Comparison of Controlled and Uncontrolled Normal Speech Rate*

Richard Gregorski, Linda Shockey,

and Ilse Lehiste

*Sponsored in part by the National Science Foundation through Grant GN 534.1 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

Comparison of Controlled and Uncontrolled Normal Speech Rate

Richard Gregorski, Linda Shockey, and Ilse Lehiste

Temporal studies have employed basically two methods for elicitation of speech rate: 1) controlled, i.e., externally induced through the use of a pulsating beat, and 2) uncontrolled, i.e., internally generated by the subject with the instruction to maintain a constant rate. Peterson and Lehiste (1960) in investigating the influence of tempo on the duration of syllable nuclei had their subjects "speak in synchronism with a periodic pulse." Lindblom (1963) used periodic clicks to manipulate speech rate in examining vowel reduction under varying tempos. Kozhevnikov and Chistovich (1965) in their experiment on the effect of rate on relative speech durations employed as a rate control a low-frequency periodic oscillation generator which was triggered by the subject's initiation of articulation. However, in their experiment to determine the number of articulatory programs in a sentence of two syntagmes, no external device was used to control rate; instead, the speaker was "instructed to adhere during all pronunciations to one and the same rate of speech." In their experimental check of syllable command hypotheses using multiple repetitions of a sentence, the subjects performed the task first at a rapid rate and then at a slow rate; no external control appears to have been employed. Nooteboom and Slis (1969) in their speech rate study had their subjects freely choose their fast, normal, and slow rates. Lehiste (1970b) in her study of the temporal organization of monosyllabic and disyllabic words in English had her subjects maintain a "subjectively constant rate."

To our knowledge, the comparability of the durations of speech units produced at a subjectively determined rate and those produced at a rate controlled by an external source has never been determined. If significant differencee exist between temporal patterns occurring in speech produced by the two methods of elicitation, obvious questions arise. For example, to what extent could we then generaliz about the temporal organization of speech from the previously mentioned studies executed with non-comparable methods? Would not the differences perhaps suggest two types of programming: 1) a basic language program including speech-unit organization and natural rhythm information, and 2) a synchronization program whose task is to adjust the language program until its natural rhythm is synchrono with the external rhythm?

It was the purpose of this experiment to determine the comparability of controlled and uncontrolled normal speech rate for both a sentence and a word spoken in isolation. Aggie was chosen fo the word, and I bag Aggie, for the sentence. The major criterion for selecting these utterances was their relatively segmentable structure when converted into oscillographic displays, and not their high semantic content. Two native speakers of English were instructed to produce both the word and the sentence about 150 times each at a comfortably constant normal rate. From recordings of these productions oscillograms were made by use of a Frøkjaer-Jensen trans-pitch meter and an Elema-Schönander Mingograph (100 mm/sec). Durations of individual segments and pauses were measured to the nearest 1/2 millimeter (i.e., 5 milliseconds). The mean duration, standard deviation, variance, and coefficient of variation ($\frac{\sigma}{M}$) were computed using an IBM 360 computer for all possible combinations of adjacent segments.

A Seth Thomas electronic metronome was used to implement the control method. To obtain the pulse rate for the controlled utterances, the mean duration for each speaker's interstress interval for both the word and the sentence of the uncontrolled productions was converted into an equivalent pulse interval on the metronome. Since for both speakers the natural sentence stress fell on the /æ/ of Aggie, it was decided to synchronize the click with this stress. The speakers were instructed to repeat the production task, only this time synchronizing the /æ/ of Aggie with the click of the metronome. The same segmentation procedures and statistical analyses that were used for the uncontrolled utterances were applied to the controlled ones. The differences between the coefficients of variation of the controlled and uncontrolled sets were computed (see Tables I-VI in the Appendix).

Figure I presents the coefficient of variation comparisons of Speaker PM's controlled and uncontrolled Aggie spoken in isolation. There was an average difference of 2% in the coefficients of variation for segments. Notice that there was no difference between the coefficients of variation of the stressed /æ/'s; in absolute terms there was only a 10 millisecond difference in their mean durations. The syllables, word and word + pause likewise had average coefficient differences of about 2%. There was a 6% difference for the pauses.

