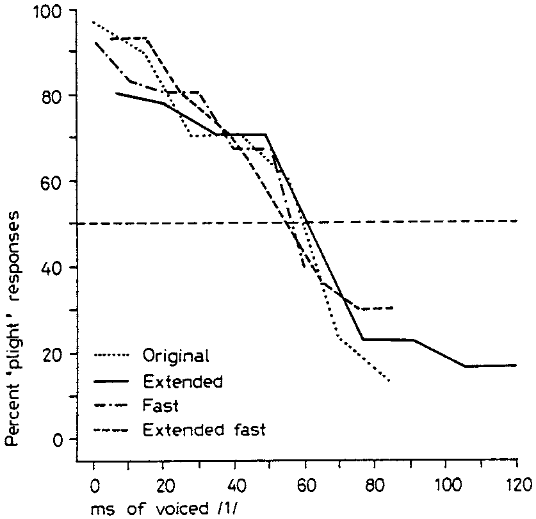

Fig. 6. Absolute duration of voiced /l/ (n = 30). The same data presented in figure 5 are plotted here as a function of absolute /l/ duration. As is seen in this figure, for the /l/ durations and rates used, absolute rather than relative duration seems to be the crucial factor in listener judgments. The cross-over point here is 55-65 ms voiced /l/. This is consistent with the value found in experiment 2.

that for certain ranges of absolute duration, the relative duration of a segment with respect to its surround makes a difference, but these data so far indicate a different story: that absolute duration is a more effective indicator of sonority than is relative duration.

Conclusions

If it is assumed that the auditory term 'sonority' can account for the syllabic versus nonsyllabic distinction of certain segments, then this study provides evidence for acoustic correlates of this term. Sonority, like other perceptual terms such as 'pitch' or 'loudness', has multidimensional acoustic correlates. The experiments presented here bear on the roles of duration, amplitude, voicing, hiss, and silence as they relate to sonority. The results of these experiments support the following hypotheses: (1) duration is a more effective cue to sonority than is amplitude, (2) amplitude may play a role when duration is ambiguous, (3) when duration is manipulated, voiced segments tend to be more sonorant than hiss-excited segments, which in turn appear more sonorant than silence, (4) absolute duration is more important to perceived sonority than relative duration.

Acoustic or auditory correlates have been proposed (but not tested) for the perception of syllabic peaks versus margins. FISCHER-JØRGENSEN [1975] suggested that liquids are auditorily weaker than vowels since most of their energy is concentrated in the first formant. FANT [1969/1973] suggested a weighted sum of the intensities of F_1 and F_2 compared to that in adjoining segments. GAITENBY and MERMELSTEIN's [1977] weighting function, which favors the frequencies between 500 and 4,000 Hz, implies a similar acoustic-auditory emphasis, although in this case the weighting is done in order to analyze syllabic stress rather than internal syllable structure. Left for further research is the implementation of these suggestions in a test of the hypotheses proposed in the present study.

Acknowledgment

I wish to thank LEN KATZ, ANDREA LEVITT, LEIGH LISKER, and MICHAEL STUDDERT-KENNEDY for discussion, criticism and many helpful comments, and BRSG Grant RR–05596 to Haskins Laboratories for support.

References

Bell, A.: If native speakers can't count syllables, what can they do? (Indiana University Linguistics Club, Bloomington, Ind. 1975).

Bell, A.: Segment organization phenomena and their explanations; in Bell, Hooper Syllables and segments (North-Holland, NewYork 1978).

Bell, A.; Hooper, J.: Issues and evidence in syllabic phonology; in Bell, Hooper Syllables and segments (North-Holland, NewYork 1978).

Bloomfield, L.: Language (Holt, NewYork 1933).

Chomsky, N.; Halle, M.: The sound pattern of English (Harper&Row, NewYork 1968).

Delattre, P.: Tendances de coupe syllabique en français, PMLA *55:* 579–595 (1940).

Delattre, P.: L'aperture et la syllabation phonétique; in French and comparative phonetics, pp. 163–167 (Mouton, The Hague 1966). (Originally published in 1944).

Delattre, P.: Change as a correlate of the vowel-consonant distinction. Studia Linguistica *18:* 12–25 (1965).

Fant, G.: Distinctive features and phonetic dimensions. Speech sounds and features (MIT Press, Cambridge, Mass. 1973). (Article first appeared in 1969.)

Fischer-Jørgensen, E.: Trends in phonological theory (Akademisk Forlag, Copenhagen 1975).

Gaitenby, J.; Mermelstein, P.: Acoustic correlates of perceived prominence in unknown utterances. Haskins Lab. Status Rep. Speech Res., SR 49, pp. 201–216 (Haskins Laboratories, New Haven 1977).

Greenberg, J.: Is the vowel-consonant dichotomy universal? Word *18:* 73–81 (1962).

Hjelmslev, L.: The syllable as a structural unit; in Proc. Third ICPS, pp. 266–272 (1338).

Hockett, C.: A system of descriptive phonology. Language *18:* 3–21 (1942).

Hockett, C.: A course in modern linguistics (Macmillan, NewYork 1958).

Hooper, J.: An introduction to natural genaerative phonology (Academic Press, NewYork, 1976).

Jones, D.: The phoneme (Heffer, Cambridge 1950).

Lebrun, Y.: Sur la syllabe, sommet de sonorité. Phonetica *14:* 1–15 (1966).

Lisker, L.; Abramson, A.: A cross-linguistic study of voicing in initial stops. Word *20:* 384–422 (1964).

Mermelstein, P.: Automatic segmentation of speech into syllabic units. J. Acoust. Soc. Am. *58:* 880–883 (1975).

Pike, K.: Phonetics (University of Michigan Press, Ann Arbor, Mich. 1943).

Price, P.: The syllable; MA paper University of Pennsylvania (unpublished, 1978).

Sievers, E.: Grundzüge der Phonetik (Breitkopf & Härtel, Leipzig 1893).

Studdert-Kennedy, M.: Speech perception; in Lass Contemporary issues in experimental phonetics, chapter 8 (Academic Press, NewYork 1976).

Trager, G.; Bloch, B.: The syllabic phonemes of English. Language *17:* 223–246 (1941).

Trager, G.; Smith, H., Jr.: An outline of English structure. Studies in Linguistics, Occasional Papers 3 (Battenburg Press, Norman, Okla. 1951).

Venneman, T.: On the theory of syllabic phonology. Ling. Ber. *18:* 1–18 (1972).

Wedin, S.; Leanderson, R.; Wedin, L.: Evaluation of voice training. Folia phoniat. *30:* 103–112 (1978).

Received: August 4, 1980; accepted: September 25, 1980

P.J. Price, Haskins Laboratories, 270 Crown Street, *New Haven, CT 06510* (USA)