

Dept. for Speech, Music and Hearing
**Quarterly Progress and
Status Report**

**Perception of segmental
duration -**

Carlson, R. and Granström, B.

journal: STL-QPSR
volume: 16
number: 1
year: 1975
pages: 001-016



**KTH Computer Science
and Communication**

<http://www.speech.kth.se/qpsr>

I. SPEECH PERCEPTION

A. PERCEPTION OF SEGMENTAL DURATION *

R. Carlson and B. Granström

Abstract

Reasonable models of segmental duration have been proposed by several research groups. These models are mostly based on acoustic analysis. However, the question still remains: how accurate does the descriptive model have to be if the perception sets the limits? Using different techniques we have tried to estimate the perceptual need for accuracy in speech production as well as speech synthesis.

In the second part of the paper we have tried to map the perceptual dependence between segments within a word. This has been done with the aid of production experiments using synthetic speech. Questions that will be dealt with include the importance of segment type and position relative to the stressed syllable nucleus.

* This paper will be presented at the Symposium on Dynamic Aspects of Speech Perception to be held in August 1975 in Eindhoven, The Netherlands.

Introduction

Human speech is characterized by sloppy and precise articulatory movements with the acoustical speech wave as result. The listener has to pin-point the crucial events in this wave as well as to neglect unimportant distortions in order to perceive the message. The durational structure of a sentence has been studied on different levels. These studies show astonishingly small variance in certain portions as well as marked variation in others. This paper deals with three questions within the area of speech perception. a) How accurately can a listener perceive segment duration? b) Is this process influenced by segment type as well as the duration of neighboring segments? c) If a segment duration is changed, what compensatory effects does a listener expect?

In the classical study Speech: Articulation and Perception, KOZHEVNIKOV and CHISTOVICH(16) presented evidence that the durational structure of a sentence is preplanned and that the open syllable is an important unit in this process. The articulatory process is changed by sequential commands at preplanned times. This would result in compensatory effects in speech production giving rise to a lengthening of one segment when the following one is shortened or vice versa. Similar results have been presented by LEHISTE (17). These findings have been put into question by OHALA (21) claiming that a "chain model" ("make things after each other") might be as relevant as "the comb model" ("keep the preplanned structure"). Furthermore OHALA discusses the influence of imprecise segmentation, speaking rate and the size of analyzed portions when considering the advantages and disadvantages of these models. If the "comb model" is the most basic one we would expect compensatory effects in speech perception in order to minimize the influence of local disturbances during the decoding of the message. If the opposite model is true there is no immediate need for such a behavior. As should be shown we find evidence for both models in our data and the purpose is not to reject one or the other.

Reasonable models of vowel duration have been proposed by several research groups. These models are based on analysis and production experiments (5, 7, 18, 19, 20), and have been shown to convey perceptually important cues, signalling position and stress within the word and the sentence (4). In a similar way the structure of clusters and syllables has been studied (9, 10, 15, 22). However, the question still remains: how accurate does the descriptive model have to be if the perception sets the limits?

We know from psychoacoustical experiments that the duration DL increases when the duration increases (1, 2, 3, 8). However, it is not a simple linear relation. Pause duration is harder to perceive than the duration of a sinusoid. Does this finding imply that the perception of variation in the duration of a voiceless occlusion is less acute than the perception of variation in the duration of a sound? As shown by FUJISAKI et al. this is not true (8). In most psychoacoustical experiments two simple stimuli are compared so that estimations are made on the basis of little, if any, information given by surrounding sounds. However, in speech we do not seem to estimate the duration of each segment as an isolated part. In tables presented by FUJISAKI et al. we find evidence that the DL is even better than expected from psychoacoustical experiments when the duration is important for phoneme classification. The pause/sound difference has disappeared. To conclude, psychoacoustical experiments cannot be used in general to explain the perception of speech segment duration.

In several experiments HUGGINS(11, 12, 13, 14) has shown that compensatory changes in certain segment durations could be made, and utterances were still classified as "normal". His results indicate that segment durations are not all important in the perception if certain conditions are fulfilled. However, in one position a compensatory change could not restore the "normal" responses. A consonant followed by a stressed vowel was not a sequence in which compensation could help. HUGGINS summarizes his results by putting forward the vowel onset hypothesis, i. e. the vowel onsets of stressed vowels have to be correctly timed and in the interval inbetween a compensation may occur. A distortion in a thus defined sequence could not be compensated by a change in the preceding or following. Experiments with speech production synchronized by external tapping may be interpreted as support of these results (24).

Questions and results described above will be tested and discussed below together with other problems concerning the perception of segmental duration.

Discrimination and identification tests

Using different techniques we have tried to estimate the perceptual need for accuracy in speech production as well as speech synthesis. Most of the experiments reviewed in the introduction shed light on the boundary condition. Far more has to be done to illuminate the acceptability and discriminability of segmental durations. In the following three experiments will be described that were designed to explore this white spot on the phonetic map.

The material

In all experiments ordinary Swedish words were used. The words carried accent II, i. e. primary stress on the first vowel and secondary on the second one. The vowels were open, phonologically short /a/ segments.

The words were recorded in an anechoic chamber and manipulated with a "tape splicing" computer program. In this program the speech signal could be inspected and marked so that e. g. certain parts of the signal could be deleted or doubled by commands from the operator. By means of this method test tapes with manipulated versions of the original were produced.

The stimuli were presented at a comfortable sound pressure level to listeners wearing head sets (HD 414). Words, segment durations, manipulated segments, number of subjects, number of responses for each stimulus and so on are presented for each experiment in Table I-A-I.

Experiment 1

In this experiments two words were tested (plasta, plasma) (see Table I-A-I). Segment durations were changed in order to study discriminability. The durations of the "s" and the "t" in /plasta/ were furthermore changed keeping the total duration of the cluster constant. The experiment was designed as an ordinary ABX test where X could be either A or B. The subjects listened to triads and had to write down if X sounded like A or B. The reference occurred at least once in each triad. The test was balanced so that no within-the-triad sequential effects could influence the result, i. e. each manipulated stimulus resulted in four triads (ABA, ABB, BAA, BAB). The triads were randomized and presented in sequences of 100 with a pause after 50. The first and last two triads were rejected. The result is presented in Fig. I-A-1.

