A usual extreme case: Pause reports of informal spontaneous dialogue

JODY ARLINGTON, SEBASTIAN M. BRENNINKMEYER, DANIELLE ARN,
RITA GRUNDHAUSER, and DANIEL C. O'CONNELL
Georgetown University, Washington, D.C.

Native speakers ($N = 49$) of American English (with and without advanced fluency in German), French, and German were asked to report pauses and estimate their duration in English and German texts of informal spontaneous dialogue. In previous research, stimulus materials for pause reports have consisted of very "orderly" discourse: monologic or dialogic, but of media quality and without the overlap, laughter, extraneous noise, rapid articulation, and slang typical of informal spontaneous dialogue. Addition of these characteristics makes the task quite difficult and diminishes both accuracy (hits/possible hits) and efficiency (hits/[hits + false alarms]) of pause reports. In both English and German texts, estimates of pause duration at actual pause positions were significantly longer than those at false alarm positions where there were no actual pauses in the text. Significant differences in accuracy across language groups were limited to the German corpus; fluency in the German was associated with greater accuracy.

Before the linguists prefix our title with an asterisk, let us explain. In the archival literature, manageable corpora of discourse have been used for the investigation of pause reports. Experiments have involved media-quality recordings of formal, memorized, syntactic, monologic, and dialogic discourse. Chiappetta, Monti, and O'Connell (1987) used Italian radio homilies. Stuckenberg and O'Connell (1988) used English and German radio homilies. Carpenter and O'Connell (1988) and Ulasevich, Leaton, Kramer, Hieke, and O'Connell (1991) used poetry read by professional actors or the poets themselves. Gardner, McMillan, Townsend-Handscomb, Barrett-Bates, and O'Connell (1990) used a scene from *Hamlet* performed by professional actors in English and in German translation. Friedman and O'Connell (1991) used political interviews taken from American and German television. Monti, O'Connell, and Chiappetta (1990), however, used a stimulus impoverishment method in which subjects "surmised" pauses for a printed text without any auditory input whatsoever.

In none of these investigations was the typical conversational situation, with its slang, interruptions, laughter, rapid articulation, overlap, extraneous sound, and other attendant chaos, engaged. In other words, the stimulus materials used in the extant research have been extreme with respect to the usual, everyday discourse that listeners are most frequently exposed to.

Correspondingly, informal spontaneous dialogue becomes an extreme case relative to previously used stimulus materials. Our hypothesis is that the characteristics of informal spontaneous dialogue will render pause reports more difficult. This difficulty will be manifested in a wide variation of estimates of duration and a consequent lack of significant differences between subjects with different native languages. Accuracy, defined as the percentage of pauses accurately reported over the number of actual pauses to be reported, is expected to diminish with the addition of these characteristics. Likewise, efficiency, defined as the percentage of pauses accurately reported over the sum of pauses accurately reported and pauses reported where there are none (false alarms), is also expected to diminish accordingly.

There are a number of suggestions of these hypotheses throughout the literature. For example, the highest accuracy index (93.5%) obtained in any of this research was found by Ulasevich et al. (1991), who used as stimulus materials a reading by William Carlos Williams of his poem "The Botticellian Trees." The mean duration of the pauses in this reading was 1,030 msec, the longest mean duration in any of this research; the articulation rate (3.20 syllables per second of on-time or actual speaking time) was also the slowest articulation rate in any of this research to date.

METHOD

Two samples of informal spontaneous dialogue, one in English between two native speakers of English and one in German between two native speakers of German, were used as stimulus materials. Descriptive statistics for these corpora are presented in Table 1.

Experimental subjects for the study were: native speakers of American English with no proficiency in German (7 men and 9 women) or with advanced proficiency in German (5 men and 4 women), native speakers of German (7 men and 8 women), and native speakers of French (5 men and 4 women). The native speakers of German and French were all fluent in English.

Instructions were read aloud and presented as a typed cover sheet for the English and German texts as follows:

Our purpose in this study is to learn something about speech pauses. In order to do so, we ask you to help us by locating such pauses and estimating their duration. You will be asked to listen to two recorded

Correspondence should be addressed to D. C. O'Connell, Department of Psychology, Georgetown University, Washington, DC 20057.

