

DURATION VERSUS SPECTRUM IN SWEDISH VOWELS: SOME PERCEPTUAL EXPERIMENTS²

I. Introduction

In many languages the feature of relative vowel length is said to be phonemically distinctive. This would seem to mean that a listener perceives an utterance in one of two or more possible ways, depending on how long he heard a vowel of a certain phonetic quality (timbre) to be sustained.³

A pair of vowels said to be minimally distinguished by length may indeed be essentially the same in vowel quality as far as can be determined by the trained ear or even by instruments.⁴ But there may be some pairs in which a more or less striking difference in vowel quality is associated with length difference (Straka, 1962). Such qualitative differences are usually written off as allophonic, conditioned by the phoneme of length, vowel gemination, or whatever feature the phonemic analysis in question posits. That is, the linguist may be influenced in his choice of phonological features by a desire for simplicity and symmetry of pattern;⁵ nevertheless, there is the possibility that in at least

¹ Also at Fonetiska Institutionen, Lund, Sweden.

² An earlier version of this article was read before the Ninth Annual Conference on Linguistics of the Linguistic Circle of New York, March 14, 1964.

³ For general treatments of this subject, see Malmberg, 1944; Durand, 1946; Fischer-Jørgensen, 1960, pp. 78—79; Lazicius, 1961, pp. 114—134.

⁴ *Quality* is a perceptual attribute; instruments can be used to measure only its major acoustical correlate, the sound *spectrum*. Another terminological distinction to be made is between *duration* and *length*. The duration of a vowel segment is the time it actually lasts; its length is the time it is perceived to last. As in all other psychoacoustic correlations, it is to be understood that a physical change does not necessarily produce a corresponding perceptual change of the same magnitude. See Fry, 1956, especially p. 171.

⁵ For a discussion of the phonemic interpretation of vowel length, see Hockett, 1955, pp. 76—80, and Jones, 1950, chaps. XXII—XXIII.

a portion of such a vowel system, vowel *quality* rather than vowel *length* bears the major communicative burden. It seems to us that in this ambiguous kind of situation, the primacy of one phonetic feature over the other can be established only through an experimental approach.

Swedish is conventionally analyzed as having eighteen vowel phonemes¹ in stressed position, nine of which are long and nine short (e.g., Malmberg, 1956). The minimal contrasts to be accounted for are listed below in fairly narrow IPA transcription together with key words in Swedish orthography.

Long Vowels

Symbol	Key Words
i:	rit, vit
e:	vet, bet
ɛ:	rät, vät
a:	rat-, fal
y:	ryt, byt
ø:	röt, föl
u:	rot, bot
o:	våt, fåt
u:	Rut, ful

Short Vowels

Symbol	Key Words
ɪ	ritt, vitt
e	rett, vett, bett
ɛ	rätt, vätt ²
a	ratt, fall
ɤ	rytt-, bytt
œ	rött, föll
ʊ	rott, bott
ɔ	rått, vått, fått
ʊ	rutt, full ³

There has been much discussion in the Swedish linguistic literature of the relevance of length versus quality,⁴ with the consensus suppor-

¹ Elert, 1964, pp. 42—43, objects that such a view results in an unusually high number of vowel phonemes for a language. Some dialects are said to have seventeen vowels; see footnote 7.

² In many dialects, for instance Central Swedish, there is no distinction between short [e] and short [ɛ], which are both pronounced [ɛ], yielding a system of nine long vowels and eight short ones. In the Stockholm area neutralization also occurs between long [e:] and long [ɛ:], which are both pronounced [e:], except before [r], where the distinction is maintained.

³ The symbols *u* and *ʊ* are taken from the Swedish *landsmålsalfabet* (dialectal alphabet). The vowel [u:] can be described with reference to the IPA Cardinal Vowel [u], which is a close central rounded vowel. The Swedish sound is more fronted and less close and has a characteristic rounding of the lips similar to that for whistling. Its short partner [ʊ] is a central half open vowel, less rounded than [u:]. It is sometimes transcribed with the IPA symbol ɵ.

