Production and perception of vowel length in spoken sentences

Citation for published version (APA):

Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, 67(1), 276-287. <https://doi.org/10.1121/1.383737>

DOI:

[10.1121/1.383737](https://doi.org/10.1121/1.383737)

Document status and date:

Published: 01/01/1980

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Production and perception of vowel length in spoken sentences

S. G. Nooteboom^{a)} and G. J. N. Doodeman

Institute for Perception Research, Eindhoven, The Netherlands

(Received 11 January 1979; accepted for publication 11 September 1979)

A series of experiments was conducted to determine (1) the accuracy with which vowel segment durations in spoken sentences can be represented in auditory sensory storage and (2) the extent to which phoneme boundaries in the identification of phonemic vowel length in Dutch are affected by syntactic and/or auditory-phonetic context. In a preliminary production test it was found that durations of both long and short Dutch vowel phonemes in monosyllabic words embedded in sentences are systematically affected by word positions in the sentences. In an initial perceptual experiment, phoneme boundaries and slopes of identification curves were measured for 12 listeners in five different test utterances in a binary forced choice identification test. Perceptual accuracy of vowel duration perception as determined from the slopes of the identification curves corresponds on the average to a just-noticeable difference (JND) of about 5 ms with a test segment duration of about 90 ms. Phoneme boundary values are systematically affected by context in ways predictable from syntactic structure and the auditory-phonetic environment. In a second perceptual experiment it is shown that a major effect on phoneme boundary can be brought about by perceived properties of the test utterance following the monosyllable containing the test segment. In a third perceptual experiment it is shown that the difference between phoneme boundaries in utterance final syllable and in embedded syllable is related to the presence or absence of a perceived speech pause following that syllable. The results of these experiments are interpreted in terms of a simple decision model with a noisy auditory representation of embedded vowel duration, lasting a few hundredths of milliseconds, and a noiseless internal criterion for vowel length identification which is systematically affected by the auditory-phonetic environment.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

What is the accuracy with which acoustic durations of vowel segments in spoken sentences are auditorily represented? How is the internal criterion for distinguishing between phonemically short and phonemically long vowels on the basis of acoustic vowel duration affected by other perceived properties such as the syntactic structure and the specific auditory-phonetic structure, of the sentences these vowels belong to? Those two questions form the subject matter of the present paper.

A measure of the accuracy with which a particular stimulus property is auditorily represented during a shorter or longer period of time can be the smallest difference in the physical magnitude of the stimulus property that can be perceived (e.g., at a 75% probability level). In classical psychophysics this measure is often obtained by using well-practiced observers and a very simple comparison task, generally not going beyond a two-interval binary forced-choice comparison (2I2AFC) task. Such a task, however, is not applicable to properties of speech sounds embedded in longer utterances, because of overloading of sensory storage by the surrounding speech material (cf. Carlson and Granström, 1975). This methodological problem can be circumvented by replacing the external reference signal used in a 2I2AFC task by an internal criterion which is constantly available to the observer. Huggins (1972), for instance, had his listeners use an internal criterion of "normal" versus "long," or "normal" versus "short" in measuring JND's of speech sound durations in spoken

sentences. Klatt and Cooper (1975) used a magnitude estimation technique, in which listeners had to compare embedded segment durations with each of eight response criterion thresholds derived from mapping their preferred duration on the middle of an arbitrary nine-point scale. In the present experiments we have based our measurement of the accuracy of auditory representation of vowel duration on the listeners' internal criterion for identification of short versus long vowel phonemes, the advantage being that this criterion, being overlearned, is potentially stable, and that the experimental procedure closely resembles the classical 2I2AFC test, except that a single-interval task is used instead of a two-interval task. This makes it feasible to compare our identification data with data on sound duration discrimination in the literature. A similar procedure is reported by Fujisaki, Nakamura, and Imoto (1975).

Our second question concerns potential effects of properties of the surrounding speech material on the internal criterion for identifying short and long vowel phonemes on the basis of acoustic vowel duration. A physical measure of this internal criterion is the stimulus duration at which 50% short vowel responses and 50% long vowel responses are given by the listeners. This we call the phoneme boundary of vowel length perception. We have tested the hypothesis that the internal criterion, measured as the phoneme boundary, is systematically affected by the expected or preferred durations which the listeners can create for both the short and the long vowel, on the basis of syntactic structure, and possibly the specific auditory-phonetic structure of the attended utterance. We have not measured expected durations directly. Results reported by Huggins (1972), Nooteboom (1973), Klatt and Cooper (1975), and Fujisaki, Nakamura, and Imoto (1975) show that there is

^{a)}Also at the Department of General Linguistics, University of Leyden, The Netherlands.

a close correspondence between the expected duration of a particular segment in an utterance, and the duration actually produced by a speaker in that position. For a survey of acoustic, physiological, and perceptual evidence concerning the effects of a number of linguistic and extralinguistic factors determining speech segment durations, the reader is referred to Klatt (1976). The perceptual relevance of syntactic effects on speech segment durations is demonstrated by Lehiste, Olive, and Streeter (1976) for English and by De Rooij (1979) for Dutch.

In order to study to what extent context effects on the internal criterion for vowel length identification can be explained by syntactic structure alone, we have collected data on the production of short and long vowel durations by several speakers in the same frame sentences to be used in the identification experiment. This production experiment is reported in Sec. I. Section II is devoted to the main perceptual experiment, in which we have measured identification functions of short versus long vowel identification in five different positions in spoken sentences. The results of this experiment, apart from providing data on the accuracy of auditory representation of vowel duration, demonstrate that phoneme boundaries are a function of syntactic structure. These results do not allow us to decide whether the effect of syntactic structure on phoneme boundaries is mediated by the perceptual results of the specific auditory-phonetic structure of the attended utterances, or, rather, is an immediate effect of the perceived syntactic structure. The perceptual experiments in Secs. III and IV demonstrate that phoneme boundaries can be shifted in a systematic and predictable way by perceived properties of the speech material immediately following the test segment in the utterance, without changing the perceived syntactic structure. The combined results of all experiments will be discussed in Sec. V.

I. PRODUCTION EXPERIMENT

A. Method

1. Sentence material

As the test vowels we used two vowel phonemes, the Dutch /a/, which is a phonemically short vowel, and the Dutch /a:/ which is a phonemically long vowel. These vowels were always preceded by /t/ and followed by /k/, thus forming either the Dutch monosyllabic word TAK (English *branch*) or the Dutch monosyllabic word TAAK (English *task*). These two words will be referred to throughout as the test words. The Dutch vowels /a/ and /a:/ differ acoustically not only in duration, but also somewhat in spectral composition (average values of 50 male speakers of F_1 , F_2 , and F_3 are 679, 1051, 2619 Hz for /a/ and 795, 1301, 2565 Hz for /a:/; cf. Pols, Tromp, and Plomp, 1973). However, if a Dutch long /a:/ is acoustically shortened it is possible to obtain 100% /a/ perceptions. (On the other hand, by lengthening a natural Dutch /a/ one does not obtain consistent /a:/ perceptions for all listeners. One reason for this asymmetry may be that possibly the relative contribution of spectral composition to /a/-versus /a:/

perception rapidly increases with increasing vowel duration.)