Figure II presents the coefficient of variation comparisons of Speaker LS's controlled and uncontrolled Aggie. Her average coefficient differences for both segments and syllables were about 1 1/2%. There was a .3% difference for the word.

Figures III and IV present the coefficient comparisons for Speaker PM's controlled versus uncontrolled sentences. Segments, syllables, and words as groups had average coefficient differences of 1-2%. There was a 1% difference for the sentence and a .1% difference for the sentence + pause.

Figures V and VI present Speaker LS's sentence comparisons. Segments, syllables, and words as groups had average coefficient differences of 1-2%. There was a 1% difference for the sentence and a 3% difference for the sentence + pause.

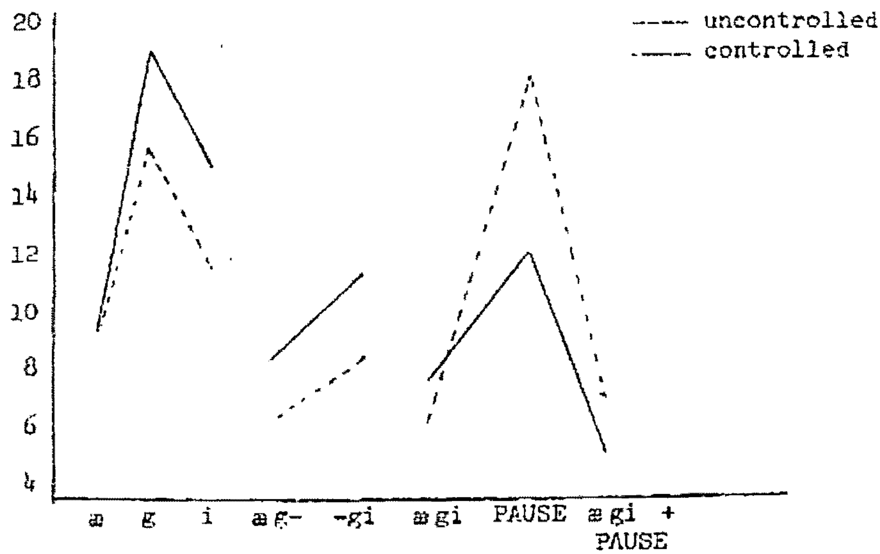

Figure I. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for Aggie produced by speaker PM.

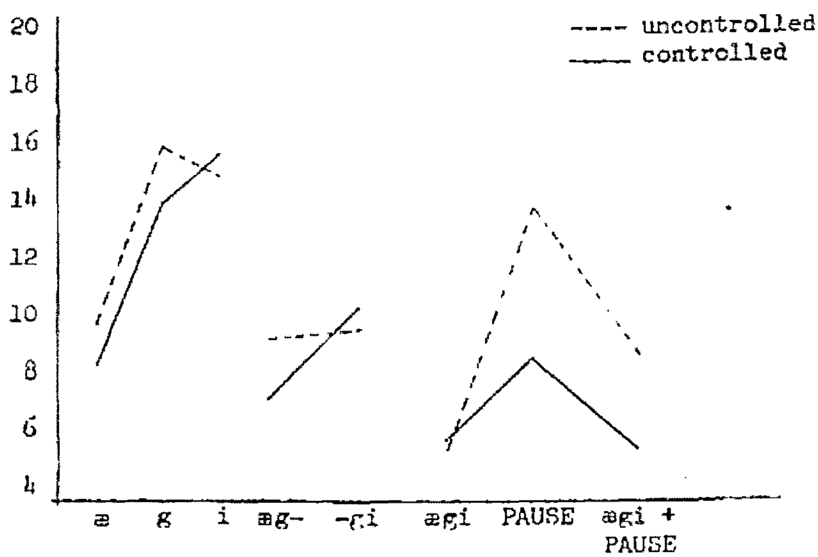

Figure II. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for Aggie produced by Speaker L.S.