The graphs show the number of incorrect responses. Random responses at 50 %. The 5 % and 1 % levels are also indicated. All comparisons are made inbetween the reference, marked in the middle of each graph and a distorted stimulus marked along the X-axis. Four points should be made about the result.

- (a) The vowel is clearly more discriminable than the consonants (Fig. I-A-1c). That is correct even in relation to the voiced nasal (Fig. I-A-1e).
- (b) A compensatory change of the phoneme boundary inbetween s and t is harder to perceive than a change of a single consonant duration (Fig. I-A-1a, 1b). Compensatory effects exist at the level of discrimination.
- (c) Some subjects complained describing the test as "boring" and "difficult".
- (d) How much does the test design influence the result? The short term memory of the subjects has to be used during the similarity estimation and three stimuli might introduce memory restrictions.

Experiment 2

In this experiment segments in five test words were studied. Three words "lassa", "platta", "rammma" containing phonologically long consonants were used and compared to a single word "astma" containing all three consonants as a medial cluster. An expanded test on the word "plasta" was included.

Because of the remarks (c) and (d) in experiment 1 the test design was changed to an ordinary AX procedure. The subjects listened to stimulus pairs and marked

	Test type	Word	Variation in single segment	Compensatory variation between segments	Number of subjects	Answers/manipulated stimulus	Answers/reference stimulus	Figure
Experiment 1	ABX	pl <u>a</u> ta	s, t	st	2	16	0	IA-1a-b
	ABX	pl <u>a</u> sma	a, s, m	—	4	16	0	IA-1c-e
Experiment 2	AX	l <u>a</u> ssa	s	—	7	14	112	IA-2a
	AX	pl <u>a</u> ttta	t	—	7	14	112	IA-2b
	AX	r <u>a</u> mma	m	—	7	14	112	IA-2c
	AX	<u>a</u> stma	a, s, t, m	st, tm	8	16	336	IA-2a-f
	AX	pl <u>a</u> ta	a, s, t	as, st	13, 14	26, 28	536	IA-3d-f
Experiment 3	X	pl <u>a</u> ta	s	as, st	14	42	336	IA-3a-c

Table I-A-I

on the answering sheet if they perceived them as same or different. This type of test demands a fair portion of equal pairs in order to balance the two alternative responses. The increase of stimuli is a drawback. However, the subjects do not give a 100 % "similar" on equal pairs and the responses on these could be used for estimating the chance level and setting of significance levels. Fig. I-A-2, I-A-3d-f present the result. Significance levels could be related to Fig. I-A-1. The result is summarized below and will be more generally discussed in the conclusions.

- (a) As expected from experiment 1 the perception of the vowel (Fig. I-A-2d, I-A-3d) is more critical than the consonants.
- (b) Compensation within a cluster decreases the possibility of perceiving duration variation (Fig. I-A-2e, 3e, 3f). This is also the case for the unvoiced stop voiced nasal boundary which in this case is a syllable boundary (Fig. I-A-2f).
- (c) No general trend could be found for the single consonant/consonant cluster relation. The "s" seems to be rather unaffected of the position despite the change of the actual reference duration. This is in disagreement with psychoacoustical findings. The "t" follow these findings but the "m" is contrary to them.
- (d) Compensation inbetween the vowel and the fricative (Fig. I-A-3d) does not influence the perception. This might be interpreted as a manifestation of the perceptual need for "correct" vowel length, a need stronger than perceiving the duration balance between syllables. This will be studied more in the production experiment below.
- (e) Finally, some comments on peculiar points in Fig. I-A-3e. The two leftmost points in the compensated word give contradicting information. Comparing the same points in experiment 1 (Fig. I-A-1a) we find evidence to suspect that the second leftmost point in Fig. I-A-3e is misleading. A more detailed analysis gives the result that one of two stimuli carries 80 % of the dissimilar responses for this point.

Experiment 3

This experiment was designed to test the acceptability of stimuli used in some of the discrimination tests. Manipulated versions of the word "plasta" were recorded and presented to the subjects. The acceptability was estimated on a scale from zero ("acceptable") to ten ("not acceptable"). More details are presented in Table I-A-I. The results are presented in Fig. I-A-3a-3c and should be compared to the result of the discrimination test in Fig. I-A-3d-3f.

Conclusions can be summarized in four points.

- (a) As expected from earlier experiments (HUGGINS) the duration of the vowel influences the acceptability more than that of the consonants.
- (b) The compensated manipulation of the "st" cluster doesn't at least lower the acceptability level. However, a positive change is hard to defend with our data.
- (c) The discrimination method is far more sensitive to durational change but the close agreement inbetween the two experiments is obvious.
- (d) Astonishingly great duration variation is accepted.

Production experiments with synthetic speech

As mentioned in the introduction we have tried a production paradigm to look into the durational interaction within a word. An advantage with a production experiment is that it gives a direct figure of preferred length and also an accuracy

