University of Notre Dame Australia
ResearchOnline@ND

Medical Conference Papers

School of Medicine

2005

Language Production: A complex dynamic system with a chronometric footprint

Kim Kirsner

John Dunn

Kathryn Hird
University of Notre Dame Australia, khird@nd.edu.au

Follow this and additional works at: http://researchonline.nd.edu.au/med_conference

Part of the Medicine and Health Sciences Commons

This conference paper was originally published as:
Kirsner, K., Dunn, J., & Hird, K. (2005). Language Production: A complex dynamic system with a chronometric footprint. *7th International Conference on Cognitive Systems.*

This conference paper is posted on ResearchOnline@ND at http://researchonline.nd.edu.au/med_conference/31. For more information, please contact researchonline@nd.edu.au.

Language Production: a complex dynamic system with a chronometric footprint

Kim Kirsner[1], John Dunn[2] and Kathryn Hird[3]

ABSTRACT

In this paper we outline a new approach to the study of language production. Central to this approach is the assumption that communication takes place in a dynamic environment in which cognitive resources are deployed to achieve 'Right-Time' as distinct from 'Fast-as-Possible' solutions. This is based on the assumption that language production includes a single, integrated, interactive process that recruits and coordinates information from a variety of internal, external and interactive sources to build each speech segment. The output of this process is reflected in the longer of the two log-normal pause duration distributions observed in spontaneous speech (Kirsner, Dunn, Hird, Parkin & Clark, 2002). The methodology described here permits the inspection of temporally defined processes under natural speaking conditions. The procedures do not rely on the assumption that language is the product of independent components that can be studied under static, de-contextualised conditions. Results from aphasia, amnesia and bilingualism will be used to illustrate the new paradigm.

Key words: language production, speaking, memory, amnesia, aphasia, dynamic systems, modularity, pause duration, segmentation, natural language

1 Is decomposition sufficient?

Decomposition and modularity have played a central role in the cognitive and neural sciences for several decades. However the paradigm associated with these tools cannot easily be refined to handle cognitive processes that unfold in time, particular when the relevant process involves coordination across a range of biological, physiological, cognitive and linguistic processes. The temporal distribution and management of spoken language is the central challenge. In this article we discuss some of the problems associated with approaches based on decomposition and modularity, and table a class of model that might be required to describe and predict spontaneous language.

Although the notion of modularity can be traced back to Hughlings Jackson's early writing, and decomposition is *the* modal paradigm for the cognitive sciences (Arbib, 1989; Marr, 1982), there is no lack of debate about the limitations of paradigms that rely on these concepts. The case for decomposition and modularity is based in part on the growing body of evidence of specific associations between cognitive function and cortical localization. However, in the last analysis, the success of this program depends on agreement about the inferential procedures that can be used to define the taxonomy (e.g., Dunn & Kirsner, 1988), and the taxonomy per se, and agreement has not been achieved.

Perhaps the most compelling problem involving modularity concerns the scale of the hypothetical modules. Are we dealing with macro-modules that involve vast *systems* such as language and memory, or are we dealing with micro-modules that involve grammatical functions, individual concepts, or even neurons? The problem of scale also besets research into localization in the brain. The available techniques measure spatial regions that range from less than 1 mm to the whole brain (and cover 7 log units) for temporal periods that range from milliseconds to days (and cover 10 log units) (see van Horn, 2002). Which scales involve modules?

The uncertainty associated with scale is clear in Levy's (1996) rejection of 'big' modularity (i.e., modularity of the language family) as an explanation of childhood communication disorders, and her preference for "accessing privileges" over 'small' modularity in her account of those disorders. Argument that 'big' modularity has been successful because text book topics have been relatively stable for 40 years (e.g., Hubbard, 2002) have no value in this debate, a position we share with Uttal (2001). But how can we

[1] School of Psychology, University of Western Australia, Australia WA 6009, kim@psy.uwa.edu.au
[2] School of Psychiatry and Clinical Neuroscience, University of Western Australia
[3] School of Psychology, Curtin University of Technology

adopt a definition of modularity unless we agree on the scale of the module? Sperber (2002) adopted an inclusive approach to the problem. According to Sperber (2002, with reference to Sperber, 1994), "I was arguing that domain-specific abilities were subserved by genuine modules, that modules came in all format and sizes, including micro-modules the size of a concept, and that the mind was modular through and through". Clearly, such a catholic definition can only be embraced if domain-specificity is discarded.

A related and more public problem involves the definition of modules. Fodor (1983), for example, provided a list of features that might be used to characterize modules, a list that comprised: domain-specific, innately specified, informationally encapsulated, fast, hardwired, autonomous and not assembled. However, Coltheart (1999), asserted that Fodor did not intend his list to be treated as a 'definition', and sought to restrict consideration to just the first of these features. According to Coltheart (1999) a module is a "cognitive system is whose application is domain specific, where domain specific is met if and only if it only responds to *stimuli of a particular class* - face, word, object or voice" (italics supplied), a claim which echoes Arbib's earlier argument about "gross modalities" (Arbib, 1989). However, while "faces, words, objects and voices" obviously involve *stimuli of particular classes*, and provide a solid starting point even if we have to overlook the fact that curves are present in objects as well as letters (Arbib, 1989), it is evident that Coltheart (1999, Figure 1 and accompanying text) wants to us to assume that shape-specific letter representations, abstract letter representations, and phonemic and lexical codes also constitute distinct domains even though they do not involve *stimuli of particular classes*, or even stimuli if it comes to that. The modularity debate is far from resolved, and it is evident that Fodor too is far from certain (Fodor, 2000).

