

Interaction Between Test Word Duration  
and Length of Utterance

Ilse Lehiste

The present paper is part of a general study of speech prosody in which I have been engaged for a number of years.<sup>1</sup> The study concerns itself primarily with durational aspects of spoken English. The specific topic discussed below is the interaction between test word duration and length of utterance.

It has been found that in Swedish and Dutch, the duration of a syllable nucleus decreases as the number of syllables which remain to be produced in the word at the beginning of the syllable concerned increases.<sup>2</sup> Lindblom and Rapp, analyzing nonsense words uttered in isolation by speakers of Swedish, found that the durations of stressed long vowels ranged from about 350 milliseconds in monosyllables to about 200 msec when three syllables followed. Analyzing nonsense words spoken in isolation by Dutch informants, Nootboom observed durations of long vowels ranging from more than 200 msec in monosyllables to about 100 msec in the first syllable of words with four syllables. The question naturally arises whether the phenomenon is restricted to word level, or whether the principle might apply at the level of sentences. A further question that seems worthy of exploration is the question whether the results might conceivably be different if semantically acceptable words are used instead of nonsense words. Partial answers to both questions are presented in this paper.

Four sets of test words were used in the study. Two of the sets were similar to those used by Lindblom and Nootboom. These lists consisted of monosyllabic, disyllabic and trisyllabic words made up of the syllables big and bag in one list and bick and back in the other list. All possible stress placements were represented. The lists contained 34 words each. The third list contained 34 English words, selected to match the described nonsense words with regard to syllable length and stress placement. The fourth list (subdivided into 4a and 4b) contained ten words in which the unstressed syllable be was combined with the stressed syllables big and bag in disyllabic and trisyllabic words, and ten similar words in which the unstressed syllable be was combined with the stressed syllables bick and back. List four thus comprised 20 words; all four lists together contained 122 words.

These test words were placed in three frames: a short frame, "Say ... instead", and two long frames, in which the test word appeared either near the beginning of the utterance or near its end. The first long frame was "Sometimes it's useful to say the word ... instead". The second long frame was "The word ... is sometimes a useful example".

In the short frame and the first long frame, the test words were followed only by the disyllabic word "instead"; the words were thus at an equal distance from the end of the utterance. However, in the short frame they were preceded by one syllable, and in the long frame by nine syllables. In the second long frame, the test words were preceded by two syllables and followed by nine syllables.

The lists of words were read by three informants in the three given frames. Each informant produced 366 utterances, for a total of 1098 utterances. The informants were graduate students familiar with recording equipment and used to a laboratory environment. The records were made in an anechoic chamber, processed through a Frøkjær-Jensen trans-pitch meter and intensity meter, and displayed on a Mingograf operated at a speed of 10 centimeters per second. Measurements were made from Mingograms using generally known techniques. The duplex oscillograms produced by the experimental setup served as the principal basis for segmentation.

Since the main concern of the present study is the interaction between word duration and length of utterance, the duration of syllable nuclei within the different syllables of the test words will not be treated in this context. The basic units will be frames and word lists. Average word durations will be reported for each list; it should be kept in mind that words of one, two and three syllables have been averaged together within each list, and the average word duration for a given list is thus a somewhat abstract concept.

Tables 1, 2 and 3 present the average durations of the test words in the four lists as a function of the length of the frame.

TABLE 1

Average durations, in milliseconds, of test words produced in three frames by speaker SG.

	List 1 Words	List 2 bigbag	List 3 bickback	List 4a bebig	List 4b bebick
Frame 1: Sometimes it's useful to say the word ... instead	541	656	663	567	558
Frame 2: The word ... is sometimes a useful example	551	668	701	563	528
Frame 3: Say ... instead	586	755	761	607	601

TABLE 2

Average durations, in milliseconds, of test words produced in three frames by speaker LS

	List 1 Words	List 2 bigbag	List 3 bickback	List 4a bebig	List 4b bebick
Frame 1: Sometimes it's useful to say the word ... instead	546	575	584	477	470
Frame 2: The word ... is sometimes a useful example	537	565	571	463	496
Frame 3: Say ... instead	562	658	615	534	532

TABLE 3

Average durations, in milliseconds, of test words produced in three frames by speaker PM

	List 1 Words	List 2 bigbag	List 3 bickback	List 4a bebig	List 4b bebick
Frame 1: Sometimes it's useful to say the word ... instead	567	771	770	612	596
Frame 2: The word ... is sometimes a useful example	599	862	806	624	614
Frame 3: Say ... instead	539	842	831	663	639

Figure 1 summarizes the information for the three speakers. In the tables, List 4 is separated into 4a (containing stressed syllables with voiced final plosives) and 4b (containing stressed syllables with voiceless final plosives). A representative disyllabic word is given at the top of each column to illustrate the word types contained in each list.