Five frame sentences were chosen such that the test words occurred in different syntactic positions. These frame sentences, together with the syntactic boundary types following the test words and the English translations of the sentences, are presented in Table I.

The diagrams of the syntactic surface structures are presented in Fig. 1. The most inclusive nodes of the word groups on both sides of the boundary following the test words are encircled. In sentence 1 there is only a preceding node, covering the whole sentence. If we assume that the degree of final lengthening is a simple function of hierarchical importance of the most inclusive nodes surrounding the boundary, ($\neq$) being more important than (S), we may expect to find decreasing durations of the vowels of TAK and TAAK from sentence 1–4, sentence 5 giving the same durations as 4.

Of course other predictions could be made on the basis of "branching depths" (Cooper, Lapointe, and Paccia, 1977), or prosodic variables such as the number of syllables or words following in the phrase (Lindblom and Rapp, 1973; Nooteboom and Cohen, 1975). In all cases we expect to find extensive differences due to differences in position of the word in the sentence. To check this assumption, and to obtain durations of /a/ and /a:/ in the various positions which can be compared to later perceptual measurements, we collected some production data.

2. Subjects

Four male Dutch speaking students of the Eindhoven University of Technology served as paid volunteers in the production experiments. All were in their early twenties, were linguistically and phonetically naïve and had no speech or hearing impairments.

3. Procedure

The speakers were tested individually in a sound-insulated booth. Each speaker was presented with a randomized list of sentences in which the ten sentences (five sentence types factorially combined with two test words) occurred five times each. Each speaker was instructed to read each sentence first to himself and then to read it aloud with a natural intonation and tempo. If dissatisfied with his performance he could repeat the sentence. The utterances were recorded by means of a Philips dynamic microphone (LBB9050/05) and a Revox A77 tape recorder. From the tape recordings oscillograms were made with a Honeywell "visicorder", type 3508 with a Honeywell accudata preamplifier. Speed was 100 ms/cm. In the oscillograms the beginnings and ends of the vowels of TAK and TAAK were in nearly all cases determined from the steepest changes in the amplitude envelope. It was only if the words were in sentence final position, that this criterion was not always applicable to the vowel segment ending and we then determined vowel ending from the last glottal pulse visible in the oscillogram. Measurements were generally accurate to the nearest glottal pulse (± 4 ms).

TABLE I. Sentences used in the production experiment. The two alternative test words are capitalized. Syntactic boundary types immediately following the test words are given in brackets following each sentence. The English translation is given below each sentence.

(1) Kees kreeg een nieuwe TAK/TAAK. (Kees got a new branch/task.)	(S-##)
(2) Die nieuwe TAK/TAAK zei Kees is te groot. (That new branch/task said Kees is too big.)	(NP-S)
(3) Die nieuwe TAK/TAAK is te groot voor Kees. (That new branch/task is too big for Kees.)	(NP-VP)
(4) Die nieuwe TAK/TAAK voor Kees is te groot. (That new branch/task for Kees is too big.)	(NP-PP)
(5) Kees kreeg een nieuwe TAK/TAAK op zijn schouders. (Kees got a new branch/task on his shoulders.)	(NP-PP)

B. Results and discussion

A two-way analysis of variance showed a significant effect of sentence types on the produced durations of the vowels in both TAK ($p < 0.001$) and TAAK ($p < 0.001$). In both cases there was also a significant effect of speakers ($p < 0.001$). Only with TAAK a significant interaction between sentence types and subjects was found ($p < 0.01$). To reduce the effect of different speakers we normalized the data for the vowels in TAK and TAAK separately by multiplying each individual vowel duration

by the grand mean of all vowel durations divided by the mean of all vowel durations of the individual speaker. Normalizing factors varied from 0.84 to 1.13. We will restrict our analysis to the normalized data. Mean durations of vowels in TAK and TAAK for the five sentence types, and their standard deviations, are presented in Table II.

As expected, we find systematic differences in durations of both the vowel in TAK and the vowel in TAAK, in different syntactic positions. The differences cannot, however, be easily explained from the hierarchy of syntactic boundary types or from the branching depth hypothesis of Cooper, Lapointe, and Paccia (1977). When listening to the utterances we have the impression that sentence types (2) and (4) are generally produced with the same prosodic structure, with no clear prosodic boundary after the test word. Such a prosodic boundary would be perfectly acceptable in (2), but not, according to our intuition in (4). Speakers have apparently optionally deleted the prosodic boundary at the beginning of an S-dominated structure in sentence type (4). Final lengthening is preserved at the (NP-VP) boundary in (3). The difference between (4) and (5) can be explained if we assume that the vowel is shortened as a function of the number of unstressed syllables following the test word. In (2) and (4) there is one unstressed syllable before the next stressed one, in (5) there are two. In (3) there are also two unstressed syllables following the test word, but here shortening would be blocked by the (NV-VP) boundary (cf. Huggins, 1975; Cooper, Lapointe, and Paccia, 1977).

Can we now predict a listener's expected durations

TABLE II. Mean durations and standard deviations of the vowels in TAK and TAAK in five sentence types spoken by four speakers five times each. Syntactic boundary types immediately following the test words are indicated. The following pairs of sentence types differ significantly (student Newman Keuls, following a two-way analysis of variance $p < 0.05$): TAK, 1-5; 3-5; TAAK, 1-5; 1-4; 1-3; 1-2; 2-5; 3-5; 4-5; 3-4.

	TAK		TAAK	
	$\bar{m}$	sd	$\bar{m}$	sd
(1) (S-##)	91 ms	9 ms	182 ms	12 ms
(2) (NP-S)	82 ms	4 ms	163 ms	12 ms
(3) (NP-VP)	90 ms	9 ms	169 ms	11 ms
(4) (NP-PP)	82 ms	8 ms	161 ms	13 ms
(5) (NP-PP)	76 ms	5 ms	148 ms	8 ms

FIG. 1. Diagrams of the syntactic surface structures of the sentences used in the production tests. Test words are capitalized, the most inclusive nodes of the word groups on both sides of the boundary following the test words are encircled.

from these production data? It should be noted that, although on the average the data follow a nice pattern which is the same for /TAK/ and /TAAK/ the variation in the data is such that individual vowel durations are hardly predictable from the syntactic structure of the sentence alone. Even the mean vowel durations of individual subjects are not fully predictable, nor do the mean vowel durations of TAK and TAAK follow exactly the same pattern for each subject. This may be seen from Fig. 2, where we have plotted for each subject per sentence type the mean vowel duration in TAK against the mean vowel duration in TAAK. Although on the average the vowel duration in TAAK is almost exactly twice the vowel duration in TAK, the relation is not a monotonous function for individual subjects.

