JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 9, 75–78 (1970)

On Judging Pauses in Spontaneous Speech[1]

JAMES G. MARTIN

University of Maryland, College Park, Maryland 20742

Spectrograms of spontaneous speech revealed that syllables preceding a judged-pause location were usually longer than those following, whether or not a silent interval was present. Most judged-pause locations were junctures, but syllable length governed judgments independently of juncture cues.

Many psycholinguists use hesitation phenomena in spontaneous speech as a source of data and hypotheses about language structure (for example, Martin, 1968) or use (for example, Levin, Silverman, & Ford, 1967; Maclay & Osgood, 1959; Martin, 1967; Martin & Strange, 1968a; Tannenbaum, Williams, & Hillier, 1965). Of the hesitation phenomena, unfilled pauses are the most frequent and are commonly identified by physical record of silent intervals of arbitrary length (for example, Boomer, 1965; Goldman-Eisler, 1958; Hargreaves & Starkweather, 1959; Levin, Silverman, & Ford, 1967; Tannenbaum, Williams, & Wood, 1967; Verzeano & Finesinger, 1949), less commonly, by listener judgment (Maclay & Osgood, 1959; Martin, 1968; Martin & Strange, 1968a, b; Tannenbaum, Williams, & Hillier, 1965). The purpose of this methodological note is to identify some grammatical and acoustic correlates of unfilled pauses, to compare listener judgment against physical measures, and to suggest that for many purposes the latter are often superfluous.

METHOD

The speech corpus was a composite tape of 60 recorded utterances from 27 different speakers describing Thematic Apperception Test Cards (Martin & Strange, 1968b). The selection criterion for utterances was a high pause rate prior to high-information (content) words; hence, the sample is not representative of spontaneous speech under all conditions.

Two scorers working independently, research assistants who were naive about the purpose of any of the studies for which the tape was constructed, listened to each utterance at least twice, marking four types of hesitations on transcripts as they listened. Only unfilled pauses are considered here. Unfilled pauses occurring together with a filled pause between the same two words also were not counted.

The speech was physically analyzed by a Kay sound spectrograph yielding separate frequency and amplitude displays on the same time line. Identifiable stretches at background noise level along the amplitude record were marked as silent intervals. Since background noise while constant was substantial, this measure in effect overestimates the length of "real" silent intervals. Partly for this reason silent intervals of less than 50 msec. were ignored, with no effect on the conclusions following.

Syllables. Syllable lengths preceding and following a pause were measured both for scorer-identified and for spectrograph-identified pauses. Three syllables in each direction were measured, unless another pause or the beginning or end of an utterance was reached.

[1] Supported in part by Grant MH 10400, NIMH, USPHS, and by a grant to the Center for Language and Cognition by the Bio-Medical Sciences Support Committee, University of Maryland.

The spectrograph technician measuring syllable and pause durations knew nothing of the purpose of the measures.

Grammatical junctures. Certain pause locations were examined on the transcript and judged by the author to be either a grammatical juncture or not. The judgments were made blind, that is, with no knowledge (*a*) of the neighboring syllable lengths or (*b*) of whether the listeners or the spectrograph had marked the pause at that point. The judgments were liberal; not only infrequent clause and sentence boundaries but phrase boundaries as well were counted as junctures.

RESULTS

In terms of overlap, agreement in pause location between the two scorers was 91.5%.[2] Each silent interval recorded by machine and marked by at least one scorer was counted as scorer-spectrograph agreement; the overlap was 90.0%.

Preliminary analysis gave little reason to report scorer agreement and disagreement separately. Pauses were classified three ways: Listener-Machine if the spectrograph plus at least one scorer "heard" it, Listener Only if at least one scorer recorded it but the spectrograph did not, and Machine Only if a silent interval recorded by spectrograph was not marked as a pause by either scorer.

Table 1 shows, not surprisingly, that there were silent intervals at most of the places (87%) where listeners heard a pause. They ranged from 50 to 4970 msec., whereas the unheard silent intervals ranged from 50 to 110 msec. Over 21% of the Listener-Machine pauses were shorter than 200 msec., smaller than many arbitrary criteria in the literature for physically measured pauses.

Syllables. Of greater interest, however, were the pauses recorded by listeners when there was no silent interval (Listener Only) and the listener's failure to record them when there actually was silence (Machine Only). On listening to the former, it often sounded as though the speaker slowed down momentarily, then picked up speed again. Such a change in tempo seems to be the case; when there was no silent interval at a judged-pause location the syllable preceding was longer than the syllable following in 17 out of 23 cases ($p < .04$, sign test). Similar results held when there was real silence at a judged-pause location, that is, syllables preceding Listener-Machine pauses were generally longer than those following ($p < .01$). In the case of the Machine-Only pauses, however, the reverse was true; most of the syllables preceding these silent intervals were shorter than those following ($p < .02$). In short, syllable-length differences, preceding minus following, where usually opposite in sign at Listener-Only and Machine-Only pause locations and could be the cue for pause judgments.