Does *informational encapsulation* provide a better platform for modular definition? According to Fodor (1983), modules are encapsulated from very particular kinds of information, involving beliefs, desires and utilities, a definition which explicitly limits the role of high level processes, and provides a platform for the assertion that explanations of communication disorders must be restricted to proximal as distinct from distal variables (Jackson & Coltheart, 2002). However, Marslen-Wilson and Tyler (1989) argue that the "immediacy with which meaning as distinct from linguistic form becomes available" compromises the claim that linguistic processing is modular. A similar problem is present in repetition priming in word recognition. Although repetition priming is form specific under de-contextualized conditions, when the influence of expectations is precluded, the situation is changed dramatically when a context is created, and expectations are allowed to work their magic. For example, although presentation of the word BANK leads to significant facilitation when that word is repeated an hour or two later during the same experiment, this can be limited or eliminated by manipulation of the context. For example, is BANK is preceded by MONEY and RIVER oh the first and second occasions, respectively, facilitation would be absent, although it would be present if MONEY or RIVER were to be used prior to the first and second presentations (Bainbridge, Lewandowsky & Kirsner, 1993). Lexical analysis is therefore anything but domain-specific or informationally encapsulated under these conditions. On the contrary, it is sensitive to the expectations of the reader.

It is not our intention to mount an argument for or against modularity however. We have a more limited objective. It is our contention that a paradigm based on decomposition and modularity cannot provide a *sufficient* account of language production. The critical issue concerns those cognitive contributions which *cannot* be characterized or measured by procedures which involve the presentation or emission of isolated words or even sentences. The most important of these contributions involves the temporal coordination of language production during spontaneous speaking. Tasks that involve the presentation and production of single words cannot provide conditions under which the full range of biological, physiological, cognitive and linguistic processes is required, and temporal coordination cannot therefore be assessed. This is not simply a matter of context. Rather, it concerns the way in which we create and manage speech, silence, prosody and gesture during spontaneous speaking.

The critical issue involves *coordination*. There is no modular account of the temporal organization of language production, and it is difficult to anticipate any way in which the static and de-contextualized tasks that constitute the bread and butter of modular research could be refined to measure and evaluate the processes responsible for the temporal management in spontaneous speaking. Spontaneous speaking can be measured only by measures that involve and reflect spontaneous speaking per se.

Essentially the same issue has been raised in a rather different form by Port and Leary (2002). According to Port and Leary, the gulf between descriptive psycholinguistics and the temporal realities of speech is unbridgeable, and they advocated a new discipline dubbed “embodied linguistics” to meet the challenge. A related but rather different type of challenge has emanated from the work on complex dynamic systems. According to van Gelder (1996), for example, complex behaviour needs to be defined in terms of transitions and dependencies, and it is precisely these characteristics that are not offered by de-contextualised paradigms. Yet another but related challenge can be discerned in recent work on motor systems (e.g., Gracco, 1990). The central issue in arguments about motor planning and execution is that there are so many muscle groups and contingencies in the final common pathway for speech that it is impossible to model it without treating it as a single functional unit in which inter-dependency and interaction are the rule rather than the exception. Our approach is consonant with these arguments. In the balance of this article we will provide a provisional description of a dynamic system that might underpin natural language production.

2 Beyond modularity

Our provisional solution to this problem involves a proposal advanced by Bechtel (1997). Bechtel, like Port and Leary (2002), accepted the critical assumption that cognitive events unfold in time, and that new models must be developed to connect the component process to the needs of the overall system. Our approach involves the performance of the overall system, and the language production system in particular. The architecture adopted by us is in some respects similar to that adopted by Bechtel, with provision for both dedicated sub-systems and a coordinating mechanism that unfolds in time. For characterization however, we are going to avoid the notion of modularity altogether, and refer to the specialist processes as *advocates*. It is our assumption that dedicated advocates accumulate information associated with specific real-time needs, and that they then act as advocates for those needs in an interaction involving the entire set of advocates. An illustrative list of the needs associated with pause duration might include the following: breathing, intention and its corollary, message planning, retrieval of information from secondary and semantic memory, discourse status including the discourse records of the speaker and interlocutor, lexical search and syntactic and phonemic construction. The parallel with Selfridge’s pandemonium model (Selfridge, 1959) is deliberate. Each of the advocates can be thought of as a demon demanding that its needs be met.

For illustrative purposes, and to provide a point of departure for debate, it is hypothesized that, first, each *advocate* tables a specific request in the form of a value for the next pause and, second, the duration of the forthcoming pause is the product of the values advanced by the complete set of advocates. For example, if the breathing, intention, message planning, memory retrieval, discourse status, lexical, syntactic and phonemic advocates demanded 3, 5, 2, 2, 5, 4, 1 and 2 units of energy respectively (where the minimum and maximum values are 1 and 5), the product would be 2400. The absolute values are of no interest of course. The critical issue concerns the distribution produced by this equation. If it is assumed that the requested values are determined independently and randomly for each trial, the resulting distribution for a set of pauses would be log-normal. The first empirical question therefore concerns the pause duration distribution observed in spontaneous speaking.

3 Measurement issues

The measurement and characterization of pause duration distributions is not a simple issue. There are several questions.