⁴ See Elert, 1964, for a detailed treatment of the problem, including extensive references to the literature on Swedish and other languages.

ting length. It would seem reasonable, then, to describe the Swedish vowel system as consisting of nine vowels plus a phoneme of length. For at least two reasons, however, this solution is seldom adopted: (1) There is generally something of a qualitative difference between a long vowel and its short counterpart; indeed, in some pairs this is so noticeable and so persistent that it remains when the length difference is neutralized, as may happen in unstressed position.¹ As a result, the transcription used is often a compromise (e.g., Malmberg, 1956), giving some pairs as differing in both quantity and quality and others in quantity alone. The members of the *e*- and *ε*-pairs are said to differ so little in quality that they are distinguished even in rather narrow transcription only by means of a length mark.² As part of the qualitative difference, the long vowels have been described as being more or less diphthongized, especially long close vowels, which may be accompanied by a fricative component (Hammarström and Norman, 1957). Also there is often a weakening of intensity in the latter part of the long vowel. (2) In most dialects long vowels are followed by short consonants and short vowels by long consonants, a fact that is reflected in Swedish orthography and in some transcription practices. The durational difference in consonants, however, is hardly perceptible in some dialects; therefore, it is often considered to be a redundant feature, varying with the duration of the vowel.³

The duration ratios between the long and short members of the Swedish vowel pairs have been calculated. Elert (1964, p. 109) shows the duration of Swedish short vowels to be on the average 65 % of that of the long ones. Hammarström (1952) found a V/V: ratio of 81 % for monosyllabic words and 71 % for disyllabic words. He found, moreover, that when 120 utterances of the words *staka* and *stacka* were measured, 35 % of the stressed vowels overlapped, i.e., they occurred within a boundary zone where their duration alone could not give definite information as to whether they were in fact "long" or "short" (Hammarström, 1956).

¹ Malmberg, 1949, remarks, "La différence de timbre est toujours là pour assurer à la voyelle son caractère de longue ou de brève, même dans les cas où les durées objectives seules ne suffiraient pas."

² We have followed this convention in our list of contrasts; nevertheless, formant-pattern differences typical of each of these two pairs (Fant, 1953, 1959) may be large enough to be audible.

³ For measurement data, see, e.g., Lisker, 1958, pp. 297—299 and Elert, 1964. Witting (1959, p. 111) views as redundant not only this feature but also timbre differences between long and short vowels.

To sum up: Quantity has traditionally been considered a relevant feature of Swedish vowel phonemes. The pairs so distinguished show qualitative differences, some of them somewhat greater than others. Consonants after "short" vowels tend to be long.

II. Experiments

In view of the foregoing considerations, post-vocalic consonant length seemed least promising as a primary differentiating feature; nevertheless, before fastening our attention on vowel length and quality, we ran an experiment to determine whether the feature of consonant length could safely be rejected as a cue. For this purpose, and for use later in the vowel length experiments, words minimally distinguished by all the vowel pairs in question were embedded in carrier sentences and recorded on tape by a male Southern Swede. The southern dialect was used because Southern Swedish subjects were readily available for the perception tests.

Consonant Duration. Of the minimal pairs of words recorded, spectrograms revealed that consonant constrictions or closures were by and large longer after short vowels, although the difference was often slight. There were also a few cases in which the consonant after a short vowel was the same as, or even shorter than, the consonant after the corresponding long vowel. The greatest difference in consonant duration, one of 55 msec, was found in the pair *stöta/stötta* embedded in the carrier frame *Du får — bättre*, yielding, respectively, the sentences 'You must push better' and 'You must prop up better'. The stop closure after the [ø:] of *stöta* was 85 msec long and after the [œ] of *stötta*, 140 msec long.

A tape segment of 70 msec duration was cut from the silent interval corresponding to the closure of the long [t] in *stötta* and introduced into the closure of the short [t] in *stöta*. In this way the original long [t] was shortened to 70 msec and the short [t] lengthened to 155 msec. Two Southern Swedish listeners on whom the stimuli were tested were uninfluenced in their linguistic identifications by the changes made. That is, although the utterance with the lengthened [t] closure sounded "somewhat absent-minded, verging on hesitation", the word was still labeled *stöta*, while the utterance with the shortened [t] closure was said to be "brisk", but the word was labeled *stötta*.