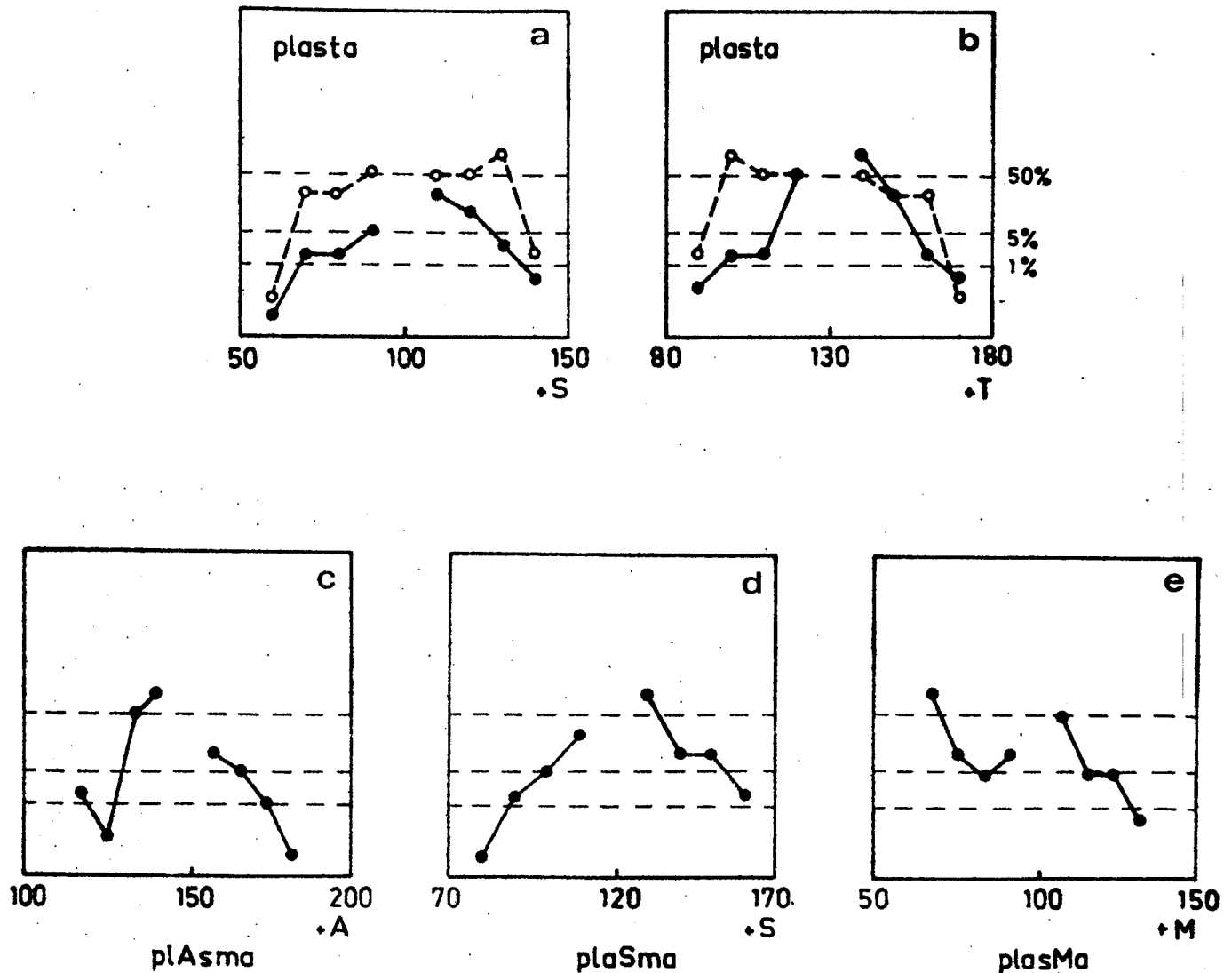


Fig. I-A-1. Results from the ABX test in experiment 1. In the triads one of the first two stimuli was always unaltered. The horizontal axis shows the duration in msec of the manipulated segment indicated below the right side of the squares. The number in the middle is the original length of the segment. The vertical axis shows the number of incorrect responses. The 50 % line indicates chance level. Points below the 5 % and 1 % line respectively are significantly different at these levels. Solid lines refer to manipulations of one segment only. Dotted lines in Fig. "a" and "b" correspond to stimuli where the total length of the cluster "st" was kept constant (compensation).

