Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5–8 September 2003, Göteborg University, Sweden.
Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 90, ISSN 0349–1021, pp. 13–16.

Fluency: Time for a paradigm shift

Kim Kirsner†, John Dunn† & Kathryn Hird‡

† University of Western Australia
‡ Curtin University of Technology

Abstract

Pauses in spontaneous speaking constitute a rich source of data for several disciplines. They have been used to enhance automatic segmentation of speech, classification of patients with acquired communication disorders, the design of psycholinguistic models of speaking, and the analysis of psychological disorders. Unfortunately, however, although pause analysis has been with us for more than 40 years, their interpretation has been compromised by several problems [6]. The first problem is that the pause distribution is skewed, making mean duration a poor measure of central tendency. The second problem is that there are at least two components to the pause duration distribution, a problem that has been confounded by the fact that most authors have assumed that short pauses can be ignored. The third problem is that many scholars have used an arbitrary criterion to separate the pause components, thereby adopting statistics that reflect errors of commission or omission.

In this paper we review recent work that resolves each of these issues and illustrates the application of the new paradigm to a variety of problems. Our research indicates that, first, there are at least two pause duration distributions, each of which may be sensitive to theoretically interesting variables; second, the distributions are log-normal, thereby opening the way to appropriate measures of central tendency and dispersion, and, third, the distributions can be reliably separated by application of signal detection theory, and the proportion of misclassifications minimised and estimated. This paper reviews recent research using the new approach to pause analysis.

1. Introduction

The objective of this paper is to review problems that have compromised pause analysis, and table provisional solutions to those problems. The first problem concerns the shape of the pause duration distribution. Because the distribution is skewed, it provides a poor platform for conventional statistical analysis. The fact that the pause distribution is skewed was first reported by Quinting [9] however his paper has had little or no impact on pause analyses in either clinical or research work.

A typical pause duration distribution is shown in Figure 1. It shows the pause duration distribution for a 20 minute autobiography by an English first language speaker. PRAAT was used to measure the duration of all pauses greater than 20 msec. The mean, median, mode, standard deviation and range for this distribution are 240, 69, 32, 434 and 20–5156 msec, respectively. The distribution is obviously skewed, and the traditional measures of central tendency and dispersion are therefore inappropriate. The scale of the problem is indicated by the fact that negative numbers are encountered within one standard deviation of the mean. The distribution meets the conditions that Limpert, Stahel & Abbt [8] specified for the use of log-normal procedures; that is, the mean values are low, the variance is large, and values cannot be negative.

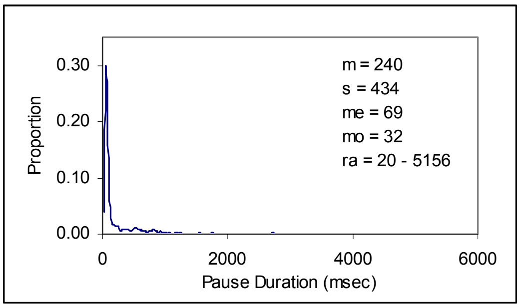

Figure 1: Pause distribution (msec) for 20 minute autobiography from individual participant.

The second problem involves the arbitrary rejection of short pause data in research involving spontaneous speech. This convention was adopted following Goldman-Eisler's seminal work [3], on the basis of which it was argued that ultra-short pauses (below about 250 msec) reflected processes qualitatively different from longer pauses (above 250 msec). The distinction originally involved the contrast between 'articulation' and 'hesitation' pauses [3], and the argument was applied more or less universally despite evidence that the majority of pauses in the 130–250 msec range at least could not be attributed to articulation [5].

The third problem involves the wide variety of criteria that have been used by different authors to identify theoretically significant pause durations. Goldman-Eisler [3] adopted 250 msec as the most appropriate value to separate 'articulatory' and hesitation' pauses, and while this value has proved popular in subsequent research, speech scientists have also used a variety of values ranging from 100 msec to more than one second [7]. For comparative purposes it is imperative that speech scientists adopt a uniform approach to the criterion problem.

A fourth and related problem involves the certainty that each individual will have a unique criterion or, worse, each individual will have a criterion that will actually fluctuate according to topic, task, time of day, age, general health, and neurological status. This problem poses a particularly significant challenge because it can only be answered by adopting measurement procedures that specify the criterion for each individual or, more probably, each speech sample.

The procedure that we have adopted to solve these problems involves two steps. The first step is based on the proposition that log transformations are appropriate for characterising data when distributions are skewed, variances are large, and negative values inadmissible. Figure 2 depicts the pause data from Figure 1 following log transformation (ln) of the original values. The data do not conform exactly to the obvious prediction based on Limpert, Stahel & Abbt [8]. Instead of observing a single log-normal function; the observed pattern involves at least two log-normal functions, a pattern reported independently by Campione & Veronis [1] and Kirsner, Dunn, Hird, Parkin & Clark [6].