A general observation may be made concerning the data for all three speakers: test words tend to be longest in the frame "Say ... instead". For speakers SG and LS, this is the case for all lists; for speaker FM, the test words are longest in the frame "Say ... instead"

in two out of four instances. In this frame as well as in the frame "Sometimes it's useful to say the word ... instead", the test words were followed by the same word, "instead". If the duration of the words depends on the number of syllables that remain to be produced in the utterance, test words should have the same duration in both frames. However, with only one exception (out of 12 instances), test words were found to be longer in the frame "Say ... instead". It seems obvious that the number of syllables remaining to be produced in the utterance does not fully determine the duration of the test words.

The frame "The word ... is sometimes a useful example" places the test words in a position in which nine syllables remain to be produced in the utterance. If the hypothesis to be tested holds, the test words should be shortest in this frame. This is true in one case out of four for speaker SG and in no instances for speaker PM. Only speaker LS has three cases out of four in which the test words are shortest in the frame in which the largest number of syllables follow the test word.

Individual variations are leveled off when all four lists and all three speakers are averaged together. Figure 1 shows the results graphically. The average durations, in milliseconds, are given inside the bars reproduced on the figure. The overall average duration of the test words was greatest in the frame "Say ... instead", noticeably smaller in the frame "The word ... is sometimes a useful example", and slightly smaller still in the frame "Sometimes it's useful to say the word ... instead". This result appears somewhat paradoxical: if the hypothesis would hold, we would expect the words to have the same duration when only the word "instead" follows, and we would expect the words to be shortest in the frame in which nine syllables follow rather than two. Clearly the results cannot be explained in terms of the number of syllables that remain to be produced in the utterance.

The apparent paradox can be solved by looking at the duration of complete utterances. Tables 4, 5, and 6 present average durations of the frame as a function of test word type and list for each of the three speakers; Figure 2 summarizes the information for all three speakers and four lists.

TABLE 4

Average durations, in milliseconds, of test words and frames in utterances produced by speaker SG

Frame and list	Duration of preceding part	Duration of word	Duration of following part	Total duration
Frame 1: Sometimes it's useful to say the word ... instead				
Words	1482	541	580	2603
bigbag	1470	656	509	2636
bickback	1460	663	506	2628
bebig	1467	567	559	2592
bebick	1478	558	540	2576
Overall average	1471	597	539	2607
Frame 2: The word ... is sometimes a useful example				
Words	233	551	1564	2348
bigbag	247	668	1545	2460
bickback	248	701	1566	2515
bebig	245	563	1598	2406
bebick	249	528	1603	2380
Overall average	244	602	1575	2421
Frame 3: Say ... instead				
Words	191	586	601	1379
bigbag	177	755	567	1499
bickback	197	761	606	1564
bebig	197	607	648	1452
bebick	197	601	611	1409
Overall average	192	662	607	1461

TABLE 5

Average duration, in milliseconds, of test words and frames in utterances produced by speaker LS.

Frame and list	Duration of preceding part	Duration of word	Duration of following part	Total duration
Frame 1: Sometimes it's useful to say the word ... instead				
Words	1629	546	574	2750
bigbag	1576	575	549	2701
bickback	1578	584	534	2696
bebig	1566	477	558	2600
bebick	1572	470	540	2582
Overall average	1584	531	551	2666
Frame 2: The word ... is sometimes a useful example				
Words	268	537	1588	2393
bigbag	253	565	1540	2358
bickback	223	571	1504	2298
bebig	239	463	1546	2248
bebick	238	496	1528	2262
Overall average	244	526	1541	2311
Frame 3: Say ... instead				
Words	216	562	603	1381
bigbag	204	658	601	1463
bickback	170	615	576	1361
bebig	183	534	623	1340
bebick	166	532	594	1292
Overall average	188	580	599	1367

TABLE 6

Average durations, in milliseconds, of test words and frames in utterances produced by speaker PM

Frame and list	Duration of preceding part	Duration of word	Duration of following part	Total duration
Frame 1: Sometimes it's useful to say the word ... instead				
Words	1760	567	529	2856
bigbag	1612	771	506	2889
bickback	1619	770	486	2875
bebig	1638	612	498	2748
bebick	1608	596	478	2682
Overall average	1647	663	499	2809
Frame 2: The word ... is sometimes a useful example				
Words	277	599	1647	2523
bigbag	301	862	1664	2827
bickback	298	806	1666	2770
bebig	312	624	1677	2613
bebick	298	614	1674	2586
Overall average	297	701	1665	2663
Frame 3: Say ... instead				
Words	181	539	549	1269
bigbag	194	842	525	1561
bickback	185	831	503	1519
bebig	175	663	544	1382
bebick	164	639	548	1351
Overall average	180	703	534	1417

For all three speakers, the duration of the whole utterance (comprising the test word and the frame) was shortest for "Say ... instead", followed by "The word ... is sometimes a useful example". When the word durations are averaged over the different lists, the duration of the words is inversely correlated with the length of the total utterance, so that the test words appear longest in the shortest utterance ("Say ... instead") and shortest in the longest utterance ("Sometimes it's useful to say the word ... instead"). This observation is supported by the fact that the duration of the word "instead" is likewise inversely correlated with the length of the utterance: in the short utterance, the duration of "instead" is greater by approximately 50 milliseconds, which is a difference of the same order of magnitude as was found for the test words.