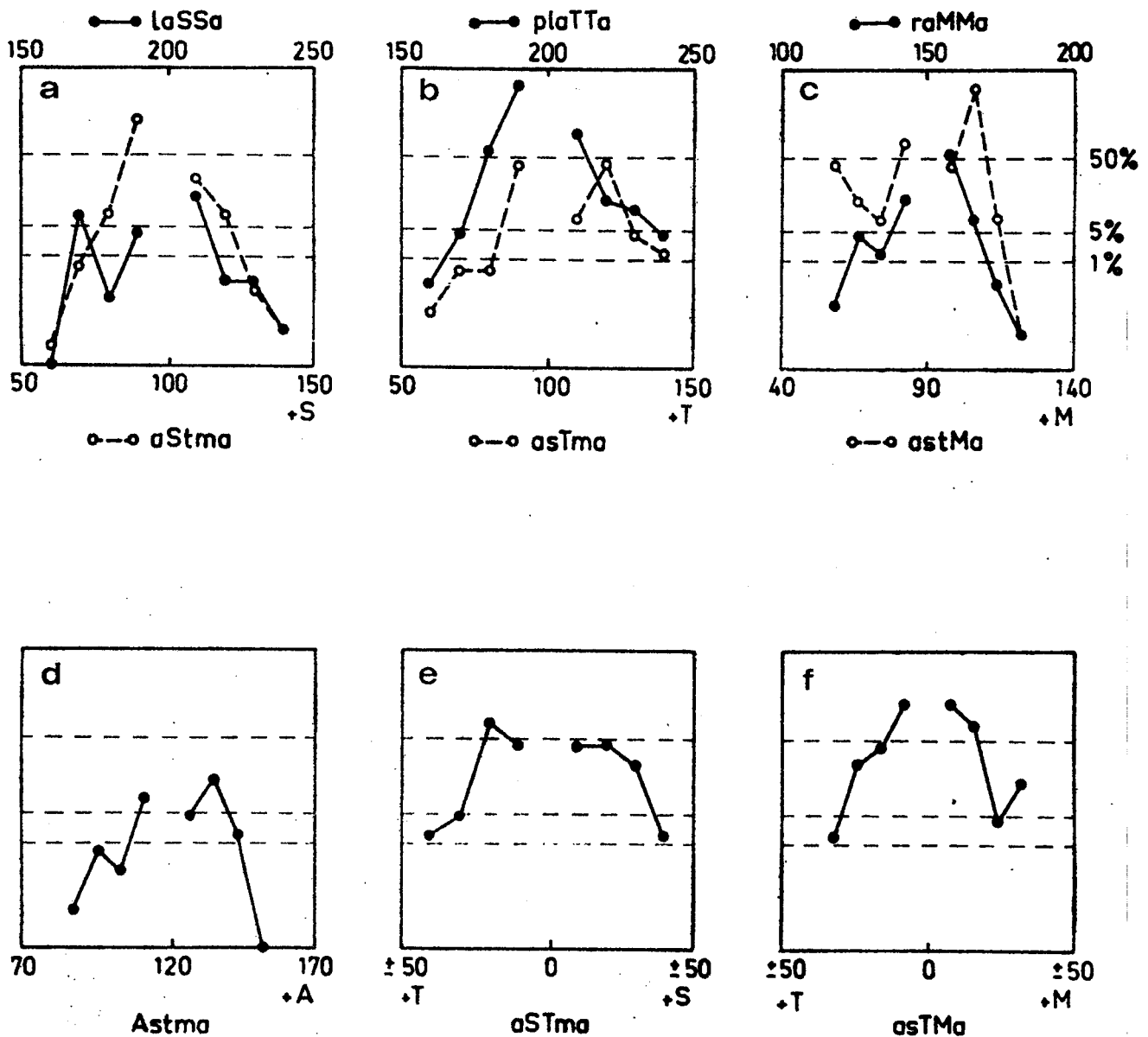


Fig. 1A-2. Results from the AX test in experiment 2. At least one stimulus in a pair was unaltered. The horizontal axis shows the duration in msec of the manipulated segments, indicated with capitals. Note that there are two scales in Fig. "a", "b", and "c" referring to single consonants (top, solid line) and cluster consonants (bottom, dotted line). The number in the middle is the original length of the segment. In Fig. "e" and "f", where the total duration of "st" and "tm" is kept constant, the horizontal scale is the duration relative to the original length. The absolute values of these segments can be seen in Fig. "a", "b", and "c", bottom scale. The vertical axis is number of "same" responses. The 50 % line indicates mean number of same responses for equal pairs. Responses below the 5 % and 1 % respectively are significantly perceived different at these levels.

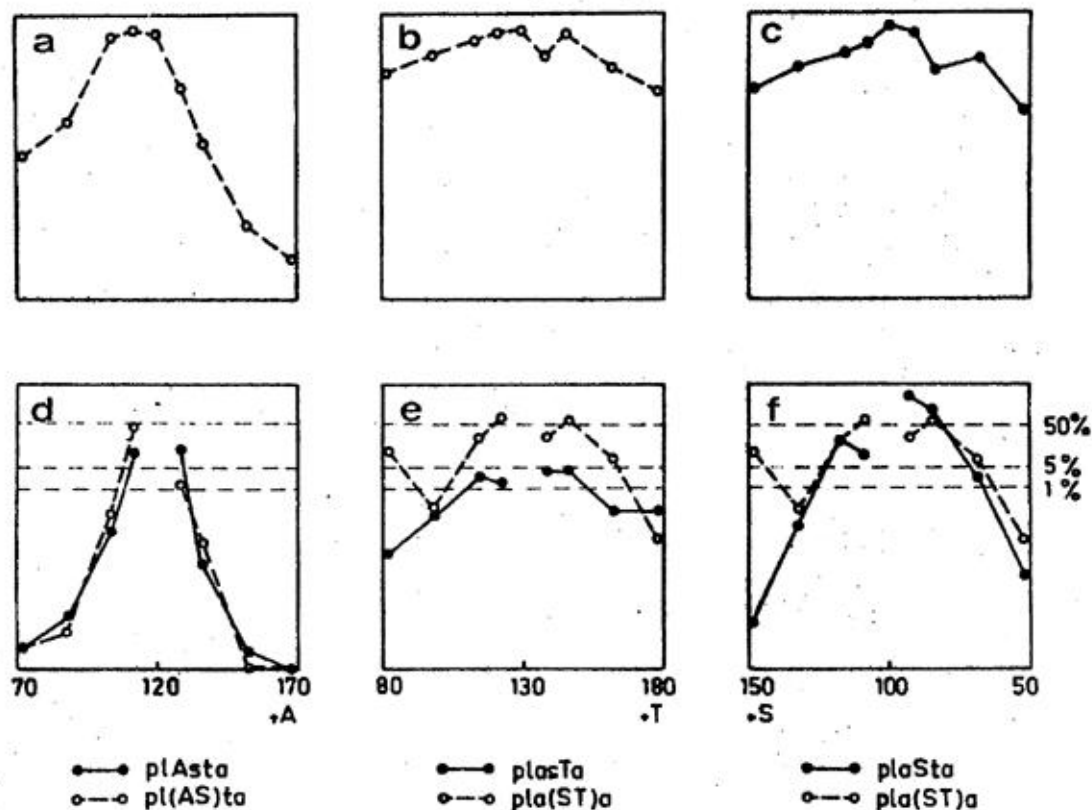


Fig. I-A-3 "d", "e", "f". Result from the AX test in experiment 2. The display is explained under Fig. I-A-2. The solid lines refer to manipulations of one segment only. Dotted lines correspond to stimuli where the total duration of "as" and "st" is kept constant. Note that the horizontal axis in Fig. "f" is reversed for comparison. Fig. "a", "b", "c" show the result from experiment 3. The vertical axis is judged acceptability. The horizontal axis and the stimuli used are the same as in Fig. "d", "e", and "f", respectively.

estimate. Synthetic speech has been used previously in production experiments for verifying the perceptual importance of vowel quality (23) and word structure (19) in estimating vowel length. A very low spread (s.d. about 5 msec) was demonstrated in these experiments indicating its usefulness.

Method

In the experiments we used an expansion of our system for synthesis by explicit, ordered rules (6). By means of a joy stick external variables could be introduced in the rules. The joy stick coordinates were displayed on a screen for visual feedback.