We may conclude that individual vowel durations may be predictable from syntactic structure in a probabilistic way, but not with any degree of precision for a particular utterance. It remains to be seen whether listeners derive their expected vowel durations from syntactic structure or from other properties of utterances. Of course the relevance of this question depends on the accuracy with which a listener's expectations can have consequences for his perceptions.

II. PERCEPTION EXPERIMENT I

A. Method

1. Stimulus preparation

One token of each of the five sentence types described in Sec. II, containing the test word TAAK, was recorded, as spoken by the first author, via a Philips LBB9050/05 microphone and a REVOX A77 tape recorder, on magnetic tape. The recording took place in a sound-insulated booth. Each sentence was spoken in a natural voice with rather unobtrusive intonation. Spectrograms (Kay Sonograph M7029A) of the utterances are presented in Fig. 3. The waveforms of the utterances were digitized and stored in 8-bit words on disk with a sampling rate of 10 kHz. Both before digitizing and upon playback the signals were low-pass filtered with a cutoff frequency at 4800 Hz to prevent the intrusion of alias components. In all utterances the vowel in TAAK together with the preceding /t/ plosive was removed with

FIG. 2. Mean duration of vowel in TAK against mean duration of vowel in TAAK for five sentence types and four subjects separately. The numbers in the graph refer to the numbers of the test sentences, different symbols refer to different subjects. $r = 0.8$; the regression line is given by $y = 0.39x + 21$; $S_y = 3$; $S_B = 0.01$ (cf. footnote 1).

FIG. 3. Wide-band spectrograms of the original recordings of the five spoken sentences from which the test utterances in perceptual experiment I were obtained. The words containing the vowel segments to be removed and replaced by test vowel segments of varying durations are indicated.

the help of a computer facility for editing the speech wave form (Willems and De Jong, 1974). The removed acoustic segment was then replaced by one of seven segments, all obtained from the same /ta:/ segment selected from one of the original utterances. This segment contained a /t/ noise burst and four pitch periods of the original /a:/ sound, thus preserving the formant transitions. The fourth period was repeated as many times as necessary to obtain the required vowel segment durations. During the last 10 ms before the required duration was reached the vowel sound was multiplied by an amplitude decay function to avoid audible spectral broadening. The intensity level of the homogeneous parts of all segments thus obtained was fixed at a value which was approximately the mean of intensity levels of the five original test vowels in the five utterances. Formant frequencies and bandwidths in hertz are F_1 766, B_1 95; F_2 1281, B_2 196; F_3 2381, B_3 239; F_4 3382, B_4 647; F_5 4462, B_5 657.

Waveforms of the shortest and longest test segments are presented in Fig. 4. Durations of the test segments used are 66, 75, 84, 93, 102, 111, and 120 ms. By fac-

FIG. 4. Waveforms of both the shortest and the longest of the seven test vowel segments to be inserted in the appropriate slots in the five test utterances.

torial combination of these seven test segments and the five utterances types 35 stimuli were obtained. A stimulus tape was prepared with ten occurrences of each stimulus. Stimuli were separated by silent intervals of 3 s. The first half of the tape contained a randomized order of five occurrences of each stimulus. The second half of the tape contained the same stimulus tokens in reversed order. This stimulus sequence proper was preceded by five more stimuli with easily identifiable vowel segments taken from the extreme ends of the range of vowel durations used in the experiment.

2. Subjects

Ten male Dutch-speaking students of the Eindhoven University of Technology served as paid volunteers in the listening experiment. They were linguistically and phonetically naïve. The two authors also served as subjects. None of the subjects had a speech or hearing impairment.

3. Procedure

Subjects were tested individually in a quiet room. The stimulus tape was played to the subject from a Philips tape recorder type N4510 through ear phones (Sennheiser HD424) at a comfortable loudness level of approximately 55 dB SPL. The subject had two push buttons in front of him, one push button marked TAK, the other TAAK. He was instructed to push the TAK button whenever he heard TAK, and the TAAK button whenever he heard TAAK. When in doubt he was to guess. Preceding each stimulus on the tape, a brief pulse train was recorded which was used to sensitize a measuring circuit during 3 s to signals from the push buttons. The pulse train itself was suppressed before the signal from the tape was fed to the ear phones. Which push button

had been selected and (in order to measure response latencies) the moment of response with respect to stimulus onset, were automatically detected and stored in punched tape together with stimulus number, for later processing.

B. Results and discussion

As there were found to be no systematic differences between the naïve and the non-naïve subjects, results will be reported for all 12 subjects. For each of the 12 subjects and each of the five test utterance types the data consist of a distribution of /a/ and /a:/ responses over ten trials on each of seven segment durations. Because of the binary forced choice nature of the task distributions of /a/ and /a:/ responses are complementary. We analyzed the distributions of /a/ responses only. Phoneme boundaries, defined as the estimated test segment durations giving rise to 50% /a/ responses, were computed for each test utterance individually by fitting a cumulative normal distribution to the data, without weighting, and including all 100% and 0% values. In three cases there were no data points between 100% and 0%, and we changed the 100% and 0% values closest to the phoneme boundary into 99% and 1%, respectively. In Table III the mean estimated phoneme boundaries are presented together with the mean estimated slopes defined as the standard deviations of the cumulative normal distributions. Standard deviations of phoneme boundaries and slopes over 12 subjects are also given.

1. Accuracy of auditory representation

The mean slopes presented in Table III may be taken as a measure of the accuracy with which durations of vowel segments, embedded in spoken sentences, are auditorily represented. However, as in many cases there are very few data points between 100% and 0%, the estimates of the slopes for individual subjects may often be imprecise. We therefore pooled the data of all subjects and all test utterances, and removed the effects of differences in phoneme boundaries between subjects and between test utterances by multiplying each test segment duration for each subject-test utterance combination by the grand mean of all phoneme boundaries (89 ms) divided by the phoneme boundary found for that particular subject-test utterance combination. The individual scores of all subjects and all test utterances

TABLE III. Mean phoneme boundaries and slopes (expressed in the standard deviations of the cumulative normal distributions used to fit the data) and the corresponding standard deviations around the means over 12 subjects. The following pairs of test utterances differ significantly (student Newman Keuls, following a two-way analysis of variance, $p < 0.01$): 1-5; 1-4; 1-3; 1-2; 2-5; 2-4; 2-3; 3-5; 4-5.