The results of the study thus indicate that the duration of test words depends on total duration of the utterance rather than on the position of the test word within the utterance. A number of other conclusions may be drawn from these results.

I have often heard the comment that test words produced in a frame are really treated by the speakers as if they were produced in isolation, and that the use of frame sentences to simulate real utterances is at best a self-deception. I would have been convinced of that if the duration of the test words would have turned out to be completely independent of the duration of the frames in which the test words were embedded. The way the duration of the test words seems to interact with the duration of the frames shows clearly that the speakers integrate the test words into the utterance at the level at which the time program for the whole sentence is generated.

The test word lists used in this study contained both real English words and words made up of nonsense syllables. As far as interaction with the duration of the frames is concerned, there was no difference in the treatment of real words and nonsense words; both were integrated with the frame in the same way. Thus the study has also produced some evidence that at least for the investigation of the durational aspects of speech, the use of frame sentences and nonsense words may be considered justified.

#### Footnotes

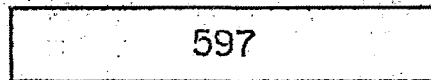
1. The study has been supported by the National Science Foundation under Grant GS-31494 #2. A preliminary version of this paper was presented at the 86th meeting of the Acoustical Society of America on October 30, 1973, at Los Angeles.

2. B. Lindblom and K. Rapp, "Reexamining the compensatory adjustment of vowel duration in Swedish words." University of Essex Occasional Papers 13 (Colchester, 1972), pp. 204-224.

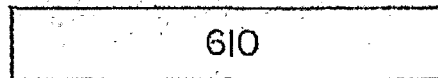
3. S. G. Nooteboom, Production and Perception of Vowel Duration: A Study of Durational Properties of Vowels in Dutch. Utrecht, 1972.



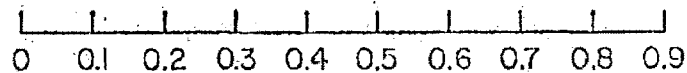
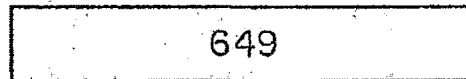
FRAME 1: SOMETIMES IT'S  
USEFUL TO SAY THE WORD \_\_\_\_  
INSTEAD



FRAME 2: THE WORD \_\_\_\_ IS  
SOMETIMES A USEFUL EXAMPLE



FRAME 3: SAY \_\_\_\_ INSTEAD



DURATION IN SECONDS

Fig. 1. Average duration of test words in different frames, averaged for three speakers.

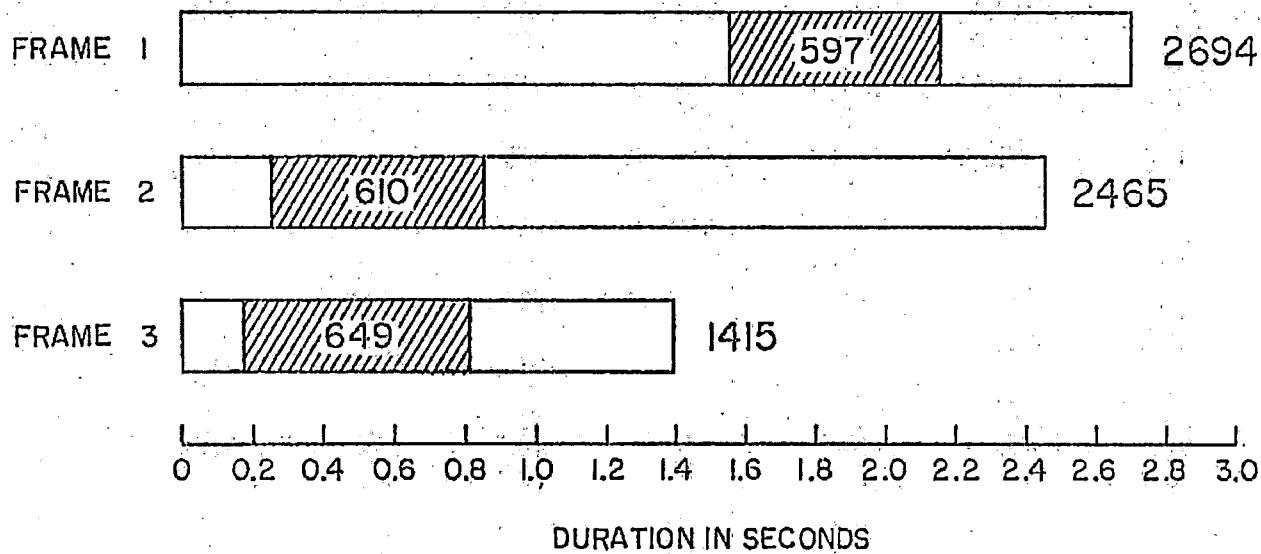


Fig. 2. Average duration of test utterances, averaged for three speakers.