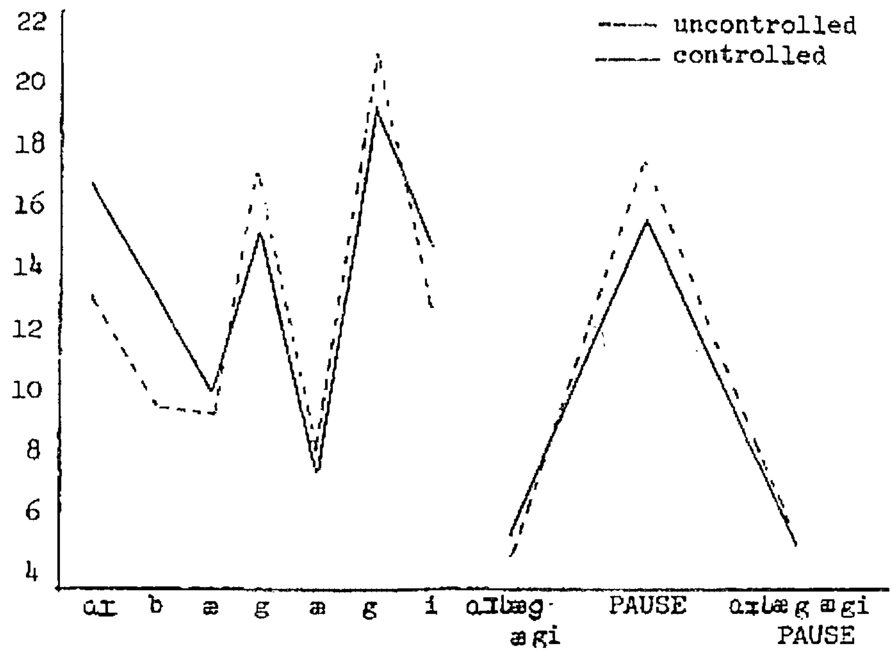

Figure III. Coefficient of Variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for I bag Aggie produced by speaker PM.

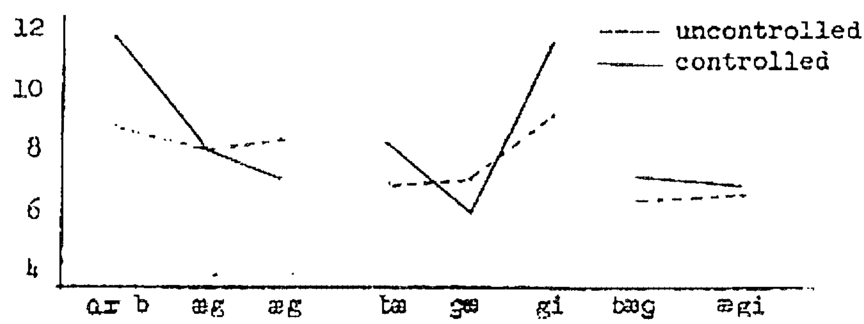

Figure IV. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for I bag Aggie produced by speaker PM.

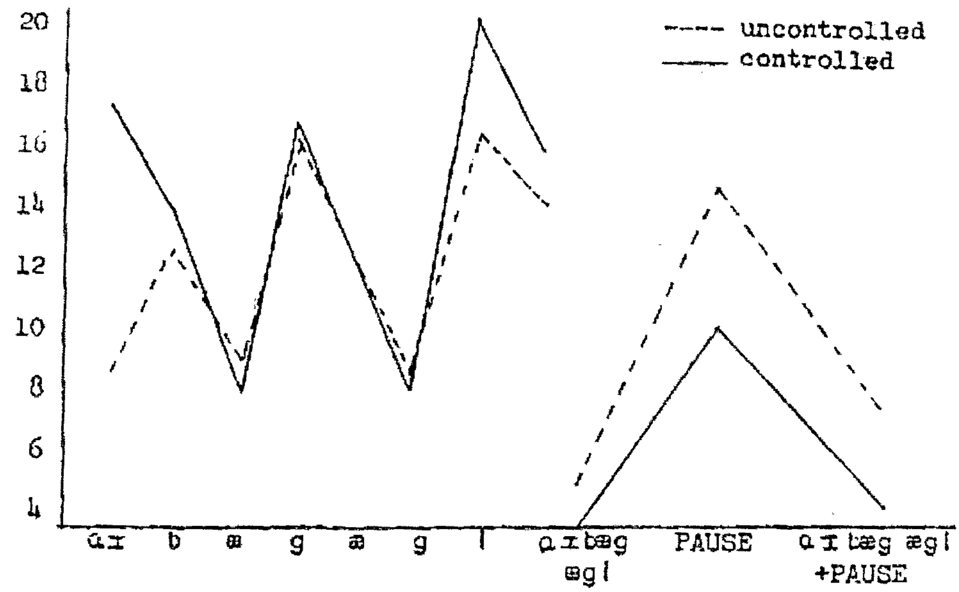

Figure V. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparisons of controlled versus uncontrolled speech-units for I bag Aggie produced by speaker LS.