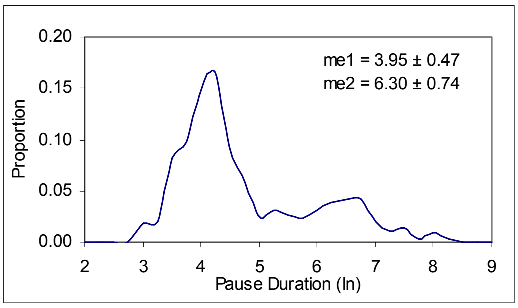

Figure 2: Pause distribution (ln msec) for 20 minute autobiography from individual participant.

The second step involved a modelling procedure supplemented by an application of signal detection theory. The modelling procedure was used to define the log-normal distributions reflected and characterised in Figure 2. As depicted there, the median and standard deviations for the components are 3.95 ± 0.47 and 6.30 ± 0.74. The real values that correspond to these medians are 52 and 545 msec.

Signal detection theory was used to define the criterion where the criterion was chosen so as to minimise the proportion of misclassifications. The criterion for this data set was 4.93 (138 msec) and the proportion of misclassifications associated with this solution was 0.026. Further analysis indicated that the distribution of speech segment durations was also log-normal, and that, when the speech segments were defined by pauses that exceeded 138 msec, the median speech segment duration was 7.04 in log time or 1156 msec in real time.

2. Data

In this section we will present selected results from four experiments involving the data analysis procedures described above. The experiments have been selected to illustrate the value of these procedures for the cognitive, communication and clinical domains, and introduce the mapping procedure that we have used to characterise the short and long pause distributions. Experiments 2, 3 and 4 were implemented in collaboration with Lesley Churchyard, Momoko Taira and Natalie Ciccone respectively.

Experiment 1. Story generation versus story recall. Participants in Experiment 1 provided five three-minute stories about friends or members of their families. PRAAT was used to measure the duration of all pauses greater than 20 msec. Figure 3 depicts the results from just two of these trials, involving generation of one story and the recall of the same story. It was hypothesized that recall would selectively influence the long pause as distinct from the short pause distribution, although we could find no precedent involving this precise manipulation. Figure 3 shows the difference in the medians between recall and generation for short and long pause durations.

The results are consistent with this prediction; while the difference in median long pause duration is generally positive, indicating longer pauses under recall than generation conditions, there is no consistent effect on the difference in short pause duration.

Experiment 2: Fluency in normal and amnesic speakers. The second experiment was originally designed to examine the impact of incidental repetition on word duration during spontaneous speech [10]. The speakers were asked to describe how they would do a number of everyday chores, including for example making a sandwich or changing a tyre.

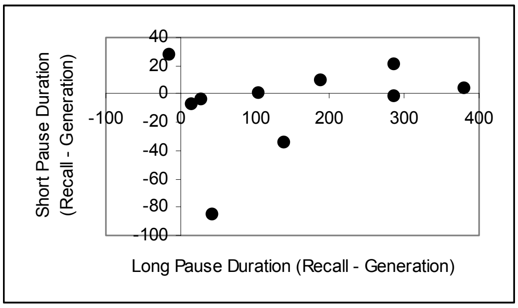

Figure 3: The differences between recall and generation for short and long pause duration.

The procedure did not include questions that would have required the participants to recall specific episodes, and it therefore involves 'implicit' or 'semantic' memory rather than 'explicit' memory.

The speech collected for the original study was re-analysed and PRAAT was used to measure the duration of all pauses greater than 20 msec. The participants were 10 institutionalised amnesic patients, all of whom presented with symptoms consistent with Korsakoff's syndrome, and ten aged matched controls. Figure 4 depicts median short and long pause duration for the participants in the control group and for two of the amnesics. The means and standard deviation are shown for the control group and, while the amnesic values fall well inside 99% confidence intervals for short pause duration, they fall well outside the 99% confidence intervals for long pause duration. It is as if the presence of amnesia has selectively influenced long pause duration in these participants despite the fact that the task involved general knowledge about familiar tasks – a semantic memory task in Tulving's terminology [11] – and did not directly challenge or require the use of explicit retrieval processes, the sine qua non of memory failure in amnesia.