The subject accepted the stimulus, causing a print-out, or rejected it, causing a new synthesized utterance with the present joy stick values introduced in the rules.

In this series of experiments the duration of one known segment was controlled by the subject. After each acceptance a new stimulus was produced by randomizing the duration of another segment and also randomly changing the zero point of the joy stick.

The basic durational structure of the utterance was copied from natural speech.

In a preliminary experiment PO, Fig. I-A-4 the word "plasta" was presented in pairs where the duration of the first "a" in the second word was randomized and the "s" duration of the same word was adjusted. This situation with a fixed external reference turned out to favor the rhythmic (i. e. vowel onset) criterion with 80 % estimated compensation (i. e. negative slope of $-.80$ of the regression line). This is, however, a quite unnatural situation and the experiment only tells us that an almost perfect compensation could occur under certain conditions.

In the further experiments we have worked with a word embedded in a sentence without an external reference, "nu ska vi _____ b aten", where the test word carries the sentence stress.

We wanted to map the perceptual dependence between segments within a word of quite common structure under a condition where the weight of different criterion should be reasonably similar to a normal listening situation. Questions that will be dealt with in connection with the individual experiments include the importance of segment type (consonant-vowel, voiced-unvoiced, same or different place of articulation) and position relative to the stressed syllable nucleus. Needless to say we will only be able to draw conclusions about perceptual demands which might (or might not) influence temporal adjustments in real speech production.

In experiment P1 presented in Fig. I-A-4 we duplicated experiment PO, i. e. the first "a" randomized and "s" adjusted but now in sentence context and adjusted to the subjects' internal reference. As can be seen no or almost no compensation occurred. Possible asymmetries were tested by randomizing "s" and adjusting "a" (experiment P2). The picture is not entirely clear (Fig. I-A-4). One subject has a compensation of 19 % but this figure relies quite heavily on the single $s = 40$ msec point which was quite an extreme stimulus. The other subject showed a slightly positive slope of the regression line. These deviations from the no compensation hypothesis are however not highly significant.

The next experiment P3 ("t" randomized, "s" adjusted) tested the compensation within a post-vocalic consonant cluster "st". Here a substantial amount of compensation was preferred, 41 % and 60 % for the two subjects (Fig. I-A-5) (the standard deviation in compensation can be estimated at 8 % and 13 % resp.). One question to be answered is whether this ease of compensation is due to the shared place of articulation.

In experiment P4 we changed "plasta" for "plaska" randomizing "k" and adjusting "s", Fig. I-A-5. As can be seen both subjects showed less compensation. The decrease is however much smaller than the remaining compensation. It might be argued that the postvocalic cluster, even though morphologically associated with