It is still possible, of course, that relative consonant length has at

least redundant cue value in certain contexts, yet the experimental results supported our belief that this feature should be rejected as a primary cue.

Vowel duration. The experimental method used was the tape cutting and splicing technique used previously, for example, in studies of Thai vowel length (Abramson, 1962, chap. II; Bastian and Abramson, 1962). The end points of each vowel were marked in wide-band spectrograms, so that the vowel included both the steady-state portions of the formants and the consonant-vowel formant transitions, but excluded obvious consonant resonances, bursts and friction patches. Successively longer segments of magnetic tape were then cut from the long member of each pair until the shortest variant was at least as short as the original short member. The cuts were made in the central part of the vowel where the formants were relatively steady, leaving the formant transitions intact. The operation of cutting and splicing produced no clicks in the tape recordings. Five tape dubbings of each variant embedded in the carrier sentence, together with five copies of the original long and short versions, also in the carrier sentence, were scrambled to form two test orders which were played to a jury of Southern Swedish students on separate occasions for word identifications. The subjects were provided with a pair of Swedish definitions from which to choose in labeling each stimulus. It seemed desirable to limit the present study to just three vowel pairs, one of which is said to exhibit little or no qualitative difference, another which shows a very striking difference in vowel quality, and a third pair with an intermediate degree of quality difference. In our attempt to find three such representative pairs that were amenable to our experimental design, we were forced to exclude certain of our informant's recordings from consideration. Long close vowels accompanied by friction¹ or weakening of intensity toward the end could not be used because of the difficulty of specifying formant positions and durations in spectrograms. Some other vowel utterances had to be excluded because there was little or no steady state in the central portion; instead, these productions tended to show considerable formant movement. Cutting fairly long segments from such a vowel is likely to produce discontinuities in formant pattern that are great enough to be audible.² The pairs fi-

¹ An additional experiment might well be run in which the presence of a fricative component is pitted against variations in duration.

² In our body of recordings, fundamental frequency changes in the key words were not large enough to cause pitch jumps.

nally chosen were [ɛ:] and [ɛ]; [u:] and [ū]; and [ø:] and [œ]. Since these three pairs seemed representative of the kinds of qualitative statements made about Swedish vowels, it was our aim to test the perceptual relevance of relative vowel duration in each pair.

Because of the foregoing limitations, as well as distributional gaps in the Swedish lexicon, the minimal pair of words recorded in a carrier sentence for each long/short pair of vowels had, of course, a fixed consonantal frame, but not the same consonantal frame as the other two pairs. Although the use of nonsense syllables might have solved this problem, the probable orthographic bias of the Swedish subjects suggested the requirement that they identify the test stimuli as meanings intended by the speaker rather than as spelled forms. In any event there was no reason to believe that such restrictions in the design of our present pilot experiments would prevent us from learning, to a first approximation, whether there was any trading relation between vowel quality and vowel length in the pairs of Swedish vowels traditionally said to be distinguished by relative length. Should the first stage of our research be productive, it would then be possible to prepare a more complicated battery of tests that would take all major contexts into account.

Measurements of the durations of the six vowels used were made on wide-band spectrograms and, as a reasonable estimate of precision, rounded to the nearest five milliseconds.¹ Our intention had been to reduce the duration of each long vowel in equal steps; however, measurements of the variants revealed a slight departure from linearity here and there. The durations of the vocalic portions of the test stimuli are given in the descriptions of the three experiments. Since we wished to vary only the physical factors that gave rise to an impression of length, but not quality, it was necessary to examine the formant patterns of the vowels (see f.n. 2). Unfortunately, determining formant frequencies from spectrograms is not an easy task (Lindblom, 1962). In addition to comparing wide-band spectrograms and narrow-band spectrograms,² we made closely spaced narrow-band sections from the beginning to the end of the formant pattern for each vowel. Our formant frequency values are derived from spectral peaks found by care-

¹ The sound spectrograph used was the Sona-Graph of the Kay Electric Company, Pine Brook, New Jersey. Precision of measurement was improved by substituting a large drum for the standard one to provide an expanded time scale on the paper.

² A frequency-scale magnifier on our spectrograph was helpful.