Test utterance	Phoneme boundary		Slope	
	$\bar{m}$	sd	$\bar{m}$	sd
1 (S-##)	100 ms	8.4 ms	5.7	2.2
2 (NP-S)	94 ms	4 ms	5.2	2.9
3 (NP-VP)	88 ms	3.8 ms	5.4	2.6
4 (NP-PP)	88 ms	3.4 ms	2.8	1.5
5 (NP-PP)	76 ms	5.5 ms	4.2	2.2

FIG. 5. Percentages of TAK judgments as a function of test vowel duration. The data are pooled over 12 listeners and five test utterances after normalization by multiplying each test vowel duration by the grand mean of all phoneme boundaries, divided by the phoneme boundary of the individual listener in the test utterance concerned. The solid curve was obtained by fitting a cumulative normal distribution to the data, without weighting and including all 100% and 0% points.

after this normalization are plotted in Fig. 5. The σ of the cumulative normal distribution fitted to these data is 4.9 ms.

If and how one may compare this accuracy in vowel duration perception with values of differential thresholds of signal duration reported in the literature, which are generally obtained with a 2I2AFC task, depends on the models one assumes to describe the subject's behavior in the two tasks. Let us assume that in the identification task a listener compares his representation X of the test segment duration with an internal criterion C which is fixed for each test utterance, and responds to all X which are smaller than C with a short vowel judgment and to all X which are greater than C with a long vowel judgment. This model closely resembles the one used in signal detection theory to describe the subject's behavior in a single-interval yes-no detection task. According to signal detection theory the sensitivity, d' , which is expressed as the separation of means of two stimulus strength distributions divided by the common standard deviation is $\sqrt{2}$ smaller than the corresponding measure of sensitivity in a forced-choice two-interval task, expressed as two times the separation of means of two stimulus strength distributions divided by the common standard deviation (Green and Swets, 1966, p. 68). Accordingly if we wish to predict the actual stimulus separation (d_{2I}) needed to obtain a given level of performance (e.g., the 75% differential threshold) in a two-interval task from the stimulus separation (d_{1I}) found at the same level of performance in a single-interval task, we can apply the following expression:

$$d_{2I} = (2/\sqrt{2}) \cdot d_{1I} = \sqrt{2} \cdot d_{1I}.$$

The standard deviation, then, of the identification functions measured in our experiment, which is about 5 ms, would correspond to a standard deviation of the psychometric function measured in a two-interval task of about 7 ms, and to a 75% differential threshold of, again, about 5 ms, at a reference duration $T=90$ ms. This

compares favorably with most data on the durational discrimination of nonspeech sounds. Fujisaki *et al.* report a standard deviation, obtained in a durational discrimination task with tone bursts of 100 ms, of 9.6 ms. Abel (1972) reports a 75% differential threshold at $T=63$ ms of approximately 10 ms for noise bursts of various bandwidths and of 6 ms for a 1000-Hz sine wave. Henry (1948) finds a 75% threshold of even 16 ms at $T=77$ ms. Considerably better results were obtained by Ruhm *et al.* (1966), who used a method of limits and extensive training of the listeners. They found a threshold of about 3 ms for both $T=80$ ms and $T=100$ ms, at normal loudness levels. Granted that there can be small differences in accuracy of duration perception due to training of the subjects and due to other differences in experimental procedure, and that our comparison between identification and discrimination data does not take into account a possibly different effect of "memory noise" (Durlach and Braida, 1969) in the two tasks, there seems to be nothing special about the accuracy measured in our identification test. It falls within the range one would expect from discrimination data reported in the literature for nonspeech sounds, but that the accuracy we measured is somewhat better than most discrimination data would suggest. This indicates that our listeners kept their internal criterion for short versus long vowel identification extremely constant within each test utterance. Let us now compare our result with other estimates of the accuracy of vowel duration perception in spoken words or sentences. Comparable results, obtained with a rather different method, were reported by Nooteboom (1973) who estimated the accuracy of vowel duration perception in synthesized Dutch three-syllable words from standard deviations in repeated adjustments to an internal criterion. Subjects were asked to reproduce 20 times the same duration of an embedded stressed vowel as accurately as possible by turning a control knob of a speech synthesizer. The relation between knob position and vowel duration and between knob angular displacement and durational change was changed after each trial. Standard deviations ranged from 2 to 7 ms for short vowels (mean durations in between 50 and 70 ms), and from 4 to 9 ms for long vowels (mean durations in between 100 and 150 ms). The results for short vowels were practically identical to those reported in Table I, presenting standard deviations in different test utterances ranging from 2.8 to 5.7 ms.

Huggins' (1972) data are also of interest. He applied an "up-down" strategy separately to "normal" versus "short" and "normal" versus "long" judgments for speech sounds embedded in spoken sentences. His data show that the subjects used two internal criteria, one for the first and one for the second type of judgment. Huggins supplies the 95% confidence limits based on 20 independent estimates, which in those cases where there is no substantial drift or bias roughly range from 1 to 3 ms. The corresponding standard deviations must have been about 2 and 7 ms with vowel durations of the order of 100 ms, again in close agreement with our data.

Fujisaki, Nakamura, and Imoto (1975), in a short vow-

el versus long vowel identification test essentially similar to ours, find in a synthesized sentence a standard deviation of the cumulative normal distributions of 7.1 ms, at a phoneme boundary of 168 ms. Data suggesting substantially lower accuracies were obtained by Klatt and Cooper (1975) in a magnitude estimation task. JND's of the segment /il/ in *deal* and of the segment /ŝ/ in *fish* were estimated from d' values between the distributions of sensation magnitude on a nine point scale for nine different stimulus durations. JND's found for /il/ (stimulus durations from 105 to 255 ms) ranged from 22 to 59 ms, depending on position in the sentence. JND's for /ŝ/ (stimulus durations from 55 to 135 ms) ranged from 25 to 98 ms, depending on position. The large JND's are attributed by the authors to the fact that a different sentence was presented on each trial, in order to come close to normal listening situations, and to the fact that judgments were not phonemic so that the listeners had no resort to a stable internal criterion. Other factors which may have contributed to the large JND's are the difficulty of the scaling task and the fact that Klatt and Cooper made no attempt to normalize for differences between subjects. Carlson and Granström (1975) applied both an ABX and an AX task to the discriminability of speech sound durations in bisyllabic words. Variability in the listener's responses was so great, probably due to overloading of sensory storage, that no JND's could be estimated.