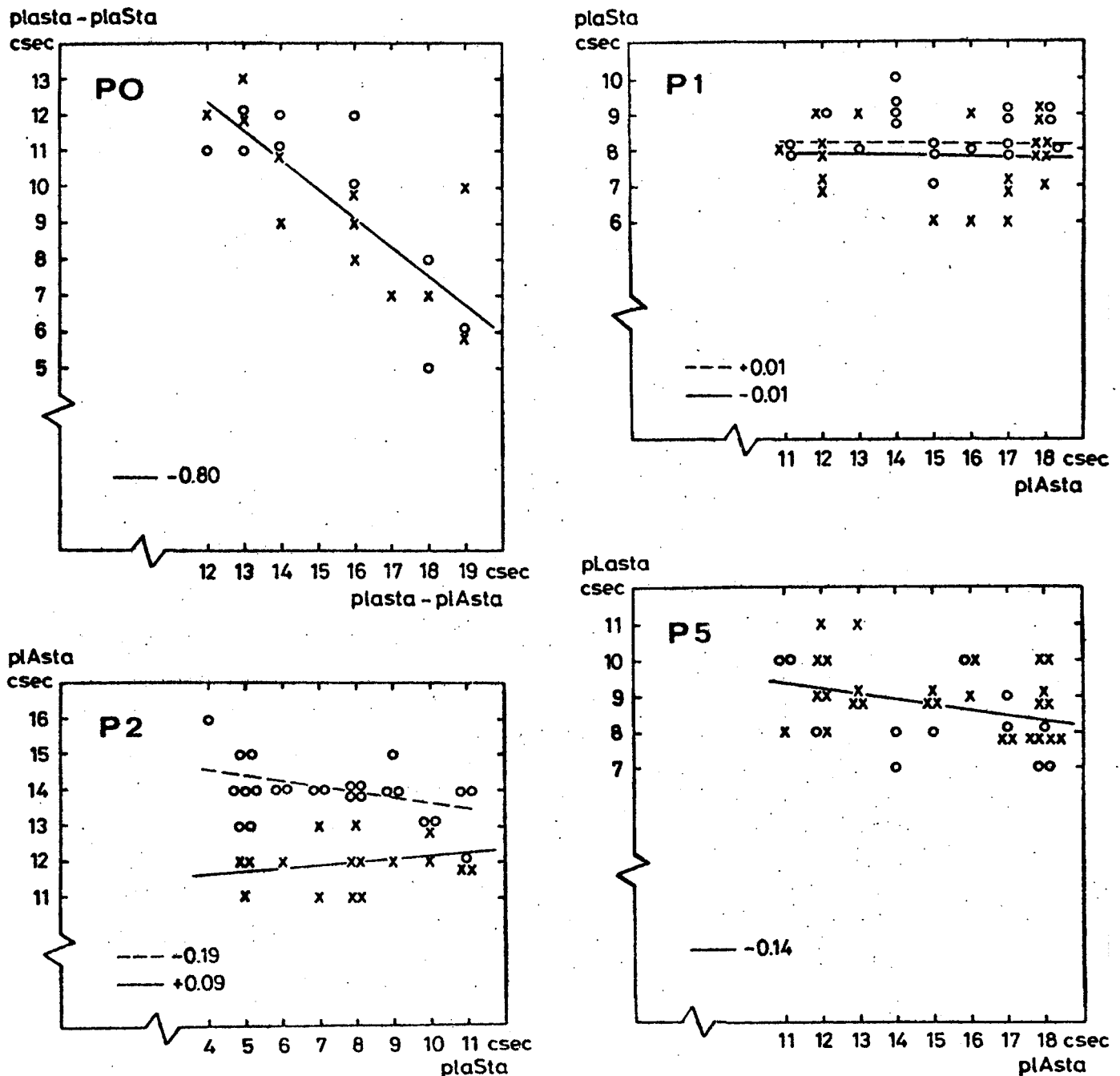


Fig. I-A-4. Vowel-consonant interaction. Results from production experiments using synthetic speech. All durations are given in 1/100 sec (csec). The horizontal axis is the randomized duration of a segment indicated by capitals. The vertical axis is the duration of the, by the subject, adjusted segment indicated by capitals. Crosses and circles pertain to different subjects. The straight lines in the figures are regression lines the slopes of which are given in the low left corners. When two lines are drawn the data from the two subjects are separated (straight line - crosses, dotted line - circles). In experiment P0 the test word is presented after a reference. In experiment P1 through P8 the test word is embedded in the sentence "nu ska vi baten".

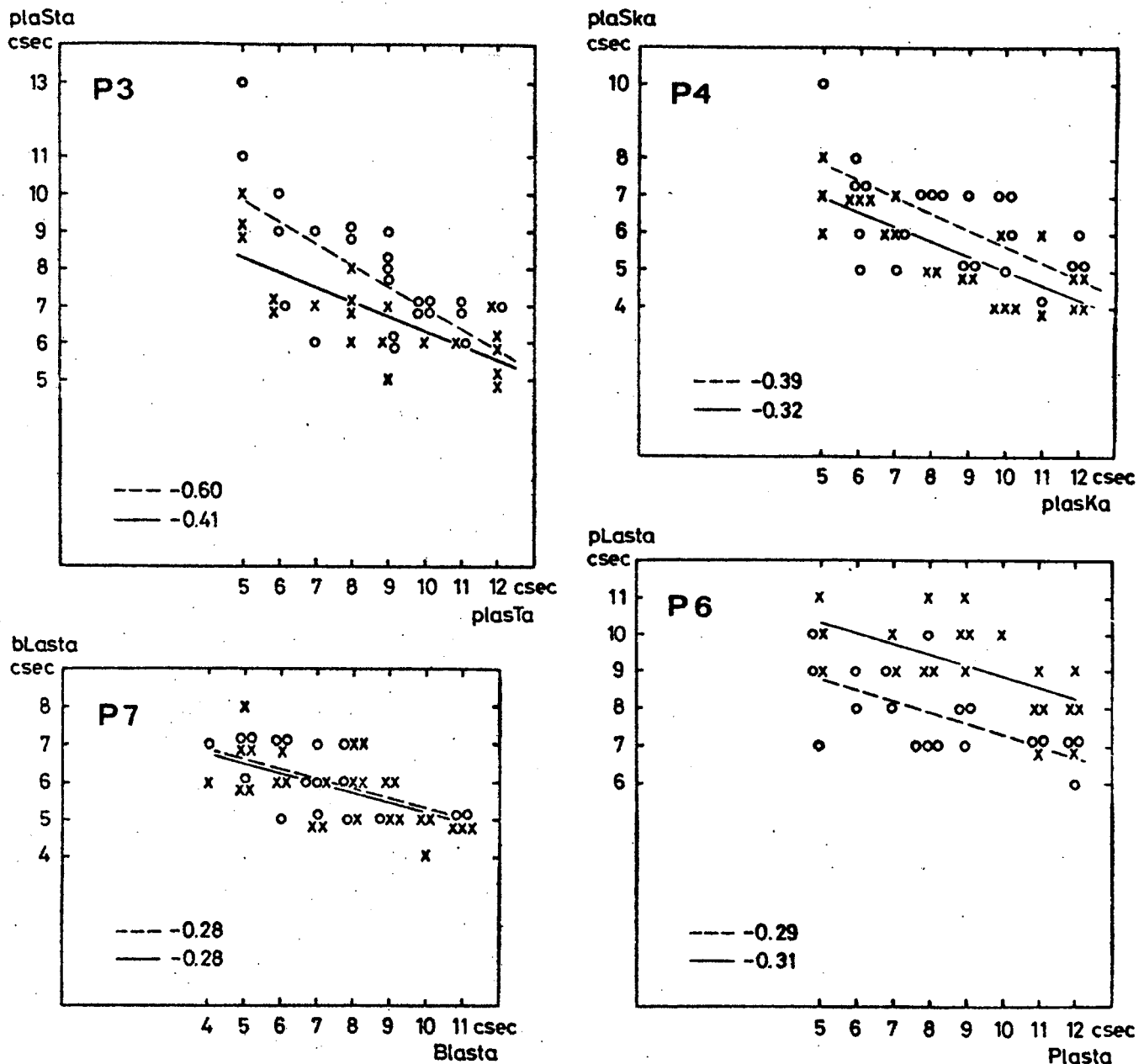


Fig. 1-A-5. Consonant-consonant interaction. Results from production experiments using synthetic speech. All durations are given in 1/100 sec (csec). The horizontal axis is the randomized duration of a segment indicated by capitals. The vertical axis is the duration of the, by the subject, adjusted segment indicated by capitals. Crosses and circles pertain to different subjects. The straight line in the figures are regression lines the slopes of which are given in the low left corners. When two lines are drawn the data from the two subjects are separated (straight line - crosses, dotted line - circles).

the preceding vowel, tends to be more related to the following vowel nucleus (i. e. open syllables as basic units (16)).

Experiment P5 (first "a" randomized, "l" adjusted, Fig. I-A-4) tested the interaction between prevocalic consonant and the stressed vowel. In this case a slight (but not highly significant) compensation was demonstrated. In this position HUGGINS (14) could not find any indication of compensation taking this as an argument for the vowel onset hypothesis. However, it is conceivable that the voiced character of the "l" made the "la" function as a unit ("l" is partly aspirated in this position but the part subjected to variations was entirely voiced). If this was true we should find a difference in the demand for compensation in a totally voiced or unvoiced or just partially voiced consonant cluster.