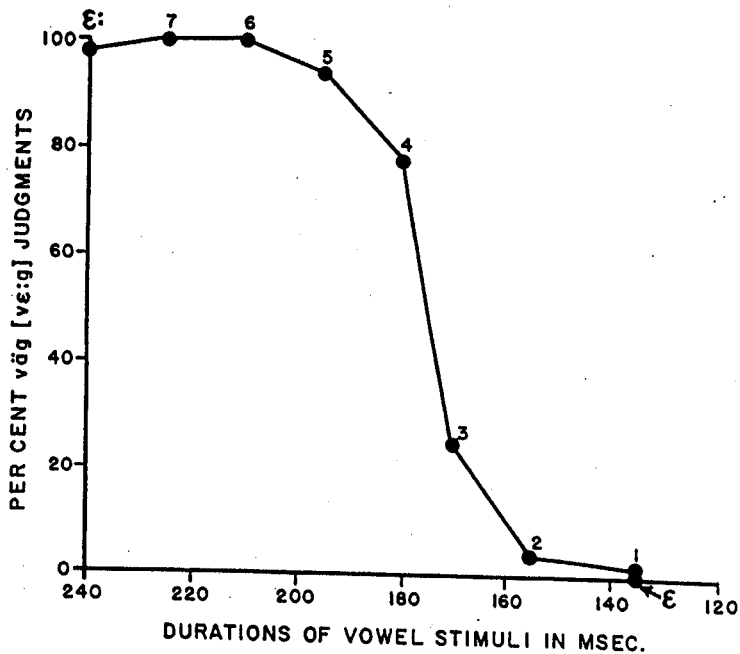


Fig. 1. Stimuli: Original *væg*, seven durational variants of *væg*, and original *vægg*. Judgments as *væg* (130 responses to each stimulus).

ful examination of the three types of instrumental display. These numbers were rounded to the nearest ten cps and should give a fairly good notion of the degree of spectral similarity between the two members of each vowel pair. It stands to reason, of course, that single tokens of each vowel uttered by a single speaker in a fixed frame are not likely to match previously published average formant frequencies derived from larger samplings (Malmberg, 1956; Fant, 1959), but in this study we are concerned only with the formants of the vowels manipulated by us.

Experiment 1.

Carrier sentence: Vår ___ lutar utåt. 'Our ___ slopes outward.'

Vowel durations: [ε:] in *væg* 'road' = 240 msec.

[ε] in *vægg* 'wall' = 135 msec.

V/V: ratio = 56 %.

These vowels were chosen to exemplify pairs in which little or no qualitative difference is heard. The vowel in *våg* was gradually shortened in seven steps, numbered serially from the shortest to the longest. We tried to make each step 15 msec long, but spectrographic checking showed that the step between variants 3 and 4 was only 10 msec long and the step between 1 and 2, 20 msec long. Subsequent test results indicated that these slight non-linearities along the physical continuum offered no compelling reason for making fresh stimuli. The shortest variant in the series was of the same duration as original short [ɛ].

Two random arrangements of these stimuli were played to thirteen subjects for labeling as 'road' or 'wall'. There were 130 responses to each stimulus. Figure 1 shows the percentage of responses to each stimulus as *våg*, the word with the long vowel. Points marking the identifications of the original members of the pair are marked by the vowel symbols $\epsilon:$ and ϵ . Between variants 3 and 4, though the difference between them was only 10 msec, the curve falls steeply from 79 % *våg*-responses to 25 %. It seems clear that length is the main cue here.

The following frequencies, rounded to the nearest 10 cps, were obtained for the first three formants at a point halfway through each original vowel of the pair:

	F1	F2	F3
	(In cps)		
$v\epsilon:g$	460	1960	2400
$v\epsilon g$	470	1920	2400

These differences leave very little room for phoneme identifications that depend on formant patterns (Flanagan, 1955). Of the two physical features under consideration, then, it is not surprising that relative duration dominates.

Experiment 2.

Carrier sentence: Det är — som du hör. 'It is — that you hear.'

Vowel durations: [u:] in *ful* 'ugly' = 175 msec.

[ü] in *full* 'full' = 140 msec.

V/V: ratio = 80 %.