Table 1
Descriptive Statistics for English and German Corpora

Statistic	Corpus	
	English	German
Total time (milliseconds)	108,620	115,880
% of offtime (pause time)/total time	22.50	22.85
Speech rate (syllables/second total time)	5.07	4.69
Articulation rate (syllables/second on-time or time of speaking)	6.55	6.09
Mean pause duration (milliseconds)	376	552
Syllables/word	1.17	1.60
Syllables/phrase (per pause)	8.20	11.10
Time (milliseconds)/phrase (per pause)	1,256	1,824
Words/phrase (per pause)	7.00	6.94

conversations—one in English and one in German. You will be provided with a text of each passage typed margin to margin in capital letters, but with no punctuation whatsoever. This text is to be used to record the location and the duration of each pause you notice. Each conversation will be played four times for you. The first two playings are to be used to mark with a slash (/) where you note the occurrence of a pause. A pause is to be marked wherever you note a complete silence. The third and fourth playings are to be used to record your estimation of the duration of each of the pauses you have marked. Please record these to tenths of a second, i.e., to 0.1 sec. For example, you might record 2.3 or 0.1 above a slash. To avoid confusion, we will remind you of your task before each playing. Now we will play a very brief practice passage four times with instructions aloud for each. Then you may ask any further questions you may have. No further questions will be answered once the experiment has started. Thanks ahead of time for your help.

A short example containing an evident pause ("To err is human/to forgive divine") was then presented to demonstrate the four steps. Subjects were run in several groups and were given the corpora in English-German or German-English order. The texts were typed margin to margin without punctuation in uppercase letters on 8.5×11 in. paper.

RESULTS

Since there was no evidence that gender or sequence was significant, data were collapsed over these variables.

For both the English and the German corpora, estimates of pause duration yielded no significant differences across speakers of the respective native languages. The only significant differences involving mean duration of estimates were the t tests for related measures between actual pause positions and false alarm positions. In the English corpus, $t(48) = 3.99, p < .01$ (560 > 409 msec; SDs = 554 and 448). In the German corpus, $t(46) = 3.87, p < .01$ (592 > 453 msec; SDs = 451 and 346).

For the English corpus, comparisons of mean number of actual pauses reported proved nonsignificant across the respective native languages of subjects.

For the German corpus, however, the overall analysis of variance was significant $[F(3,45) = 9.72, p < .01]$. The respective means and standard deviations for Germans, Americans with advanced fluency in German, French, and Americans without any fluency in German were: 24.4 > 20.0 > 15.3 > 12.9 and 6.1, 7.1, 5.1, 4.9. In terms of accuracy percentages, these means trans-

late respectively into 50.8%, 41.7%, 31.9%, and 26.8%. Individual comparisons showed that the means of Americans not fluent in German were significantly different from both the Germans and the Americans fluent in German $[t(23) = 2.97, p < .01; t(29) = 5.80, p < .01]$. The means of the French and Germans were also significantly different from one another $[t(22) = 3.72, p < .01]$. Efficiency percentages were low, but did not vary significantly across language groups; the overall efficiency index was 51.5%.

DISCUSSION

The significant differences between means for positions where there are actual pauses and positions where there are none (false alarms) are hardly surprising. They reflect the fact that the former positions are marked by actual pauses, whereas the latter offer to the subjects no genuinely silent duration to be estimated. The differences for the English and German corpora, respectively, are almost identical: 139 and 151 msec.

The absence of any significant differences for the English corpus across language groups is also quite reasonable. Both the French and the German subjects were students at an American university. Their English fluency, though non-native, was excellent. The situation with the German corpus was entirely different; the Americans without German fluency and the French group could not process the German corpus efficiently. Although the differences in mean estimates of duration across language groups were short of significance, as predicted even for the German corpus, the significant differences in hits (or accuracy) reflect this differential ability to process the language in question.

Specific positions where false alarms clustered are also of interest. For both corpora, positions where six or more false alarms occurred were examined. A number of circumstances were found to be associated with these cases. Some of these positions involved actual pauses of threshold duration (<130 msec). Some were anticipations of nearby actual pauses. In the German corpus, some were associated with laughter, extraneous noise, and/or overlap of speech. And in one instance in the German corpus, it was quite clear that subjects not fluent in the language had become lost in following the text. For example, 5 of the American subjects not fluent in German substituted a false alarm position that read *gestern/ich* for an actual pause position that read *gegessen/ich*.

The argument that accuracy and efficiency of pause reports are diminished by the use of informal spontaneous dialogue as stimulus material can perhaps be reviewed most clearly if we array the overall findings in the extant literature. Table 2 presents a comparison of results from a number of studies; in all of these studies, the subjects were native speakers of American English and the corpora consisted of both English and German speech. Although many other factors are relevant, it is clear from these data that longer mean duration of pauses, slower articulation rate, media quality of speech, monologue, formality, and memorization all contribute to higher accuracy scores, whereas interruption and overlap, laughter, extraneous noise, absence of media-quality recording, rapid articulation, shorter mean pause duration, and (in the extreme case) stimulus impoverishment that excludes an auditory signal altogether, all contribute to lower accuracy scores.