For the mixed cluster in experiment P6 ("p" randomized and "l" adjusted, Fig. I-A-5) we can see a compensation of close to 30 % for both subjects. By the way, we can see that the subjects in this case have clearly different internal references (15 msec difference in preferred "l" length). Looking at the all-voiced cluster (experiment P7, "b" randomized, "l" adjusted in "blasta", Fig. I-A-5) we find a very similar amount of compensation 28 % indicating that the voicing is not an important determinant for temporal grouping. This figure compares nicely with the 35 % mean compensation in the final non-homorganic all-voiceless cluster in experiment P4. Comparing preferred "l" durations in experiments P6 and P7 we can note an interesting detail. The durational difference between aspirated and unaspirated "l" that has been observed in natural speech (10, 15) is nicely reproduced in our experiment supporting the view that allophonic variations of this kind have real perceptual significance.

All experiments up to this point have dealt with temporal interaction between adjacent segments. It has been argued by HUGGINS (14) that most compensation occurs across syllable boundaries. To test one such interaction we randomized the final "a" and adjusted the first "a" (experiment P8, Fig. I-A-6). As can be seen no compensation was found, on the contrary, the vowel durations showed high positive correlation. The data points could best be described by one horizontal line corresponding to the minimum acceptable duration of the primary stressed first "a" and a regression line for durations of the secondary "a" exceeding 100 msec.

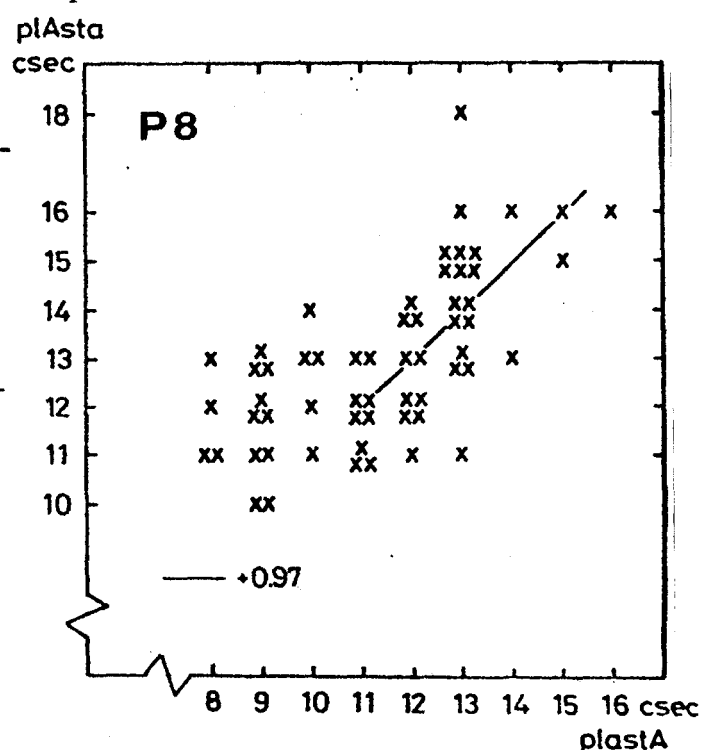
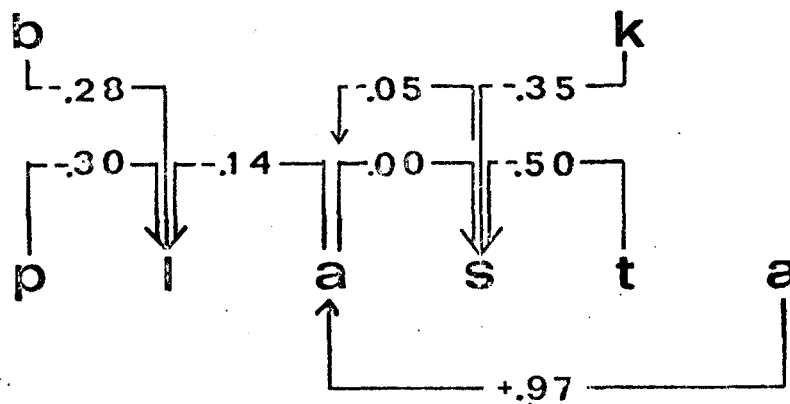


Fig. I-A-6. First A adjusted when last A randomized. See Fig. I-A-4 for further explanation.

The slope of this line is .97 pointing towards a demand on temporal balance between the two syllables. The degree of emphatic stress on the word increases with increased length of the two vowels. It should however be kept in mind that the second vowel carries secondary stress with a distinctive intonational pattern in focus position. Truly unstressed vowels may behave more like consonants and be more apt to compensation as suggested by HUGGINS (14).

The result from the production experiments P1 through P8 are summarized below in terms of mean slopes of the regression lines. Arrows point at adjusted segments.



General discussion

From several studies we know that a number of factors are important in determining the durational structure of an utterance such as segment type, stress, quantity, vowel quality, degree of emphasis, position within a word or a phrase, etc. Our prime concern has, however, been with the precision and the interaction between segments within a word asked for by the perceptual system. Even if our results are preliminary in nature and in some instances lacking a high degree of statistical significance we believe that it is possible to summarize our conclusions under the headings below.

Relation to psychoacoustics

Results from tone, noise, and silence discrimination could not explain the difference in discriminability of durations in speech. The results presented by FUJISAKI et al. that all segment types were equally discriminable were not reproduced in our experiments. It should, however, be kept in mind that their results were obtained in a phoneme boundary condition.

Accuracy of segmental duration

Vowel durations showed to be most crucial in the discrimination and rating experiments, indicating that vowel length is more important than the duration between vowels or vowel onsets. In the production experiment, however, no reliable differ-

ence was found in the accuracy of adjustment of different segment types and the spread was quite great compared to results with isolated words of simple structure. This might be explained by the fact that an expanded sentence context introduces a multitude of possible criteria the weight of which is not uniquely determined.