So far this comparison between the accuracy measured in our phoneme identification experiment and other data on the sensitivity to sound duration, both for isolated nonspeech sounds and for speech sounds in longer utterances, shows a remarkably good agreement of data obtained with different psychophysical methods. Only data obtained with methods which do not minimize the effect of sensory storage or memory limitations ("memory noise") are clearly different. We feel safe in concluding that a binary forced choice identification experiment is a reliable and easy way of measuring the accuracy with which acoustic speech sound duration is auditorily represented. For a more expanded argumentation of the basic similarity between forced choice phoneme identification and discrimination with respect to other stimulus dimensions the reader is referred to Schouten (1978).

2. Position effects on phoneme boundaries

Although within each test utterance the phoneme boundary for each subject is rather precisely given, suggesting that subjects keep their internal criterion constant, we do find in Table III systematic differences in phoneme boundaries from test utterance to test utterance, suggesting that the internal criterion is systematically affected by the structure of the test utterance. Let us assume that the listeners on the basis of some as yet unknown aspect of the test utterance construct an expected duration of both /a/ and /a:/ to fill the test segment slot in each test utterance, and that they shift their internal criterion for vowel length perception in accordance with the systematic variations in expected durations from test utterance to test utterance. From this model one would predict a positive

linear relationship between expected durations of /a/ or /a:/ on the one hand, and measured phoneme boundaries on the other. To test this prediction we have to make one additional assumption because we have no direct measure of the expected durations. Two alternative hypotheses suggest themselves: (a) Expected durations are determined solely by the abstract linguistic (syntactic, perhaps also semantic) structure of the sentences, and thus are not specific to particular realizations of these sentences; (b) Expected durations are determined by the specific auditory-phonetic structure of the attended utterance which of course in turn may be at least partly determined by the linguistic structure of the sentences. From hypothesis (a) we predict a positive correlation between the measured phoneme boundaries and the durations of both TAAK and TAK in the production experiment, because the abstract linguistic structure of each sentence type is the same in the production and the perception experiment. In order to minimize the variation due to individual differences in the internal criterion, which are of less interest to us than the relative effect of expected durations on the internal criterion within subjects, we normalized the measured phoneme boundaries of individual subjects on the grand mean of all phoneme boundaries before correlating phoneme boundaries with production data. Calculating the product moment correlation coefficient of the 12 normalized phoneme boundaries in each of the five test utterances with the mean vowel durations in both TAK and TAAK (five tokens of each of four speakers) in each of the five corresponding sentence types, gives for TAK $r = 0.71$ ($p < 0.001$; regression line given by $y = 1.09x - 2.2$; standard error of estimate $S_y = 4.3$; standard error of the slope $S_b = 0.15$) and for TAAK $r = 0.8$ ($p < 0.001$; regression line given by $y = 0.65x - 18$; standard error of estimate $S_y = 2.5$; standard error of the slope $= 0.02$). These correlations show that to some extent the vowel durations measured in the production experiment in the five sentence types and the phoneme boundaries found in the perception experiment in the corresponding test utterances, are controlled by differences between the five sentences which are common to production and perception experiment, most probably differences in syntactic position of the test words. Of course the correlations do not show that the effect of syntactic structure on phoneme boundaries is immediate. It might be mediated by the particular auditory-phonetic form of the test utterances used in the perception experiment, the production of which was also influenced by the syntactic structure of the sentences. Thus hypothesis (b) may still be correct. This hypothesis predicts a close correlation between the phoneme boundaries on the one hand and the original vowel durations in the test utterances used in the perception experiment on the other, under the assumption that listeners can recreate the duration of a speech sound embedded in an utterance when this speech sound (and its duration) is removed from the utterance. That this is a reasonable assumption follows from experimental results by Nootboom (1973) and Klatt and Cooper (1975). As the test utterances were originally produced with TAAK and not TAK we can only test the prediction that there is a close positive relation between the duration of the original vowels

in TAAK on the one hand and the normalized measured phoneme boundaries of the 12 subjects on the other. The product moment correlation appears to be $r = 0.83$ ($p < 0.001$; regression line is given by $y = 0.49x + 3.4$; standard error of estimate $S_y = 3.6$ ms; standard error of the slope = 0.04). This relation between original, but deleted, vowel durations in the test utterances and the normalized measured phoneme boundaries is graphically represented in Fig. 6.

These data confirm the hypothesis that our listeners indeed constructed fair estimates of what the vowel durations in each test utterance should have been (i.e., expected durations), and adjusted their internal criterion for vowel length perception accordingly. They do not allow us, however, to choose between hypothesis (a) and (b). It remains to be shown that not only the syntactic structure, but also the specific auditory-phonetic structure of the test utterances in the perception experiment contributes to the expected durations and so affects the phoneme boundaries. Inspection of the spectrograms of the five test utterances (Fig. 3) shows that at least in these utterances an effect of the immediately following context is more probable than an effect of the immediately preceding context, which is in line with the more general observation that the major context effects on vowel duration are from later to earlier rather than from earlier to later (Nooteboom and Cohen, 1975). From Fig. 7, displaying the response latencies measured in the present perceptual experiment, it can be seen that the increase of response latencies found for ambiguous segments alone would enable the listener to employ auditory-phonetic evidence over at least a few hundreds of milliseconds after the test segment in his phonemic decision. Below we will, in perceptual experiment II, first demonstrate that the phoneme boundary can indeed be influenced by what follows in the utterance. We will then, in perceptual experiment III, show that the phoneme boundary can be systematically affected by an aspect of the auditory phonetic structure of the test utterance that is related to the perceived prosodic form but not to the perceived syntactic structure of the sentence.

FIG. 6. Phoneme boundaries, after normalization on the mean phoneme boundary duration in each test utterance, as a function of the duration of the vowel segments in TAAK in the original recordings of these five test utterances. Note that these vowel segments have been removed and replaced by test segments of varying durations in the actual stimulus utterances. $r = 0.83$; $y = 0.49x + 3.4$; $S_y = 3.6$; $S_B = 0.04$ (cf. footnote 1).

FIG. 7. Mean response latency times of 12 listeners, in five different test utterances, as a function of test segment duration minus phoneme boundary duration.

III. PERCEPTION EXPERIMENT II

A. Method

1. Stimulus preparation

Starting from two of the digitized test utterances used in perceptual experiment I we prepared three new test utterances. The original test utterances were (2) and (5): (2) Die nieuwe TAK/TAAK zei Kees is te groot. (That new BRANCH/TASK said John is too big.) (5) Kees kreeg een nieuwe TAK/TAAK op zijn schouders. (John got a new BRANCH/TASK on his shoulders.) In these two utterances the acoustic structures of /t/ and /k/ in the test words were made identical. Durations of the silent intervals of /t/ and /k/ were made 110 and 65 ms, respectively, the noise bursts of /t/ and /k/ were approximately 10 and 20 ms, respectively. We thus obtained two of the new test utterances (2') and (5'). A third was obtained by removing all speech material after the /k/ burst in (5'). In this way a new test utterance (1') was formed differing only from (5') in what follows the test word, i.e., either the word group *op zijn schouders* (on his shoulders) or silence. The silent interval duration of 65 ms and the noise burst duration of 20 ms of the /k/ are unusually short for utterance final position. This was heard as somewhat unnatural in (1'). It ensured, however, acoustic identity of (1') and (5') from the onset of the test utterances to the end of the test word. In all three test utterances the vowel slot in the test words was filled by one of seven vowel segments obtained in the same way as in perceptual experiment I. Durations were 65, 73, 81, 97, 105, and 113 ms. Factorial combinations of test vowel segments and test utterance types gave 21 stimuli. A stimulus tape was prepared with ten occurrences of each stimulus in random order. This sequence of stimuli was preceded by five more stimuli with easily identifiable vowel segments, taken from the extreme ends of the range of test vowel durations.