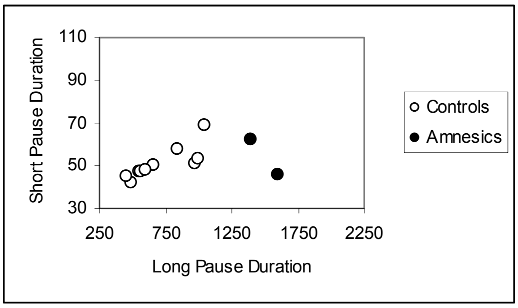

Figure 4: Short and Long Pause Duration for two Korsakoff amnesiacs and ten control participants.

Experiment 3: Fluency in Japanese First Language and English Second Language Speakers (JFL/ESL). The third experiment involved the collection of three 3-minute speech samples from each of 11 JFL/ESL speakers living in Perth, a multi-cultural but predominantly English-speaking community. The second and third samples were in Japanese and English respectively, and involved stories about the participant's favourite holiday destinations, in Japan and Australia, respectively. The results indicated that, overall, the participants had longer short pause duration medians and longer long pause duration medians in English than Japanese, and that each of these effects was statistically significant. Figure 5 is a summary of the results, showing the increase in

Proceedings of DiSS'03, Disfluency in Spontaneous Speech Workshop, 5–8 September 2003, Göteborg University, Sweden.
Robert Eklund (ed.), Gothenburg Papers in Theoretical Linguistics 90, ISSN 0349–1021, pp. 13–16.

the median durations for the short and long pauses for English relative to Japanese.

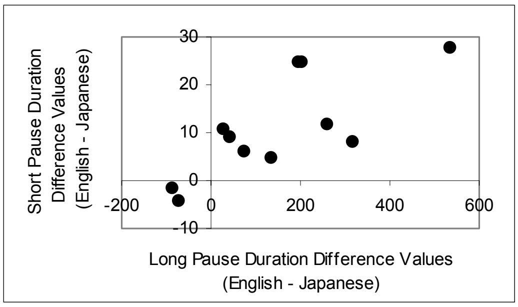

Figure 5: Difference values (English – Japanese) for short and long pause duration in story-telling.

The correlation between the short and long pause duration values observed in Figure 5 was significant, (r (10) = 0.57), but the variables were also related to, 'hours of training and experience in English', indicating that practice in the participant's second language influenced both short and long pause duration. We also found that the participant's had longer median speech segment durations in English than Japanese, at 898 versus 1044 msec, however the extent to which this is due to language differences or practice differences between the speaker's languages cannot be determined from our data.

Experiment 4: Fluency in normal and aphasic speakers. The fourth experiment involved the analysis of speech collected from eight aphasics and 13 control participants. Each person provided four narratives/ descriptions during each of each of eight sessions. PRAAT was used to measure the duration of all pauses greater than 20 msec.

The results depicted in Figure 6 are means based on the medians calculated separately for each individual for each session. The means for the control group are 67 ± 8 and 749 ± 111 msec for short pause duration and long pause duration respectively. The individual vales for the 'Broca' and 'Anomic' patients as classified by the Boston Diagnostic Aphasia Examination are both outside the 99% confidence intervals for the control participants.

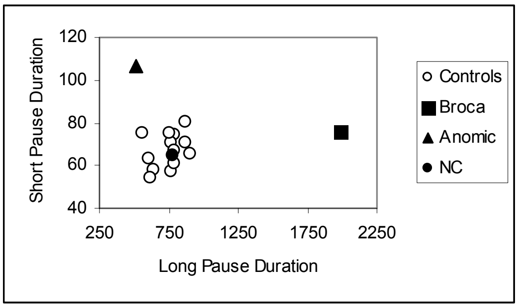

Figure 6: Short and Long Pause Duration for three aphasics and 13 control participants.

Criteria for normal participants. The research reported in this paper was designed in part to overcome the problems associated with the use of different but arbitrarily selected criteria to distinguish different types of pauses.

Figure 7 depicts the criteria for 33 speakers. Twenty of these speakers participated in the memory experiment reported above, and the other 13 were the control participants for the aphasia experiment. The mean for each individual was based on between 600 and 2000 pauses involving between three and eight separate data acquisition sessions. The mean, standard deviation and range for the criteria were 255, 83 and 98 – 490 msec, respectively. The misclassification errors associated with these values ranged from less than one percent to 16 percent. The mean criterion is remarkably consistent with the general criterion advocated by Goldman-Eisler [3], 250 msec (see arrow in Figure 6); although the spread is consistent with our assertion that adoption of a general criterion for all participants is inappropriate.