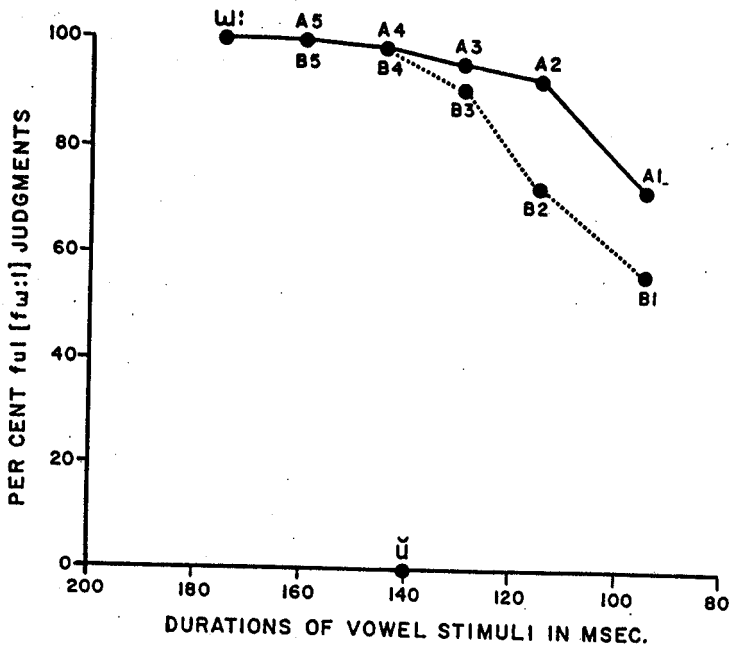


Fig. 2. Stimuli: Original *ful*, five pairs of durational variants of *ful*, and original *full*. Judgments as *ful* (130 responses to each stimulus).

These vowels were chosen to exemplify pairs in which a rather striking qualitative difference is heard. In approximately the second half of the vowel in *ful*, the gross spectrographic pattern suggested the diphthongization and rather weak intensity sometimes described for Swedish long close vowels. We were interested in seeing which of the two parts provided the stronger cue to long [u:], the fairly steady formants of the first part of the vowel or the moving formants of comparatively weak intensity of the second part. We, therefore, made two series of cuts, one from the beginning of the vowel to the middle, and one from the end back to the middle, leaving formant transitions intact. In each case five pieces of tape, increasing in duration in steps of 15 msec, were removed. Three of the variants of each series were shorter than the original short vowel of *full*.

Fifteen subjects listening to one test order and eleven of them to the other yielded 130 responses to each stimulus. In Figure 2 the series which left the fairly steady portion uncut is labeled A. 1—5, while the series

which left the diphthongal glide uncut is labeled B. 1—5. At the appropriate points on the time axis, the percentage of responses to each stimulus as *ful*, the word with the long vowel, is given. The vowel symbols indicate the responses to the two original productions. It is evident that duration is of little importance in this vowel pair. Variants equal to, or a little shorter than original [ū] are heard as [u:] nearly 100 % of the time. Even the shortest variant, 95 msec, was heard by the majority of the listeners as *ful*. In addition, a higher number of *ful* responses is given by the A-series, in which the diphthongized portion of the vowel is shortened. In our experiment, then, the relatively steady portion of the vowel is slightly superior to the rather unsteady part as a cue to [u:]. The latter finding was somewhat surprising, since the off-glide has been considered an important characteristic of long, close Swedish vowels.

In measuring the frequencies of the first three formants of this pair of vowels, it was necessary to sample the long vowel at four points in time to represent the formant movement adequately; these points are labeled as percentages of the duration of the vowel. One point was enough for the short vowel.

	% Duration	F1	F2	F3
		(In cps)		
fu:l	25	360	1650	2460
	50	360	1560	2350
	65	270	1510	2120
	85	270	1390	2020
fūl	50	410	1270	2140

It is clear from these figures that the formant configuration of the second half of the long vowel is not dramatically more dynamic than that of the first half; nevertheless, the gross impression that motivated the making of two series of test stimuli seems to be supported. In the light of experimental findings that assign little or no perceptual importance to relative duration, these formant frequencies leave the way open to the interpretation that vowel quality is the dominant feature for the listener. Identifications that depend on formant patterns are made possible not only by the overall differences between the formant frequencies of *ful* and *full* but also by the moving formants of *ful*.