2. Subjects

Seven male Dutch-speaking students of the Eindhoven University of Technology served as paid volunteers in the listening experiment. They were linguistically and phonetically naïve. Three more subjects were employees of our Institute, two (the only females) being linguistically and phonetically naïve, the other being the second author of this paper. None of the subjects had a speech or hearing impairment.

3. Procedure

The procedure was the same as in perceptual experiment I, except that no response latencies were measured.

B. Results and discussion

There were no systematic differences between the paid subjects and the others. Phoneme boundaries for the three test utterances are presented in Table IV.

The difference between (1') and (2') is not significant, the other differences are ($p < 0.001$, student Newman Keuls, following a two-way analysis of variance). We may conclude from these data that the internal criterion for vowel length perception is systematically affected by perceived properties of the test utterances following the test words. This can be seen from the difference in phoneme boundary between (1') and (5'), which utterances are acoustically identical until after the /k/ of the test word. The low value of the phoneme boundary in (1') (94 ms) as compared to that in (1) in perceptual experiment I (101 ms) is most probably due to the unnaturally short durations of the /k/ closure time and noise burst.

The phoneme boundaries in (1') and (5') may serve as reference points for a further experiment (perceptual experiment III), in which the effect of a silent interval of varying duration following the /k/ burst in (5') on the phoneme boundary is studied. Of course, in the case of no silent interval the phoneme boundary should be as in (5'). In the case of a long silent interval one may expect the same phoneme boundary as in (1'), as this situation would be very similar to the presentation of (1') followed by a spurious word group. The question of interest is how long this silent interval has to be in order to produce a phoneme boundary similar to that in (1'), just long enough to suggest a perceivable speech pause within a syntactically coherent utterance, or so long that the perception of syntactic structure is broken up, so that utterance (1') rather than (5') with a speech pause is perceived.

TABLE IV. Phoneme boundaries and their standard deviations around the means over 12 listeners in three test utterances. The following pairs of test utterances differ significantly (student Newman Keuls, following a two-way analysis of variance, $p < 0.01$): 1'–3'; 2'–3'.

Test utterance	Phoneme boundary duration	
	$\bar{m}$	sd
1' (S-##)	94 ms	7 ms
2' (NP-S)	89 ms	3 ms
3' (NP-PP)	79 ms	3 ms

IV. PERCEPTION EXPERIMENT III

A. Method

1. Stimulus preparation

Starting from test utterance (5') in perceptual experiment II we prepared five new test utterances by introducing a silent gap of 0, 100, 200, 400, or 800 ms, immediately following the noise burst of the /k/ in the test word: Kees kreeg een nieuwe TAK/TAAK (≠) op zijn schouders (Kees got a new BRANCH/TASK (≠) on his shoulders). Otherwise the stimuli were identical with those derived from (5') in perceptual experiment II. We thus obtained 35 stimuli by factorial combinations of the five test utterance types and the seven test vowel segments. A stimulus tape was prepared with ten occurrences of each stimulus in random order, plus five stimuli with easily identifiable vowel segments, preceding the stimulus sequence proper.

2. Subjects

The subjects in the identification test were the same as in perceptual experiment II. Ten more subjects, all employees of our Institute, and three of whom belonged to the speech and hearing group, listened much later to the same tape, and indicated whether or not they heard a speech pause in the utterances.

3. Procedure

The procedure of the vowel length identification test was identical with that in the other perceptual experiments. The same is true for the speech pause perception test, except that the listeners were now asked to place a cross on a response sheet after the number of each stimulus in which they heard a speech pause.

B. Results and discussion

The results of both the vowel length identification test and the speech pause perception test are presented in Table V. The results are graphically represented in Fig. 8. The interrupted horizontal lines give the mean phoneme boundaries obtained in test utterances (1') and (5') in perceptual experiment II. The open circles stand for the mean phoneme boundaries obtained in the present experiment. The solid curve is given by

TABLE V. Mean phoneme boundaries, mean percentages of speech pause perception, and their respective standard deviations over ten listeners, for each of the durations of the silent gap following the test word. The following pairs of test utterances differ significantly in the mean phoneme boundary (student Newman Keuls, following a two-way analysis of variance, $p < 0.01$): 0–800 ms; 0–400 ms; 0–200 ms; 0–100 ms; 100–800 ms; 100–800 ms; 100–400 ms; 100–200 ms.

Silent gap duration	Phoneme boundary		Percentage of speech pause perception	
	$\bar{m}$	sd	$\bar{m}$	sd
0 ms	78 ms	3.2 ms	4	5
100 ms	84 ms	5.1 ms	27	31
200 ms	88 ms	3 ms	69	23
400 ms	90 ms	4 ms	92	12
800 ms	92 ms	4 ms	96	6

FIG. 8. Phoneme boundary durations as a function of the duration of a silent gap introduced immediately after the test word TAK/TAAK in test utterance (5'). The open circles give the mean phoneme boundary durations of 12 listeners. The crosses stand for phoneme boundary durations predicted from the mean phoneme boundary durations in final syllable and in embedded syllable as measured in perceptual experiment II, plus the independently measured probability that the silent gap is perceived as a speech pause. The solid curve was obtained by curve fitting.

$$y = 78 + 14[1 - e^{-x/200}]$$

in which the value 78 corresponds to the mean phoneme boundary duration found with a silent gap duration of 0 ms, 14 corresponds to the difference between the mean phoneme boundaries found with silent gap durations of 0 and 800 ms, and 200 is found by curve fitting.