Our results support an approach that takes vowel lengths as the basic unit of the descriptive framework.

Compensatory effects

Compensation within a consonant cluster has been shown both to decrease discriminability and increase acceptability (in the production experiments). This is in agreement with the vowel onset hypothesis, keeping vowel onset times unaltered. In homorganic clusters a higher degree of compensation might be preferred possibly depending on the less pronounced articulatory constraints. However, the demand for compensation between consonants and vowels is either very weak or non-existent irrespective of the position relative to the vowel or shared voicing cue. This points to the fact that vowel length (rather than just onset time) plays a primary role in perception. Under certain conditions where the rhythmic demands are emphasized substantial compensation occurs between vowel and consonant (experiment PO).

Perfect compensation (100 %) has, however, not been demonstrated in any case supporting our view that a number of sometimes contradictory demands have to be fulfilled simultaneously resulting in a compromise.

Perceptual balance

A tendency to keep durational ratios constant would, if it occurred between segments, counteract the tendency for compensation on the segmental level but constant ratios between syllables and longer units would in fact demand a within the unit compensation. The positive correlation between vowel lengths (experiment P8) within a word strongly indicates that the temporal balance between syllable nuclei within a word is of major importance, far more so than the normalizing effect of the sentence or the intra-syllable balance.

References

- (1) ABEL, S.M.: "Discrimination of Temporal Gaps", J. Acoust. Soc. Am. 52 (1972), pp. 519-524.
- (2) ABEL, S.M.: "Duration Discrimination of Noise and Tone Bursts", J. Acoust. Soc. Am. 51 (1972), pp. 1219-1223.
- (3) BURGHARDT, H.: "Subjective Duration of Sinusoidal Tones", paper 20 H 1, 7th International Congress on Acoustics, Budapest 1971.
- (4) CARLSON, R., GRANSTRÖM, B., LINDBLOM, B., and RAPP, K.: "Some Timing and Fundamental Frequency Characteristics of Swedish Sentences: Data, Rules, and a Perceptual Evaluation", STL-QPSR 4/1972, pp. 11-19.

- (5) CARLSON, R. and GRANSTRÖM, B.: "Word Accent, Emphatic Stress, and Syntax in a Synthesis by Rule Scheme for Swedish", STL-QPSR 2-3/1973, pp. 31-36.
- (6) CARLSON, R. and GRANSTRÖM, B.: "A Text-to-Speech System based on a Phonetically Oriented Programming Language", in this issue of STL-QPSR.
- (7) CARLSON, R., ERIKSON, Y., GRANSTRÖM, B., LINDBLOM, B., and RAPP, K.: "Neutral and Emphatic Stress Patterns in Swedish", to be publ. in Speech Communication, Almqvist & Wiksell, Stockholm 1975, Vol. 2.
- (8) FUJISAKI, H., NAKAMURA, K., and IMOTO, T.: "Auditory Perception of Duration of Speech and Non-Speech Stimuli", to be publ. in Auditory Analysis and Perception of Speech, Academic Press, London 1975.
- (9) HAGGARD, M. P.: "Effects of Clusters on Segment Durations", pp. 1-50 in Speech Synthesis and Perception, Progress Report No. 5, 1971, Psychological Laboratory, Cambridge.
- (10) HOLMGREN-RAPP, K., unpublished data.
- (11) HUGGINS, A. W. F.: "The Perception of Timing in Natural Speech I: Compensation Within the Syllable", Language and Speech 11 (1968), pp. 1-11.
- (12) HUGGINS, A. W. F.: "How Accurately Must a Speaker Time His Articulations?", IEEE Transac. on Audio and Electroacoustics, AU-16 (1968), pp. 112-117.
- (13) HUGGINS, A. W. F.: "Just Noticeable Differences for Segment Duration in Natural Speech", J. Acoust. Soc. Am. 51 (1972), pp. 1270-1278.
- (14) HUGGINS, A. W. F.: "On the Perception of Temporal Phenomena in Speech", J. Acoust. Soc. Am. 51 (1972), pp. 1279-1290.
- (15) KLATT, D. H.: "Durational Characteristics of Prestressed Word-Initial Consonant Clusters in English", pp. 253-260 in QPR No. 108, 1973, Research Lab. of Electronics, Massachusetts Institute of Technology.
- (16) KOZHEVNIKOV, V. A. and CHISTOVICH, L. A.: Speech: Articulation and Perception, US Dept. of Commerce Translation, JPRS 30, 543, Washington, D. C. 1965.
- (17) LEHISTE, I.: "Temporal Compensation in a Quantity Language", pp. 929-937 in Proc. of the 7th International Congress of Phonetic Sciences, Montreal (ed. A. Rigault and R. Charbonneau), Mouton, The Hague 1972.
- (18) LINDBLOM, B. E. F.: "Some Temporal Regularities of Spoken Swedish", to be publ. in Auditory Analysis and Perception of Speech, Academic Press, London 1975.
- (19) NOOTEBOOM, S. G.: "Production and Perception of Vowel Duration", Philips Res. Rep., Suppl. 5, 1972.
- (20) NOOTEBOOM, S. G.: "Contextual Variation and the Perception of Phonemic Vowel Length", to be publ. in Speech Communication, Almqvist & Wiksell, Stockholm 1975, Vol. 3.
- (21) OHALA, J. J.: "The Temporal Regulation of Speech", to be publ. in Auditory Analysis and Perception of Speech, Academic Press, London 1975.
- (22) O'SHAUGHNESSY, D.: "Consonant Durations in Clusters", IEEE Transac. on Acoustics, Speech, and Signal Processing, ASSP-22 (1974), pp. 282-295.
- (23) PETERSEN, N. R.: "The Influence of Tongue Height on the Perception of Vowel Duration in Danish", ARIPUC 8 (1974), pp. 1-10 (Univ. of Copenhagen).
- (24) RAPP, K.: "A Study of Syllable Timing", STL-QPSR 1/1971, pp. 14-19.