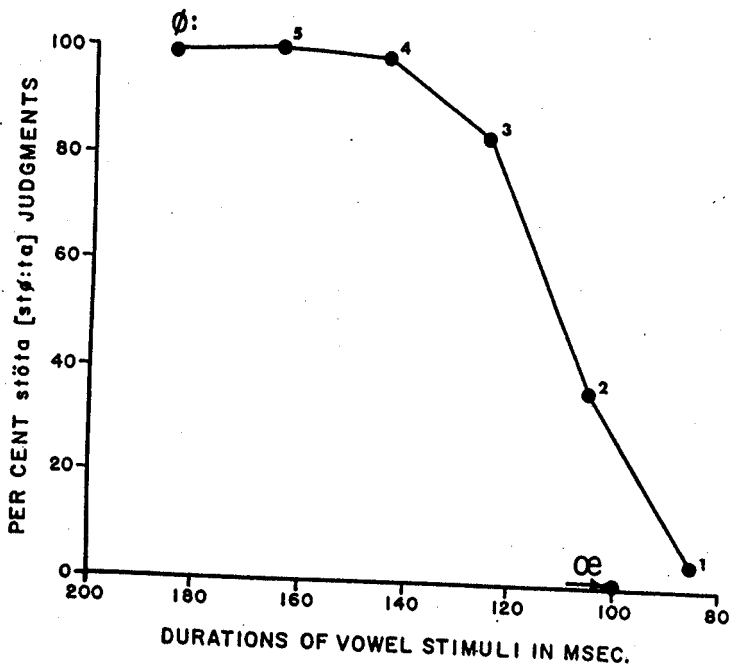


Fig. 3. Stimuli: Original *stöta*, five durational variants of *stöta*, and original *stötta*. Judgments as *stöta* (135 responses to each stimulus).

Experiment 3.

Carrier sentence: Du får — bättre, 'You must — better.'

Vowel durations: [ø:] in *stöta* 'push' = 185 msec.

[œ] in *stötta* 'prop up' = 100 msec.

V/V: ratio = 54 %.

These vowels were chosen to exemplify pairs in which a moderate but clear qualitative difference is heard, intermediate between the types used in Experiments 1 and 2. The vowel in *stöta* was gradually shortened in five steps of 20 msec each. The shortest of these variants was 85 msec long, that is, 15 msec shorter than the original short member of the pair.

Fifteen subjects listening to one test order and twelve of them to the other yielded 135 responses to each stimulus. Figure 3 shows the percentage of responses to each stimulus as *stöta*, the word with the long vowel. The vowel symbols indicate the responses to the two origi-

nal productions. Between variants 2 and 3 the curve falls from 84 % to 38 %, demonstrating that length is the main cue here. There seems, however, to be a small effect that can be attributed to the qualitative difference between the two members of the pair. Variant 2 is only five msec longer than the short vowel of original *stötta*, yet it is labeled *stöta* 38 % of the time. Indeed, even variant 1 is called *stöta* nearly 4 % more often than is *stötta*, although the latter is 15 msec longer; the margin is slim for significance, but taken together with the responses to variant 2, it is suggestive.

The following frequencies were obtained for the first three formants at a point halfway through each original vowel of the pair:

	F1	F2	F3
	(In cps)		
stø:ta	400	1590	2380
stœta	530	1590	2280

In addition to the durational distinction, then, there is a spectral difference that seems to be large enough to account for the moderate qualitative difference between the two vowels and for the small perceptual effect of this vowel quality that is evident in Figure 3.