3.1 Skew

The first problem involves the fact of the skew in the pause duration distribution in spontaneous speaking. We have been able to discover just three articles or chapters published prior to 2002 which recognized that the pause duration distribution is skewed. The first of these involved a chapter by Jaffe and Feldstein (1970, Ch 4). The chapter includes log frequency distributions for pause duration for individual speakers, vocalization duration for individual speakers, and switching duration for pairs of speakers. The functions are approximately linear, indicating that the underlying distributions are approximately lognormal. Jaffe and Feldstein were concerned with a theory of dialogue however, and they gave little consideration to the implications for psycholinguist research, and their observations about the shape of the pause duration distribution may have been overlooked in subsequent work on pause duration for this reason. The second report, a monograph by Quinting (1971), included descriptive statistics for eight aphasic participants and eight non-brain damaged participants. The average (standard deviation) of the mean, median, mode and standard deviations for the pause duration distributions for these participants were as follows: Mean = 1.22 sec (±0.35), Median = 0.87 sec (±0.21), Mode = 0.26 sec (± 0.06) and the standard deviation = 2.03 (±0.51). Two points stand out. First, the mode is actually identical to the recording limit adopted by them (i.e., 0.26 sec), an outcome which implies that a substantial fraction of pauses was overlooked by their procedure. Second, because the standard deviation (i.e., 2.03 sec) is actually greater than the arithmetic mean (i.e., 1.22 sec), they are dealing with a strongly skewed distribution. Indeed, the distribution is so skewed that the entire body of subsequent research that has relied on the arithmetic mean must be insecure.

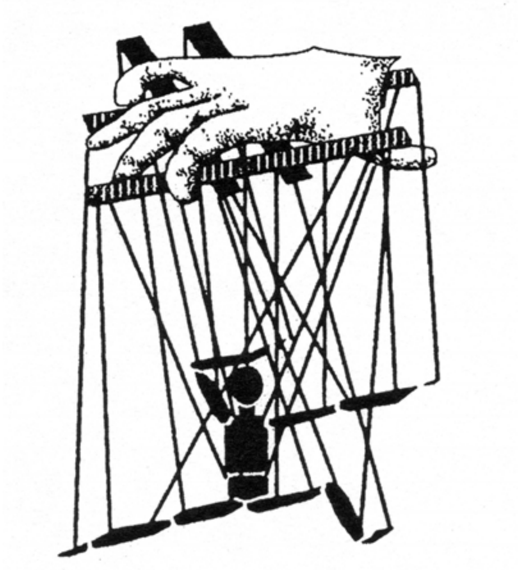


Figure 1A: Puppet controlled by independent components (figure from Turvey, 1990)

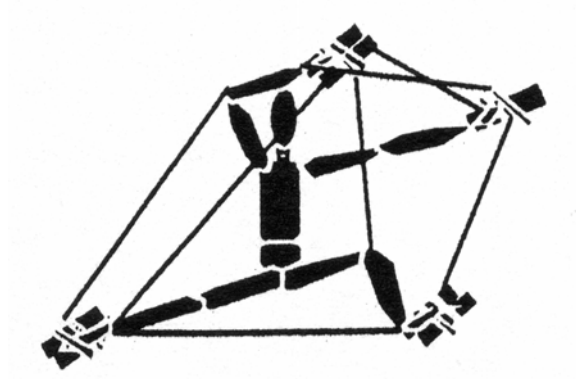


Figure 1B: Marionette involving Dynamic Interactive System (figure from Turvey, 1990)

3.2 Boundary Threshold

The second problem involves the Boundary Threshold, and the distinction between what we refer to as the *Boundary* and *Distribution Thresholds*. The *Boundary Threshold* is the shortest value classified as a pause. In some cases this value reflects the limits of the recording or segmentation system. Jaffe and Feldstein (1970) for example, were sampling vocalization at a rate of only 200 samples per minute, and the pause duration distributions described by them reflected this limitation. However in many more recent cases, rejection of pause durations below 250 msec reflects an arbitrary decision that such pauses are irrelevant to a particular research objective. Goldman-Eisler (1968), for example, advocated adoption of 250 msec as the criterion because this would in effect separate two qualitatively different types of pauses, ‘articulation’ pauses that are assumed to be shorter than this value, and ‘hesitation’ pauses that are assumed to be longer than this value. Goldman-Eisler assumed furthermore that only psycholinguistic and cognitive pauses were of interest, and she therefore discarded all pauses shorter than 250 msec. Goldman-Eisler’s solution and rationale have been widely adopted since 1968 although, in practice, many scientists have adopted their own criteria (see Hieke, Kowal and O’Connell, 1983 for a review), a procedure that has hindered comparative analyses.

In our research we have generally adopted a *Boundary Threshold* of approximately 20 msec, and assumed that all longer pauses are of interest to the acoustic analysis of speech. Even this value overlooks some ultra-short pauses, but the cost of segmentation rises dramatically below 100 msec.

3.3 Presence of two pause types and two pause duration distributions

The general claim that there are two pause types is not new. Lounsbery (1954), for example, drew a distinction between ‘juncture’ pauses and ‘hesitation’ pauses, and put the threshold at 100 msec. According to Lounsbury, junction pauses serve as aids to the listener by helping to put across the structure of the sentence whereas hesitation pauses reflect weak associations between linguistic events, and mark the beginnings and ends of speaker units. Goldman-Eisler (1958, p99) drew a similar distinction, between phonetic and hesitation pauses. According to Goldman-Eisler,

> "Greater precision of measurement (with reference to the fact that her recorder was accurate to only 100 msec) would bring into the range of distinction a different level of speech production, namely the phonetic one. (Pauses up to 0.20 or even 0.25), though rare, might occur as part of ritardando effects, or articulation shifts or between plosives.) To be quite certain that these were not included, the gaps in the visual record classified as pauses had to be equivalent to durations of not less than 0.25 seconds."