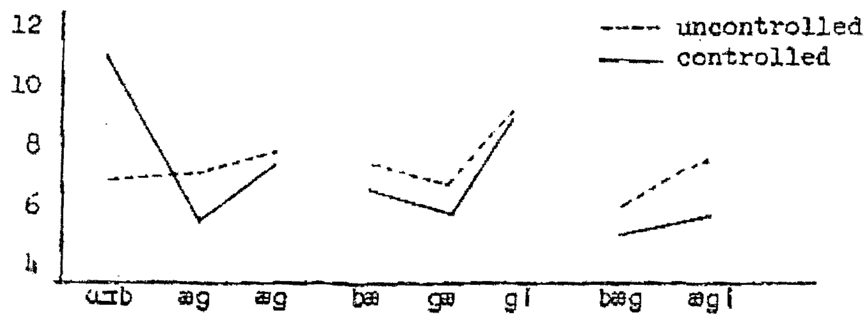

Figure VI. Coefficient of variation ($\frac{\sigma}{M}$ x 100) comparison of controlled versus uncontrolled speech-units for I bag Aggie produced by speaker LS.

To test for the significance of these coefficient of variation differences, we assumed that if the same magnitude of difference exists between two uncontrolled sets and also between two controlled sets, then such differences cannot be attributed to the control technique. We divided both the controlled and uncontrolled sets into sequential halves of about 75 tokens each. The average coefficient differences between the uncontrolled halves and also between the controlled halves were comparable to those between the entire controlled and uncontrolled sets (see Table VII in the Appendix). It thus appears that these differences are due to the natural variability of speech in a repetition task and cannot be attributed to the use of the periodic beat.

The controlled and uncontrolled sets were also examined for the direction of the differences between the coefficients of variation. We found no systematic direction to these differences for either speaker.

We conclude that in repetitions of the same words and sentences spoken at a normal rate, the two methods described here produce comparable results. However, we want to emphasize that we make no claim regarding differences between controlled and uncontrolled speech produced at other rates or using other elicitation techniques.

Appendix

TABLE I

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for Aggie produced by Speaker PM.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
æ g i	.093 .158 .116	.093 .190 .150	-- .032 .034	.022
æg gi	.063 .087	.080 .114	.017 .027	.022
ægi	.060	.076	.016	.016
PAUSE	.184	.121	.063	.063
ægi + PAUSE	.069	.050	.019	.019

TABLE II

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for <u>Aggie</u> produced by speaker LS.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
æ g i	.096 .157 .149	.083 .136 .155	.013 .021 .006	.013
æg gi	.089 .094	.067 .103	.022 .009	.016
ægi	.061	.064	.003	.003
PAUSE	.136	.086	.050	.050
ægi + PAUSE	.085	.053	.032	.032

TABLE III

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for <u>I bag Aggie</u> produced by speaker PM.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
ɑɪ b $æ_1$ g_1 $æ_2$ g_2 i	.130 .096 .092 .168 .080 .211 .126	.168 .131 .101 .147 .072 .185 .146	.038 .035 .009 .021 .008 .026 .020	.022
ɑɪb bæ $æg_1$ gæ $æg_2$ gi	.087 .067 .081 .070 .085 .093	.119 .083 .081 .061 .071 .114	.032 .016 — .009 .014 .021	.015