TABLE 1

MEDIAN PAUSE AND SYLLABLE LENGTHS (IN MSEC.)

	Listener-Machine	Listeners Only	Machine Only
N	150	23	33
Pause	470	—	70
Syllable preceding	290	320	130
Syllable following	190	200	190

Grammatical junctures. The Listener-Only and Machine-Only pauses combined were classified in terms of juncture versus nonjuncture and preceding syllable longer versus preceding syllable shorter. There was an association between these variables; 80% of the junctures but only 27% of the nonjunctures were preceded by a relatively longer syllable, $\chi^2(1) = 11.82$ ($p < .01$). Apparently speakers tended to mark the ends of phrases with elongated syllables before proceeding to the next

[2] An unpracticed group of listeners placed 88% of their pause judgments in the same places as these two practiced scorers (Martin & Strange, 1968b).

constituent. Such a result is reasonable enough, suggesting that the speaker may have slowed down momentarily in order to compose his next phrase, but it raises the question of what the listeners were using as cues for pauses, syllable lengths or grammatical juncture, when the silent intervals were short or nonexistent.

To consider this question two groups of pause locations were examined, those heard (Listener Only) which were *not* junctures ($N = 10$) and those not heard by listeners (Machine Only) which *were* junctures ($N = 7$). First, note that syllable lengths following a pause location were roughly constant, about 190–200 msec., across the three groups, and call this range of values typical syllable length for the present speaking task. Syllables preceding Listener-Only pauses at nonjunctures were longer than all three (typical length) types of following syllables, as well as those preceding Machine-Only pauses at junctures (all p's $< .05$, two-tailed, except the comparison with Listener-Only, $p < .05$, one-tailed). If grammatical juncture of itself cues pause judgments, then it appears that short preceding syllables suppressed pause judgments at junctures, while long preceding syllables induced pause judgments even though grammatical juncture cues were absent. Both cases support the conclusion that syllable length was the dominant cue. Similar results held when juncture was not considered, that is, syllables preceding Listener-Only pauses were generally longer than all groups of following syllables, while syllables preceding Machine-Only pauses were generally shorter (all p's $< .05$).

Further examination suggested that many of the Machine-Only pauses might better be regarded as only normally occurring intermittent gaps in speech. All but two of these 33 short intervals between words required consonant-to-consonant transitions, many of which might normally provide a silent interval. All but nine were followed by stop consonants, which require closure of the vocal cavity with resulting silence. Finally, note that if (median) Machine-Only pauses and syllables preceding them are added together the result about equals the typical syllable length.

To conclude, elongated syllables usually accompany and precede judged-pause locations whether or not a silent interval is present, but these locations also are usually grammatical junctures. However, elongated syllables appear to cue pause judgments independently of grammatical cues. Finally, since the extended syllables which listeners hear as unfilled pauses are logically as good an indicator of hesitation as other measures commonly used (for example, filled pauses, repeats or false starts, all of which require judgment), listener judgment seems preferable to physical recording of unfilled pauses as well except, of course, in cases where the duration of real silence is an issue.

References

Boomer, D. S. Hesitation and grammatical encoding. *Language and Speech*, 1965, **8**, 148–158.

Goldman-Eisler, F. Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 96–106.

Hargreaves, W. A., & Starkweather, J. A. Collection of temporal data with the duration tabulator. *Journal of the Experimental Analysis of Behavior*, 1959, **2**, 179.

Levin, H., Silverman, I., & Ford, B. Hesitations in children's speech during explanation and description. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 560–564.

Maclay, H., & Osgood, C. E. Hesitation phenomena in spontaneous English speech. *Word*, 1959, **15**, 19–44.

Martin, J. G. Hesitations in the speaker's production and listener's reproduction of utterances. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 903–909.

Martin, J. G. Some acoustic and grammatical features of spontaneous speech. Paper presented at the Conference on the Perception of Language, University of Pittsburgh, 1968.

Martin, J. G., & Strange, W. Determinants of hesitations in spontaneous speech. *Journal of Experimental Psychology*, 1968, **76**, 474–479. (a)

MARTIN, J. G., & STRANGE, W. The perception of hesitation in spontaneous speech. *Perception and Psychophysics*, 1968, **3**, 427–438. (b)

TANNENBAUM, P. H., WILLIAMS, F., & HILLIER, C. S. Word predictability in the environments of hesitations. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 134–140.

TANNENBAUM, P. H., WILLIAMS, F., & WOOD, B. S. Hesitation phenomena and related encoding characteristics in speech and typewriting. *Language and Speech*, 1967, **10**, 203–215.

VERZEANO, M., & FINESINGER, J. E. An automatic analyzer for the study of speech in interaction and in free association. *Science*, 1949, **110**, 45.

(Received July 11, 1969)