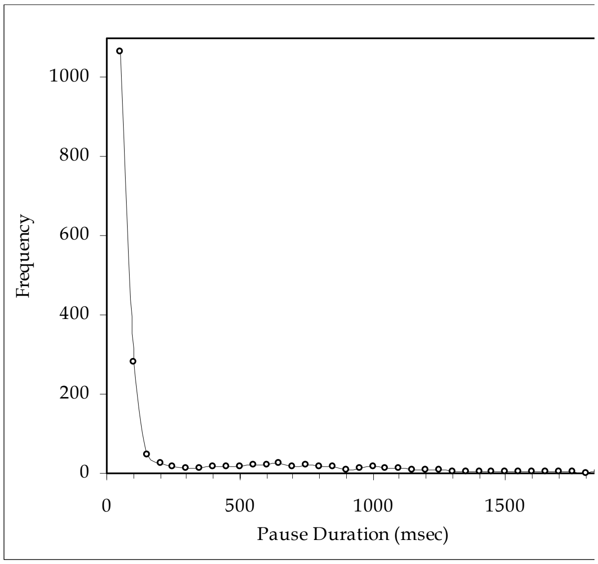


Figure 2A: Sample of pause duration distribution for participant PC: Mean = 270, Standard Deviation = 926, Median = 42, Mode = 21, Range = 16 – 14670, and number of observations = 1813. The x-axis was truncated (to 2 sec) to facilitate depiction.

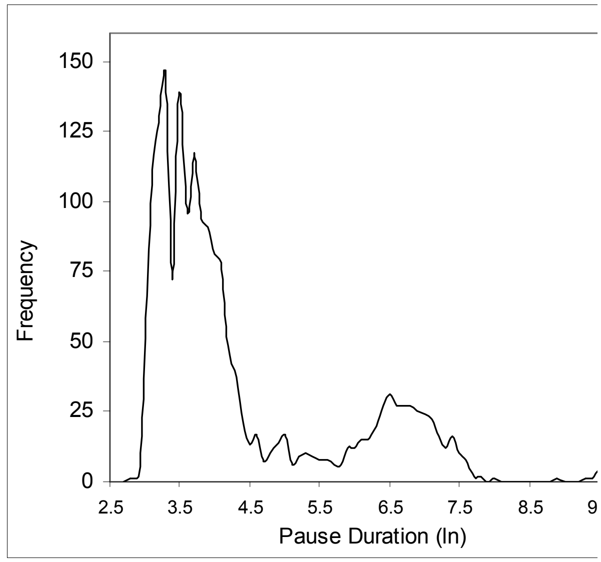


Figure 2A: Pause duration distribution for participant PC following log transformation: Mean for short pause distribution = 3.53 (34 msec) ± 0.41; Mean for long pause duration distribution = 6.34 (565 msec) ± 0.90.

However, the procedure adopted by Goldman-Eisler (1968) and many subsequent researchers involved several interesting assumptions. The first assumption is that there is qualitative change in pause type on the pause duration continuum between 20 msec and infinity. The second assumption is that the transition between pause types occurs at or near the 250 msec boundary identified by Goldman-Eisler (1968). Intriguingly, scientists have actually adopted a variety of values from 100 msec to 1000 msec or more (see Heike, Kowal & O'Connell, 1983), thereby compromising comparisons between studies. The third assumption involves the stability and generality of the point of transition. Is this a stable and universal characteristic, or does it depend on variables such as age, health, personality, mood and cognitive efficiency[4]. Figure 2A depicts the results for a single speaker. Participant PC was asked to provide an autobiography. She talked for 21: 47 minutes, and provided a virtual monolog. Application of a *Boundary Threshold* of approximately 20 msec yielded a total of 1813 pauses; that is, 1.39 pauses of all durations per

[4] We have adopted the terms *Boundary Threshold* and *Distribution Threshold* to identify the smallest measurable pause and the point of transition between the short and long pause types respectively. Because Goldman-Eisler's (1968) rejected all values below a single threshold, these functions are conflated in her work.

minute. The figure depicts frequency values for every 50 msec step from 0 to 2000 msec[5]. The figure also includes the mean, standard deviation, median and mode for participant PC. The mean is of course far lower than that reported by Quinting (1971) but that is not surprising as he used a *Boundary Threshold* of 250 msec. By eye, the break in the function occurs at approximately 150 msec rather than 250 msec.

3.4 Shape of pause duration distributions

Figure 2B depicts the same data set following lognormal transformation of the individual values. The presence of two component distributions is now evident. It is also evident that each of the distributions is approximately lognormal. Application of a signal detection model located the *Boundary Threshold* at 4.52 (92 msec), a value that yielded an estimated *Misclassification Rate* of 0.6%. The *Distribution Threshold* values for five other participants in the same study ranged from 3.87 (48 msec) to 4.93 (138 msec) while the misclassification rates for the same participants ranged from 1.3% to 4.4%. The results from PC and our other participants are consistent with the proposition that there are two pause types in spontaneous speaking, and with the further assumption that they each follow the lognormal. But they also show that the *Distribution Threshold* is anything but universal, and differs from speaker to speaker.