III. Conclusion

As far as it has gone, the present study supports the traditional view that length is a distinctive feature of Swedish vowels, although it obviously does not pervade the whole vowel system. For one pair, at least, Experiment 2 does not contradict the hypothesis that quality (timbre) rather than length is distinctive.¹ Indeed, even in Experiment 3 in which relative length is dominant, quality seems to have some discrimi-

¹ It is noteworthy that in Experiment 2 the V/V: ratio is 80 %, while in Experiments 1 and 3 the ratios are 56 % and 54 % respectively. Ewert finds (1964, p. 113) that in all types of material used in his investigation the ratio of [ü]:[u:] is higher than that of all other vowel pairs and seeks an explanation in the fact that the short vowel "has a considerably more open and back articulation and, hence, greater intrinsic duration ...". We might also speculate that, whatever the historical situation in Swedish was, as cue value gradually shifted from length to quality in this pair, the constraint upon speakers to maintain a clear durational difference lessened.

minative effect. These results favor accounting for the maximal set of contrasts presented in the Introduction in terms of a subset of eight vowel phonemes that combine with a phoneme of length to yield long counterparts and another subset of two vowel phonemes, / \ddot{u} / and / u /, distinguished by quality.

Our plan is to use this tentative restatement of the Swedish vowel system as a working hypothesis for further work in two directions: (1) Similar tape-cutting experiments will be run for the vowel pairs of this study, as well as others, in a wider variety of consonantal environments. (2) Techniques of speech synthesis will be used to vary both the spectral configurations and the durations of vowels to assess their relative contributions to word intelligibility in Swedish.

Haskins Laboratories,
New York

References

- ABRAMSON, A. S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments*. (Bloomington).
- BASTIAN, J. and ABRAMSON, A. S. (1962). Identification and discrimination of phonemic vowel duration. *J. Acoust. Soc. Amer.* 34, 743—744(A).
- DURAND, M. (1946). *Voyelles longues et voyelles brèves: Essai sur la nature de la quantité vocalique*. (Paris).
- ELERT, C. C. (1964). *Phonologic Studies of Quantity in Swedish*. (Uppsala).
- FANT, C. G. (1953). Speech communication research. *IVA 24* (Royal Swedish Acad. of Engineering Sci.), 331—337.
- (1959). Acoustic analysis and synthesis of speech with applications to Swedish. *Ericsson Technics* 1, 3—108.
- FISCHER-JØRGENSEN, E. (1960). *Almen Fonetik* 3rd ed. (Copenhagen).
- FLANAGAN, J. L. (1955). A difference limen for vowel formant frequency. *J. Acoust. Soc. Amer.* 27, 613—617 (Letter).
- FRY, D. B. (1956). Perception and recognition in speech. *For Roman Jakobson* (The Hague), pp. 169—173.
- HAMMARSTRÖM, G. (1952). Sur la durée des phonèmes en suédois. *Revista do Laboratório de fonética experimental* (Coimbra) 1, 9—27.
- (1956). Problèmes phonométriques, et autres, concernant la durée en suédois. *Rev. do Lab. de fonética experimental* (Coimbra) 3, 17—37.
- HAMMARSTRÖM, G. and NORMAN, L. (1957). Om den frikativa slutfasen vid de svenska långa vokalerna i, y, u, \ddot{u} . *Nordisk tidskrift for tale og stemme* 17, 89—98.
- HOCKETT, C. F. (1955). *A Manual of Phonology*. (Baltimore).
- JONES, D. (1950). *The Phoneme*. (Cambridge).
- LAZICIUS, J. (1961). *Lehrbuch der Phonetik*. (Berlin).

DURATION VERSUS SPECTRUM IN SWEDISH VOWELS

- LINDBLOM, B. (1962). Accuracy and limitations of Sona-Graph measurements. *Proc. Fourth International Cong. Phon. Sci.* (The Hague), pp. 188—200.
- LISKER, L. (1958). The Tamil occlusives: short vs long or voiced vs voiceless? *Indian Linguistics*, Turner Jubilee Vol. I, 294—301.
- MALMBERG, B. (1944). Die Quantität als phonetisch-phonologischer Begriff. *Lunds Universitets Årsskrift* (Lund), pp. 3—103.
- (1949). Review of M. Durand, Voyelles longues et voyelles brèves in *Studia Linguistica* 3, 39—59.
- (1956). Distinctive features of Swedish vowels: some instrumental and structural data. *For Roman Jakobson* (The Hague), pp. 316—321.
- STRAKA, G. (1962). Durée et timbre vocalique. *Z. für Phonetik und allgemeine Sprachwissenschaft* 12, 276—300.
- WITTING, C. (1959). *Physical and Functional Aspects of Speech Sounds* (Uppsala).