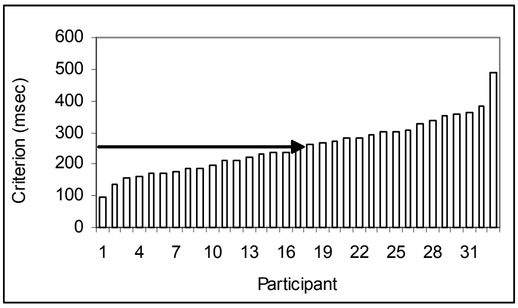

Figure 7: Criteria for 33 normal English speakers.

3. Concluding remarks

While interpretation of double dissociations requires a degree of caution [2], it is nevertheless appropriate to present our results within this frame of reference. What is the relationship between the two pause types? Do they involve independent processes for example, or do they reflect the operation of a single process at two temporally distinct moments in language production, and, if that characterization is valid, do they involve intersecting or non-intersecting sets of variables?

The results of Experiments 1, 2 and 3 are consistent with the hypothesis that the short and long pause duration distributions are functionally independent. Whereas recall instructions and amnesia selectively influence long pause duration, and we found a similar pattern for the Broca's aphasic, anomia selectively influenced short pause duration. On the other hand, the contrast between first and second language fluency was reflected in changes in both short and long pause durations, and individual differences in short and long pause duration were correlated in the memory experiments (in data not summarised above).

There are two classes of explanation for an association between short and long pause duration even if they are functionally independent. First, because both sets of pauses operate through a single and common functional unit [4], the vocal tract, variables that influence this unit are likely to produce correlated changes on each measure. This may be affected by changes in health, emotional status, arousal, tension and, significantly, variables that moderate coordination of the language production system [12]. The second class of variable concerns practice. Practice can be expected to operate on variables such as articulation pauses, speed of articulation, phonological error detection and correction and voiceless transitions, all potentially affecting short pause duration. But practice can also be expected to affect retrieval and implementation efficiency of both syntactic and lexical structures, thus potentially affecting long pause duration.

However, the functional independence of short and long pause durations suggests that they are affected by at least partially independent variables even if these variables are also moderated by higher level variables such as emotion and practice. In addition to the selective effects identified in the first three experiments, it is to be expected that variables such as intention, attention, planning, topic change, and inspiration will selectively influence long pause duration although, until appropriate data is available this hypothesis is speculative.

The implications of our research are as follows. First, the analysis of spontaneous speech requires new foundations involving the use of signal detection or other models to determine individual criteria. Second, the longstanding and widespread disinterest in short pauses must be reversed. Third, answers to questions about the process or processes responsible for short and long pauses are integral to language production, and cannot be treated as if they involve questions separate from models of this domain. Fourth, because each coordination moment provisionally involves information from component processes from different 'domains', their presence challenges modular approaches to language production.

4. References

[1] Campione, E. & J. Veronis. 2002. A Large-Scale Multilingual study of silent pause duration. `http://www.Ipl.univ.aix.fr/sp22002/pdf/campione-veronis.pdf`

[2] Dunn, J. C. & K. Kirsner. 2003. What can we infer from double dissociations? *Cortex*, vol. 39, pp. 1–7.

[3] Goldman-Eisler, F. 1968. *Psycho-linguistics: Experiments in spontaneous speech*. New York: Academic Press.

[4] Gracco, V. L. 1990. *Characteristics of speech as a motor control system. Cerebral control of speech and limb movements*. G. E. Hammond. North Holland, Amsterdam: Elsevier Science Publishers B.V, pp. 3–28.

[5] Hieke, A. E., S. Kowal & D. C. O'Connell. 1983. The trouble with "articulatory" pauses. *Language and Speech*, vol. 26, pp. 203–214.

[6] Kirsner, K., J. Dunn, K. Hird, T. Parkin & C. Clark. 2002. Time for a pause… *Proceedings Ninth International Speech Science Technology Conference*, Melbourne.

[7] Kowal, S., R. Wiese & D. C. O'Connell. 1983. The use of time in story-telling. *Language and Speech*, vol. 26, no. 4, pp. 377–392.

[8] Limpert, E., W. A. Stahel & M. Abbt. 2001. Log-normal distributions across the sciences: Keys and Clues. *Bioscience*, vol. 51, no. 5, pp. 341–352.

[9] Quinting, G. 1971. *Hesitation phenomena in adult aphasic and normal speech*. The Hague.

[10] Robertson, C. & K. Kirsner. 2000. Indirect memory measures in spontaneous discourse in normal and amnesic subjects. *Language and Cognitive Processes*, vol. 15, no. 2, pp. 203–222.

[11] Tulving, E. 1972. Episodic and Semantic Memory. In: E. Tulving & W. Donaldson (eds.), *The Organization of Memory*, New York: Academic Press, pp. 382–404.

[12] Turvey, M. 1990. Coordination. *American Psychologist*, vol. 45, pp. 938–953.