The finding that the pause duration distribution is approximately lognormal appears to have been re-discovered in three independent laboratories, in Aix (Campione & Veronis, 2002), Perth (Kirsner, Dunn, Hird, Parkin & Clark, 2002) and Madison (Rosen, Kent & Duffy, 2003). Significantly, none of these groups appear to have been aware of the earlier report by Jaffe and Feldstein (1970). The lognormal observation made by Jaffe and Feldstein is particularly interesting because they used a *Boundary Threshold* (as defined by the limitations of their system) of about 300 msec, and therefore observed a lognormal distribution for values greater than 300 msec. However, where characterization of the pause duration distributions is concerned, the level of the *Boundary Threshold* is critical. If the *Boundary Threshold* is equal to or greater than the *Distribution Threshold*, the shape, mean and standard deviation of the short pause distribution will be severely compromised. Thus, as Campione and Veronis (2002) used a *Boundary Threshold* of 60 msec, and we are observing *Distribution Thresholds* as low as 48 msec, their claim that the median of the short pause distribution is in the vicinity of 100 msec is suspect. In our case, furthermore, the means (in log) of some of the short pause distributions are actually lower than the *Boundary Threshold* adopted by Campione and Veronis (2002). It is evident then that effective parameter estimation requires the use of ultra-low Boundary Thresholds, at 20 msec or less.

3.5 Separation and characterization of the short and long pause duration distributions

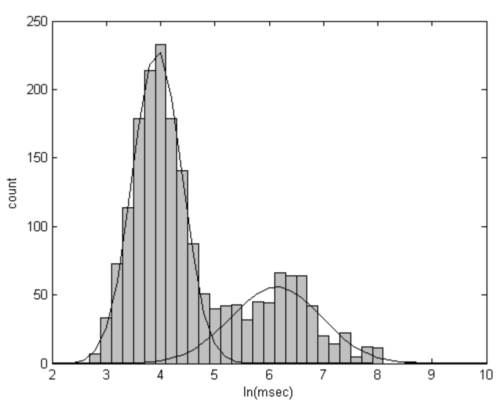


Figure 3A: Pause duration distribution and model for participant XX

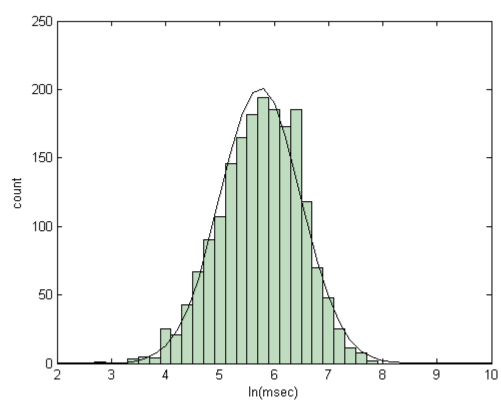


Figure 3B: Speech segment duration distribution and model for participant XX

[5] The range is actually 16 to 14670 msec but the tail of the function cannot be seen with the naked eye if the abscissa covers the full range.

Another problem concerns the design and implementation of a procedure to separate and characterize the short and long pause distributions. If it is assumed that pause duration involves a mixture of two log normal distributions, and that these distributions overlap to some extent, the procedure should be designed to minimize total misclassification rate; that is, the proportion of short pauses classified as long pauses, and the proportion of long pauses classified as short pauses. We have used the Expectation-Maximization Algorithm (McLachlan & Peel, 2000) to estimate the means, standard deviations, and the relative proportion of each distribution. The two best-fitting normal distributions (in log-time) for another participant are shown in Figure 3. RB, the participant in question was a member of the non-brain damaged control group in a study of people with acquired neurogenic communication disorders. The model distribution of short pauses has a mean of 4.18 (66 msec in real time), a standard deviation of 0.55, and a relative probability of 0.69 (i.e., short pauses constitute 69% of the total distribution of pause duration). The model distribution of long pauses has a mean of 6.66 (785 msec in real time), a standard deviation of 0.64, and a relative probability of 0.31 (i.e., long pauses constitute the remaining 31% of the total distribution).

Finally, once the component distributions have been fit to the data, it is possible to estimate an optimal *Distribution Threshold* for each data series. This problem is very similar to classical signal detection theory (Green & Swets, 1966) and we define the *Distribution Threshold* as the value on the log-time axis that minimizes the total expected misclassification rate. This is defined as the sum of the area under the model short pause distribution to the right of the threshold and the area under the model long pause distribution to the left of the threshold. The former estimates the rate at which short pauses are misclassified as long pauses, while the latter estimates the rate at which long pauses are misclassified as short pauses. For Participant RB, the optimal *distribution threshold* was 5.48 (240 msec in real time), and the total expected misclassification rate was 1.6%

3.5 Speech segment duration

Speech comprises alternating periods of speech and pause. However, it is now clear that there are two forms of pause; short pauses which last for 50 -70 msec, and long pauses which last for 500 – 700 msec. How should the pause data be used to define the speech segments? Should we include the speech segments that separate each and every pause, or should consideration be restricted to the speech segments which occur between the long pauses? The procedure advocated by Goldman-Eisler (1968) inadvertently followed the second of these procedures, and pauses of less than 250 msec were excluded absolutely. In effect, she regarded the short pause type as a part of speech rather than a pause per se. It is possible to extend this concept, and assume that the energy level associated with natural speech involves a wide range of amplitudes, some of which fall below the level of ambient noise. They are classified as short pauses under our procedure but they never-the-less form a part of the speech signal rather than a 'genuine' pause. We have proceeded on the assumption that this is correct, treated short pauses as if they are part of the speech signal, and defined speech segment duration as the period between consecutive long pauses. Figure 3B is the speech segment duration distribution following log transformation for a single participant. The function is approximately log normal.