The crosses in Fig. 8 are predictions of the mean phoneme boundaries from the data in perceptual experiment II plus the data on speech pause perception. These predictions are generated by

$$y = 79 + 15 \cdot P_{\text{sp}},$$

where 79 is the mean phoneme boundary duration in milliseconds found for the embedded condition [i.e., test utterance (1') in perceptual experiment II], 15 is the difference between the mean phoneme boundaries in the embedded and final-syllable conditions in perceptual experiment II [test utterances (1') and (5')], and P_{sp} is the probability of speech pause perception expressed as a fraction of 1, as measured independently in the present perceptual experiment. This simple probabilistic model assumes that each listener, within this experimental task, operates with only two, discrete, internal criteria, one for the embedded condition and one for the prepausal condition. Whenever he hears a speech pause he employs the latter, in all other cases he employs the first. The internal criterion in the prepausal condition is assumed to be identical with the one the subjects employed in the utterance final syllable in perceptual experiment II. A perceived speech pause apparently breaks up the auditory flow of speech and thereby separates the test word from the auditory-phonetic environment otherwise affecting the listener's internal criterion. Note that the perceived lexical-syntactic structure of the sentence used in this experiment is not affected by the presence or absence of a perceived speech pause, as long as the duration of the pause is

within the range of speech pauses normally made by speakers within sentences. The results of this experiment confirm our hypothesis that the internal criterion for vowel length identification is primarily affected by the specific auditory-phonetic structure of the surrounding speech material, and not immediately by the syntactic structure of the perceived sentence.

V. GENERAL DISCUSSION

The two main conclusions we draw from the results and discussions presented in this paper are (1) The accuracy with which vowel durations in spoken sentences are auditorily represented is at least as good as would be predicted from most data on duration discrimination of isolated nonspeech sounds with comparable durations, and corresponds to a 75% differential threshold of some 5 ms for $T=90$ ms. (2) The internal criterion listeners employ for distinguishing phonemically long vowels from phonemically short vowels on the basis of vowel duration can be flexibly and accurately adjusted to the auditory-phonetic structure of the surrounding speech material, including speech material over a few hundreds of milliseconds following the vowel concerned.

As to the first of these conclusions, the reader may object that we have not entirely proved our point. Although we have shown that auditory representation of vowel duration is quite accurate in the immediate neighborhood of a stable internal criterion for categorical perception, this does not necessarily imply that vowel durations farther away from such a criterion are represented with equal accuracy. For those vowels the accuracies measured by Klatt and Cooper (1975) (cf. Sec. II B1) may be thought to be more representative. The matter is not without practical importance. Klatt and Cooper conclude from their data that the minimum accuracy with which a particular segmental duration has to be computed in, for example, speech synthesis by rule is of the order of 25–100 ms, at least, they say, if all other durations are computed with good accuracy. Why they make the latter restriction is not entirely clear. If they really have been measuring JND's, these would apply equally to all segment durations in a synthesized speech utterance, as long as the errors are randomly distributed over the utterance. We submit, however, that their results should rather be interpreted as perceptual tolerances for one segment per utterance in the absence of a stable internal criterion. Allowing an error of somewhere between 25 and 100 ms in all segment durations of an utterance would possibly drastically upset the perception of the temporal structure of a speech utterance. The average error in computing speech segment durations should not be greater than some 5–10 ms, that is of the same order of magnitude as the sensation noise, if one does not wish to take chances in introducing perceptible differences in overall temporal structure. One may argue, of course, that the allowable error in computing speech segment durations in speech synthesis depends rather on perceptual tolerances than on the amount of duration-sensation noise. It should be recognized, however, that the rules which govern the size of the perceptual tolerance for

each segment duration in speech utterances are still unknown, and are probably language specific. In this respect work on the perception of intonation has been more successful (cf. Cohen and 't Hart, 1967; 't Hart and Cohen, 1973; 't Hart and Collier, 1975). The difference between perceptual tolerance and the effect of sensation noise is nicely illustrated by the earlier discussed data reported by Huggins (1972). Whereas he finds tolerance ranges in the perception of speech sound duration of about 20 ms to even about 100 ms, depending on type of speech sound, position of speech sound and subject, the upper and lower limits of these tolerance ranges are in most cases determined with the same accuracy as we have found in phoneme identification, thus showing a relatively small effect of sensation noise, and of course also showing that the internal criteria for "too short" or "too long" are in most cases very stable. In the absence of a stable internal criterion and without the possibility of using an external reference signal, it does not seem feasible to measure the effect of sensation noise in the auditory representation of duration. However, if the auditory system is capable of representing the duration of a vowel which is ambiguous with respect to some internal criterion, with an accuracy of a few milliseconds and during a few hundreds of milliseconds it would be rather inefficient if this capacity were reserved for ambiguous vowel durations only, and were not used, for example, to take optimum advantage of the temporal properties of the surrounding speech material in order to resolve the ambiguity.

Our second conclusion also needs some elaboration and raises a few questions. Within the experimental task employed in our perceptual experiments the listeners could in each trial identify the test utterance, which was one of a very limited number of known test utterances and thus easily and quickly identifiable, and then concentrate on the phonemic decision. We have assumed that the perceived syntactic structure of the test utterance plus its particular auditory form enabled the listener to create an expected duration for each of the two possible candidates to the ambiguous vowel segment, and that a noiseless internal criterion employed in the phonemic decision process is affected by the expected durations in a simple way: If the values of the expected durations increase, the values of the criterion similarly increase, and vice versa. The listeners are assumed to identify all test segment durations shorter than the criterion as /a/, and all test segment durations longer than the criterion as /a:/. This hypothetical perceptual strategy is supported by the data of our perceptual experiments, which furthermore provide the following additions: The context affecting the expected durations, and thereby the criterion for vowel length perception, may include speech material following the test segment up to a few hundreds of milliseconds. The changes in criterion are in accordance with known regularities in the temporal structure of speech, reflecting for example the presence or absence of final lengthening. Differences in criterion for the same test utterance between subjects are small (standard deviations often in the order of a few milliseconds), considerably smaller in fact that one finds for differences in segment durations

between speakers in the production of the same sentences, showing that the perceptual processing is severely constrained by the actual auditory-phonetic structure of the utterances. One may ask whether it is necessary to explain our results from shifts in the internal criterion. One might imagine that the internal criterion is fixed, not adjustable to context, and that the shifts in phoneme boundary are due to systematic effects of context on the listener's subjective duration of the test segment. In principle this might be tested by measuring subjective durations with a method of adjustment to an external reference, as has been done for durations of isolated sounds and silent intervals by Burghardt (1973a and 1973b). The model in which the internal criterion is affected by context predicts no measurable differences in subjective durations between the test segments in different test utterances, whereas, of course, the other model does. It is not easy to see, however, how one would apply successfully a method of adjustment employing an external reference to speech segments embedded in longer utterances without running into the same memory limitation problems encountered in applying ABX or AX discrimination tests to such segments.