Table 2 reflects only accuracy for subjects who are native speakers of English. Over all subjects, efficiency percentages for the English and German corpora were 74.7% and 51.5%, respectively. The lower efficiency for the German corpus does *not* reflect a lower efficiency contributed to by the subjects without German fluency, but rather the fact that, of the two corpora, the German had far more extraneous noise, laughter, overlap and interruption, slang, and a poorer quality of recording.

In both corpora, longer actual pause durations were underestimated and shorter pauses overestimated. Already in 1890, James (1983/1890) referred to the cross-over point for these estimates as an "indifference point" (p. 581), after Vierordt. In the English and German corpora,

Table 2
Accuracy (Hits/Possible Hits) of American Subjects, Mean Duration (*M*) of Actual Pauses (Milliseconds),
Articulation Rate (AR; Syllables/Second), and Characterization of Genre for English and
German Corpora in Recent Published Research

Research	Corpus	Accuracy	*M*	AR	Genre
Ulasevich et al. (1991)	English	93.5%	1,030	3.20	Poetry readings
	German	82.0%	610	3.47	
Stuckenberg & O'Connell (1988)	English	91.9%	665	5.04	Radio homilies
	German	84.4%	775	5.20	
Gardner et al. (1990)	English	82.5%	469	4.54	Dramatic dialogues
	German	78.5%	646	3.30	
Friedman & O'Connell (1991)	English	64.5%	475	5.82	Radio interviews
	German	29.6%	462	6.02	
Present research	English	64.9%	376	6.55	Informal dialogues
	German	26.8%*	552	6.09	
Monti et al. (1990)	English	40.0%			Unpunctuated printed poetry
	German	21.9%			

*This percentage includes only American subjects *without* fluency in German; those *with* fluency had an accuracy index of 41.7%.

respectively, actual pause durations greater than 820 and 680 msec are underestimated; actual pause durations less than 240 and 280 msec are overestimated. The various estimates of an indifference point given by James all fall within these limits.

Researchers are still being advised to analyze pauses by means of timings "undertaken with a hand held stop watch" (MacWhinney, 1991, p. 250) on the grounds that there can be "97% agreement with the content and segmentation of the original transcript produced by the experimenter" (p. 251). The problem is, however, not at all one of reliability, but of validity; the experimenter too is subject to the same inaccuracies reported in the entire research literature on pause reports and confirmed by the present research. There is no evidence whatsoever that an experimenter is any more accurate or efficient in such crude stopwatch measurements than anyone else. In fact, our evidence regarding the inaccuracy and inefficiency of pause reports indicates that only on-line instrumental measurement, referred to as the "instrumental–perceptual" method by Kowal (1991, p. 39; our translation), can yield the requisite level of accuracy and efficiency.

REFERENCES

CARPENTER, S., & O'CONNELL, D. C. (1988). More than meets the ear: Some variables affecting pause reports. *Language & Communication*, 8, 17-27.

CHIAPPETTA, J., MONTI, L. A., & O'CONNELL, D. C. (1987). Pause perception: Some cross-linguistic comparisons. *Bulletin of the Psychonomic Society*, 25, 103-105.

FRIEDMAN, L., & O'CONNELL, D. C. (1991). Pause reports for spontaneous dialogic speech. *Bulletin of the Psychonomic Society*, 29, 223-225.

GARDNER, J. C., MCMILLAN, H. S., TOWNSEND-HANDSCOMB, D., BARRETT-BATES, R., & O'CONNELL, D. C. (1990). Reporting pauses in dramatic dialogue. *Bulletin of the Psychonomic Society*, 28, 167-170.

JAMES, W. (1983). *The principles of psychology.* Cambridge, MA: Harvard University Press. (Original work published 1890)

KOWAL, S. (1991). *Über die zeitliche Organisation des Sprechens in der Öffentlichkeit.* Bern: Verlag Hans Huber.

MACWHINNEY, B. (1991). *The CHILDES project: Tools for analyzing talk.* Hillsdale, NJ: Erlbaum.

MONTI, L. A., O'CONNELL, D. C., & CHIAPPETTA, J. (1990). "Surmising" pauses. *Journal of Psycholinguistic Research*, 19, 43-55.

STUCKENBERG, A., & O'CONNELL, D. C. (1988). The long and the short of it: Reports of pause occurrence and duration in speech. *Journal of Psycholinguistic Research*, 17, 19-28.

ULASEVICH, A., LEATON, B., KRAMER, D., HIEKE, A. E., & O'CONNELL, D. C. (1991). American and Japanese subjects' reports of pause occurrences and duration. *Language & Communication*, 11, 299-307.