4 Empirical Issues

Ten parameters are relevant to the measurement of 'fluency'; the threshold (defines point of optimum separation between the short pause and long pause duration distributions), proportion misclassification (provides an estimate of the extent of overlap and therefore misclassification between the short pause and long pause distributions), the means, standard deviations and occurrence rates for the short pause and long pause distributions; and the mean and standard deviation of the speech segment duration distribution. Consideration will be given to four general issues.

4.1 Is system performance sensitive to social and situational variables?

One of the claims advanced for modular systems is that consideration is focused on if not restricted to 'proximal' variables such as word type rather than 'distal' variables such as social context (Jackson & Coltheart, 2002). This question can be considered with reference to a variety of situational questions. Two situations will be considered here. The first of these involves the contrast between the situational conditions faced by politicians. We compared the performance of politicians under two conditions: public speaking as manifest in formal speeches, and public speaking as required under interview conditions. A sample of the results is shown in Figure 4A. The figure shows that one parameter, mean long pause duration, is sensitive to this manipulation; mean long pause duration was longer for all 12 of our participants under public speaking conditions. The mean difference approached 300 msec in real time. Another situation of interest involves emotion. Undergraduates were asked to describe events that involved happy, funny, sad and frustrating circumstances. The results are depicted in Figure 4B. Mean long pause duration was significantly longer when our participants were asked to talk about sad events than happy, funny or frustrating events.

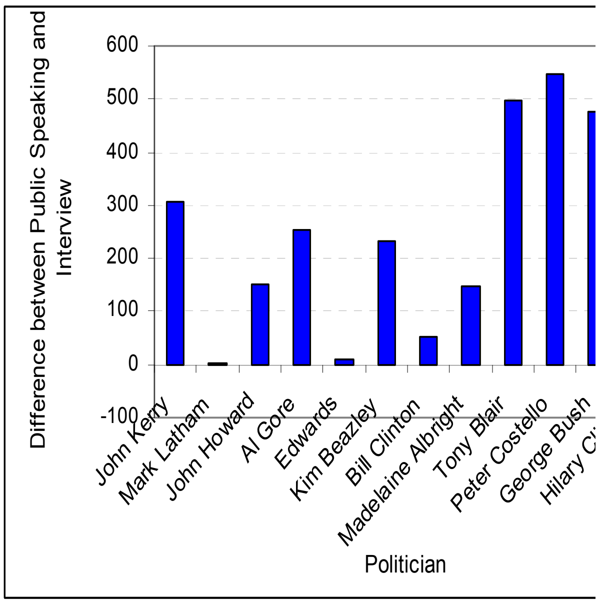


Figure 4A: Impact of speaking context on mean long pause duration I: The ordinate depicts the difference between interview and public speaking conditions.

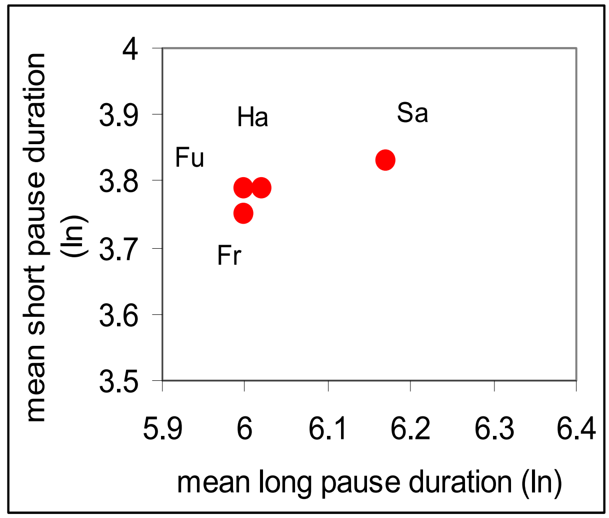


Figure 4B: Impact of speaking context on mean long pause duration II: The impact of emotion on mean short pause and mean long pause duration

4.2 Is system performance sensitive to brain damage involving acquired or degenerative disorders?

We will now turn to variables which influence 'proximal processes'; that is, the basic cognitive operations involved in the design, preparation and execution of natural language. Figure 5A depicts the impact of various forms of aphasia on mean long pause duration. Consideration will be restricted to one participant, GJ. GJ was classified as an Broca's aphasic according to the Boston Diagnostic Assessment Examination. As shown in Figure 5A, mean long pause duration for this participant exceeds the 95% confidence interval, an outcome that is consistent with the assumption that this type of disorder is characterized by nonfluent speech, few words, short sentences, and many pauses.

Figure 5B depicts the impact of Korsakoff's amnesia, an organic disorder associated with chronic alcoholism, on mean long pause duration. The figure depicts the results for nine control participants and seven amnesics, where the amnesic participants were classified as such by reference to conventional memory tests. The figure shows that mean long pause duration under natural speaking conditions falls beyond the 95% confidence interval for the control group. The actual tasks used to elicit natural speech actually involved a request to provide a 'procedural description' of routine tasks such as changing a tyre and making a sandwich.