TABLE IV

Coefficient of variation ($\frac{\sigma}{\mu}$) comparisons of uncontrolled versus controlled speech-units for <u>I bag Aggie</u> produced by speaker PM.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
aɪbæ	.070	.094	.024	
bæg	.063	.072	.009	
ægæ	.054	.057	.003	.011
gæg	.077	.062	.015	
ægi	.066	.068	.002	
aɪbæg	.068	.083	.015	
bægæ	.045	.054	.009	.007
ægæg	.057	.055	.002	
gægi	.061	.061	---	
aɪbægæ	.051	.063	.012	
bægæg	.049	.052	.003	.007
ægægi	.049	.054	.005	
aɪbægæg	.054	.059	.005	.006
bægægi	.044	.051	.007	
aɪbægægi	.047	.056	.009	.009
PAUSE	.176	.153	.023	.023
aɪbægægi + PAUSE	.050	.051	.001	.001

TABLE V

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for _I bag Aggie_ produced by Speaker LS.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
aɪ	.107	.175	.068	
b	.126	.139	.013	
æ$_1$	.091	.082	.009	
g$_1$	.158	.165	.007	.023
æ$_2$	.086	.081	.005	
g$_2$	.159	.198	.039	
i	.140	.157	.017	
aɪb	.068	.109	.041	
bæ	.072	.063	.009	
æg$_1$	.070	.057	.013	.013
gæ	.065	.056	.009	
æg$_2$	.076	.072	.004	
gi	.091	.089	.002	

TABLE VI

Coefficient of variation ($\frac{\sigma}{M}$) comparisons of uncontrolled versus controlled speech-units for I bag Aggie produced by speaker LS.

Speech-unit	Uncontrolled	Controlled	Difference	Average Difference
aɪbæ	.063	.064	.001	
bæg	.059	.052	.007	
ægæ	.056	.046	.010	.007
gæg	.060	.059	.001	
ægɪ	.071	.056	.015	
aɪbæg	.055	.056	.001	
bægæ	.048	.042	.006	.006
ægæg	.054	.047	.007	
gægɪ	.061	.050	.011	
aɪbægæ	.051	.045	.006	
bægæg	.048	.044	.004	.009
ægægɪ	.056	.040	.016	
aɪbægæg	.050	.043	.007	.009
bægægɪ	.049	.039	.010	
aɪbægægɪ	.052	.039	.013	.013
PAUSE	.145	.102	.043	.043
aɪbægægɪ + PAUSE	.074	.045	.029	.029

TABLE VII

Coefficient of variation ($\frac{\sigma}{M}$) differences between various set comparisons of speech-units for <u>Aggie</u> and <u>I bag Aggie</u> produced by speakers PM and LS.

	Set Comparison *	Segments	Syllables	Word(s)	Sentence
Speaker PM <u>Aggie</u>	UNCON/UNCON	.014	.003	.009	
	CONT/CONT	.020	.014	.013	
	UNCON/CONT	.022	.022	.016	
Speaker LS <u>Aggie</u>	UNCON/UNCON	.051	.060	.070	
	CONT/CONT	.019	.020	.007	
	UNCON/CONT	.013	.016	.003	
Speaker PM <u>I bag Aggie</u>	UNCON/UNCON	.022	.018	.011	.020
	CONT/CONT	.029	.020	.020	.009
	UNCON/CONT	.022	.015	.006	.009
Speaker LS <u>I bag Aggie</u>	UNCON/UNCON	.015	.003	.005	.006
	CONT/CONT	.016	.006	.005	.012
	UNCON/CONT	.022	.013	.011	.013

*UNCON= Uncontrolled, CONT= Controlled.

Bibliography

Kozhevnikov, V. A., and L. A. Chistovich. Speech: Articulation and Perception. Translated by J.P.R.S., Washington, D.C., No. JPRS 30, 543. Moscow-Leningrad. 1965.

Lehiste, Ilse. Suprasegmentals, Cambridge: M.I.T. Press, 1970a.

Lehiste, Ilse. "Temporal Organization of Spoken Language," Working Papers in Linguistics No. 4, 96-113. Ohio State University, Columbus, Ohio. 1970b.

Lindblom, B. "Spectrographic Study of Vowel Reduction," Journal of The Acoustical Society of America 35, 1773-1781. 1963.

Nooteboom, S. G., and I. H. Slis. "A Note on Rate of Speech," IPO Annual Progress Report, No. 4, 58-60. Institute for Perception Research, Eindhoven, Holland. 1969.

Peterson, G. and Ilse Lehiste. "Duration of Syllable Nuclei in English," Journal of the Acoustical Society of America 32, 693-703, 1960.