Another question that arises is the question of which particular properties of the auditory-phonetic environment bring about the shifts in internal criterion (or in subjective duration). In perception experiment IV we have systematically varied only one such property, the duration of a silent gap following the syllable containing the test segment. The differences between phoneme boundaries in test utterances (2)–(5) in perceptual experiment I must have been determined by other properties of these utterances, the most likely candidate being the rate of speech, as suitably defined. Experiments in which rate of speech as such is manipulated are at present in progress. Preliminary results show that phoneme boundaries are systematically affected by slowing down or speeding up the speech material, either immediately preceding or immediately following the test segment, or both. The effects are relatively small, however, presumably because the phonemic distinction between Dutch /a/ and /a:/ depends only partly on vowel duration, and partly on spectral properties. This may be compared to the relatively small effects of speech rate on VOT boundary values, as reported by Summerfield and Haggard (1972).

Finally there is the ever-recurring question of the relevance of these data, obtained in highly restricted laboratory experiments, to our understanding of actual speech communication. What we have shown in this paper, at least to our own satisfaction, is that listeners, when forced to, are capable of using small differences in speech sound duration within spoken sentences in a phonemic decision task, and that they can take advantage of their implicit knowledge of detailed temporal regularities in speech in optimizing their chance of correct recognition of the message. In perceiving high-quality speech, listeners probably do not depend so much on an accurate auditory representation of the detailed temporal structure of speech. This may be quite different, however, in the perception of speech through

low-quality communication channels or in a noisy environment, where spectral properties of speech may become badly distorted and the more noise-resistant temporal properties of speech may become indispensable to recognition. The relevance of temporal properties of speech to the recognition of synthesized speech has been shown by Huggins (1977). The importance of temporal properties of speech to the recognition of other forms of low-quality speech is still to be assessed. In this respect it is of interest, however, that it has been shown that the difference limen for sound duration is not necessarily affected by the presence of considerable sensorineural hearing loss (Ruhm *et al.*, 1966).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the help of A. J. Breimer in processing the data, and of H. Duifhuis and H. Timmers in interpreting the results and discussing earlier drafts of the manuscript.

¹The product moment correlation coefficient is given by

$$r = (\sum xy - 1/n \sum x \sum y) / ((n-1) \sigma_x \sigma_y),$$

the standard error of estimate by

$$S_y = \sigma_y (1 - r^2)^{1/2},$$

and the standard error in the slope by

$$S_B = \left[\frac{1}{n-2} \cdot \frac{\sigma_y^2}{\sigma_x^2} \cdot (1 - r^2) \right]^{1/2}.$$

- Abel, S. M. (1972). "Duration Discrimination of Noise and Tone Bursts," *J. Acoust. Soc. Am.* **51**, 1219-1223.
- Burghardt, H. (1973a). "Die subjektive Dauer schmalbandiger Schalle bei verschiedenen Frequenzlagen," *Acustica* **28**, 278-284.
- Burghardt, H. (1973b). "Ueber die subjektive Dauer von Schallimpulsen und Schallpausen," *Acustica* **28**, 284-290.
- Carlson, R., and Granström, B. (1975). "Perception of Segmental Duration," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. G. Neeboom (Springer-Verlag, Berlin).
- Cohen, A., and 't Hart, J. (1967). "On the Anatomy of Intonation," *Lingua* **19**, 177-192.
- Cooper, W. E., Lapointe, S. G., and Paccia, J. M. (1977). "Syntactic Blocking of Phonological Rules in Speech Production," *J. Acoust. Soc. Am.* **61**, 1314-1320.
- De Rooij, J. J. (1979). "Speech Punctuation: an Acoustic and Perceptual Study of Some Aspects of Speech Prosody in Dutch," Doctor's thesis, University of Utrecht, The Netherlands.
- Durlach, N. I., and Braida, L. D. (1969). "Intensity Perception. I. Preliminary Theory of Intensity Resolution," *J. Acoust. Soc. Am.* **46**, 372-383.
- Fujisaki, H., Nakamura, K., and Imoto, T. (1975). "Auditory Perception of Duration of Speech and Non-Speech Stimuli," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York).
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).
- 't Hart, J., and Cohen, A. (1973). "Intonation by Rule: a Perceptual Quest," *J. Phonetics* **1**, 309-327.
- 't Hart, J., and Collier, R. (1975). "Integrating Different Levels of Intonation Analysis," *J. Phonetics* **3**, 235-255.
- Henry, F. (1948). "Discrimination of the Duration of a Sound," *J. Exp. Psych.* **38**, 734-743.
- Huggins, A. W. F. (1972). "Just Noticeable Differences for Segment Duration in Natural Speech," *J. Acoust. Soc. Am.* **51**, 1270-1278.
- Huggins, A. W. F. (1975). "On Isochrony and Syntax," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York).
- Huggins, A. W. F. (1977). "Speech Timing and Intelligibility," to be published in the proceedings of Attention and Performance VII.
- Klatt, D. H. (1976). "Linguistic Uses of Segmental Duration in English: Acoustic and Perceptual Evidence," *J. Acoust. Soc. Am.* **59**, 1208-1221.
- Klatt, D. H., and Cooper, W. E. (1975). "Perception of Segment Duration in Sentence Contexts," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. G. Neeboom (Springer-Verlag Berlin).
- Lehiste, I., Olive J. P., and Streeter, L. A. (1976). "Role of duration in disambiguating syntactically ambiguous sentences," *J. Acoust. Soc. Am.* **60**, 1199-1202.
- Lindblom, B., and Rapp, K. (1973). "Some Temporal Regularities in Spoken Swedish," *Papers of the Institute of Linguistics, University of Stockholm*.
- Neeboom, S. G. (1973). "The Perceptual Reality of Some Prosodic Durations," *J. Phonetics* **1**, 25-45.
- Neeboom, S. G., and Cohen, A. (1975). "Anticipation in Speech Production and its Implications for Perception," in *Structure and Process in Speech Perception*, edited by A. Cohen and S. G. Neeboom (Springer-Verlag, Berlin).
- Pols, L. C. W., Tromp, H. C. R., and Plomp, R. (1973). "Frequency Analysis of Dutch Vowels from 50 Male Speakers," *J. Acoust. Soc. Am.* **53**, 1093-1101.
- Ruhm, H. B., Mencke, E. O., Milburn, B., Cooper, W. A., and Rose, D. E., (1966). "Differential Sensitivity to Duration of Acoustic Signals," *J. Speech Hear. Res.* **9**, 371-384.
- Schouten, M. E. H. (1978). "The Case against a Phonetic Mode of Perception," submitted for publication in *Acta Psychol.*
- Summerfield, A. Q., and Haggard, M. P. (1972). "Speech Rate Effects in Perception of Voicing," *Speech Syn. Percep.* **6**, 1-12.
- Willems, L. F., and De Jong, Th. A. (1974). "Research Tools for Speech Perception Studies," *Institute for Perception Research Annual Progress Report* **9**, 77-81.