5 Theoretical considerations

5.1 A provisional hypothesis

As discussed in the introduction one interpretation of the lognormal distributions observed in language production is that they reflect a multiplicative interaction among some or all of the many variables tugging at the language production system (e.g., Limpert, Stahel & Abbt, 2001). This idea was captured pictorially in Figure 1B. In our case it is hypothesized that the process which determines long pause duration as well as the other language production parameters reflects a multiplicative interaction involving not only the cognitive and linguistic variables listed above but physiological variables such as breathing. The argument constitutes a simple hypothesis about the dynamics of language production. While it might be possible to develop compromise models of language production involving both modular and dynamic stages or components, the onus is on scientists who choose to advance a modular viewpoint to explain the log normal distributions observed for long pause duration in particular.

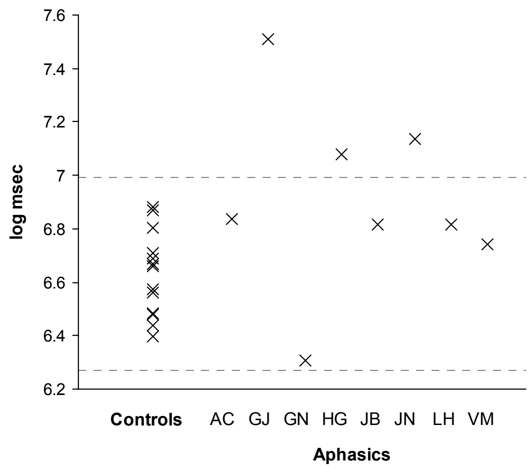


Figure 5A: Mean long pause duration for a control group and for 8 participants with aphasia. The bars indicate 95% confidence intervals for the Control group. Participant GJ was classified as a 'Broca's aphasic.

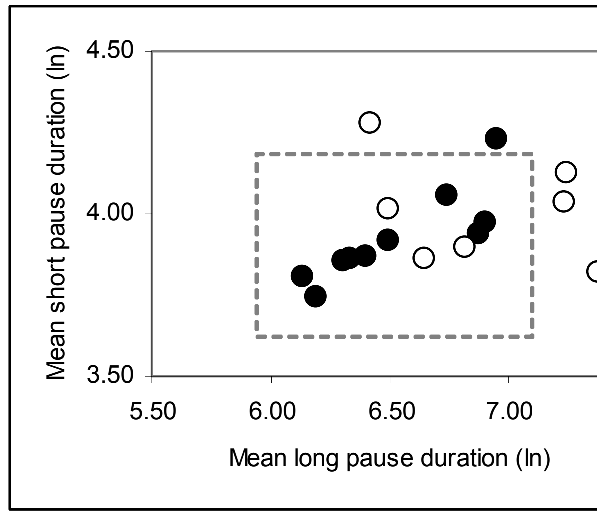
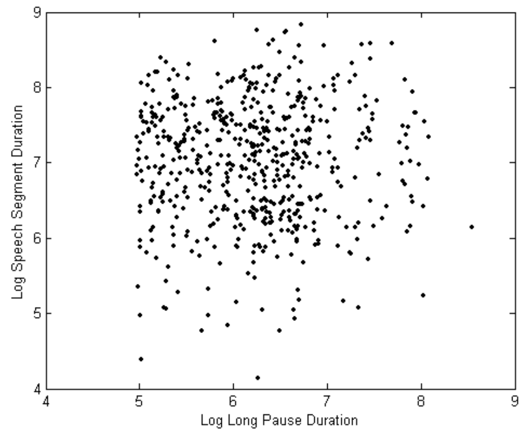


Figure 5B: Relationship between the duration of successive pause and speech periods in a 20 minute autobiography

5.2 The role of silence in speech

What is the role of 'silence' in natural language? If it is assumed that mean long pause duration is a measure of silence, the obvious interpretation is that it is a measure of thought or, more narrowly, cognitive work. There is a large field of cognitive claimants for this role. The field includes the preparation of 'intentions', development of a speaking plan, attention to monitoring requirements, selection of lexical items, preparation of a syntactic scaffold for the forthcoming utterance, and the construction and maintenance in memory of an appropriate phonological string. The last two of these variables at least can be used to predict a positive association between mean long pause duration and mean speech segment duration. Figure 6A was prepared to test this prediction. It was our expectation that there would be a positive association between variables. Our prediction was not supported; it is evident that the two variables are unrelated, an outcome that lends no support to the assumption that pause duration is determined by cognitive work.

Figure 6B provides information about the dynamic range of normal speech. The points in the centre of the figure indicate mean short pause duration and mean long pause duration for 13 non-brain damaged speakers. Some of the aphasics and amnesics that we have tested fall outside the range for normal speaking for the indicated variables, and for other variables such as speech segment duration. The ovals depict the approximate 95% 2-dimensional confidence intervals for the same 13 speakers, and each point therefore serves as a mid-point for one oval. The range of variation in regard to the short and long pause duration axis is considerable. The range depicted below covers the 95% confidence intervals for all speakers. The range for short pause duration is from 4.03 log msec (56 msec) to 4.42 log msec (83 msec). The equivalent value for long pause duration is from 4.37 log msec (79 msec) to 8.60 log msec (5432 msec). Presumably we are capable of processing and understanding information despite the way in which it is distributed in time.



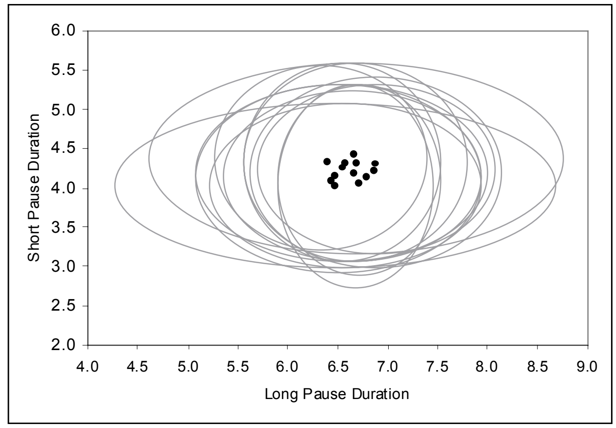


Figure 6A: The relationship between the duration for each speech segment (y-axis) and for the long pause which immediately preceded each segment (x-axis). The data is from a 20-minute autobiography from a student and involves about 1800 pause – speech pairs.

Figure 6B: Individual speaking patterns for 13 non-brain damaged speakers. The points represent depict the means in a two dimensional space for each speaker. The ovals depict the 95% confidence intervals in a two dimensional space for each speaker.

6 Concluding remarks

The findings summarized above indicate that the temporal dynamics of language production are sensitive to a wide range of variables. A quick list includes physiological variables such as breathing variables, social variables involving audience and situational characteristics, cognitive variables such as event reconstruction and lexical search, linguistic variables such as syntactic construction. We have hypothesized that the language production parameters described in this article reflect a complex dynamic system that depends on some of interaction between these and other variables. Another to way handle this process is to assume that it is dominated by *right time* principles. Unlike performance in so many of the experimental tasks devised by cognitive scientists, where fast-as-possible processing is required, communication is the objective of communication, and people adapt and modify pause duration to achieve that objective, and that they do so in a way that is indifferent to fast-as-possible principles. The precise nature of the system that manages this process is unclear. But a complex dynamic process is one of the obvious candidates.

REFERENCES

Arbib, M. (1989). Modularity and interaction of brain regions underlying visuomotor coordination (p333 - 364). In Garfield, J.L. (Ed.). Modularity in knowledge representation and natural language processing. Cambridge. MA. MIT Press.

Bainbridge, J.V., Lewandowsky, S., & Kirsner, K. (1993). Context effects in repetition priming are sense effects. Memory & Cognition., 21, 619-626.

Bechtel, W. (1997). Dynamics and Decomposition:Are They Compatible? Proceedings of the Australian Cognitive Science Society.

Campione, E. & Veronis, J. (2002). A Large-Scale Multilingual study of silent pause duration. http://www.lpl.univ-aix.fr/sp2002/pdf/campione-veronis.

Coltheart, M. Modularity and cognition. Trends in Cognitive Sciences, 1999, 3, 115-120.

Dunn, J.C., & Kirsner, K. (1988). Discovering functionally independent processes: The principle of reverse associations. Psychological Review, 95, 91-101.

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly Journal of*

Experimental Psychology, 10, 96-106.

Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech.New York. Academic Press.

Gracco (1990).Characteristics of speech as a motor control system. In G.E. Hammond (Ed.), *Cerebral control of speech and limb movements*. Elsevier Science Publishers B.V. North Holland.

Green, D. M. and Swets, J. A. *Signal detection theory and psychophysics*. New York: Wiley, 1966.

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The trouble with "articulatory" pauses. *Language and Speech, 26*, 203-214.

Jackson, N. E. & Coltheart, M. (2002). Distinguishing proximal from distal causes is useful and compatible with accounts of compensatory processing in developmental disorders of cognition Behavioral and Brain Sciences, 25, 758-759.

Jaffe, J., & Feldstein, S. (1970). Rhythms of dialogue. New York: Academic Press.

Kirsner, K., Dunn, J., Hird, K., Parkin, T. & Clark, C. (2002). *Time for a pause*. Paper presented at the 9th Speech Science Technology Conference, Melbourne.

Levy, Y. (1996). Modularity of Language reconsidered. Brain and Language, 55, 240-263.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and Clues. *Bioscience, 51(5),* 341-352.

Lounsbury, F.G. (1954). Transitional probability, linguistic structure, and systems of habit-family hierarchies. Osgood, C. & Sebeok, T.A.(Eds.), Psycholinguistics: A survey of theory and research problems (pp93-101). Baltimore: Waverley Press.

Marr, D. (1982). Vision. Freeman.

Marslen-Wilson, W. & Tyler, L.K. (1989). Against modularity (p37-62). In Garfield, J.L. (Ed.), Modularity in knowledge representation and natural language processing. Cambridge. MA. MIT Press.

McLachlan, G.J. & Peel, D. (2000). Finite Mixture Models. New York. Wiley.

Port, R.F. & Leary, A.T. (2000). *Speech Timing and Linguistic Theory*. http://www.cs.indiana.edu/hyplan/port/pap/PAPER.feb.18.port.leary.pdf

Quinting, G. (1971). *Hesitation phenomena in adult aphasic and normal speech.* The Hague.

Rosen, K.M., Kent, R.D. & Duffy, J.R. (2003). Lognormal distribution of pause length in ataxic dysarthria. Clinical Linguistics and Phonetics, 17 (6), 649-686.

Selfridge, O.G. (1959). Pandemonium: A paradigm for learning. In D. V. Blake and A. M. Uttley, editors, *Proceedings of the Symposium on Mechanisation of Thought Processes*, pages 511-529, London, H. M. Stationary Office.

Sperber, D. (2002). In defense of massive modularity. In Dupoux, E. *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler*. 2002, Cambridge, Mass. MIT Press. 47-57.

Turvey, M. (1991). Coordination. *American Psychologist*, 938-953.

Uttal, W.R. (2001), The New Phrenology, Cambridge, MA: MIT Press.

Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behavioural and Brain Sciences, 21*, 615-665.