2010

Objective measurement of fluency in natural language production: A dynamic systems approach

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

Kim Kirsner

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

Part of the Medicine and Health Sciences Commons

This article was originally published as:
http://doi.org/10.1016/j.jneuroling.2010.03.001

This article is posted on ResearchOnline@ND at
http://researchonline.nd.edu.au/med_article/100. For more information, please contact researchonline@nd.edu.au.
Dear Author,

Any queries or remarks that have arisen during the processing of your manuscript are listed below and highlighted by flags in the proof. Please check your proof carefully and mark all corrections at the appropriate place in the proof (e.g., by using on-screen annotation in the PDF file) or compile them in a separate list.

For correction or revision of any artwork, please consult http://www.elsevier.com/artworkinstructions.

**Articles in Special Issues:** Please ensure that the words ‘this issue’ are added (in the list and text) to any references to other articles in this Special Issue.

<table>
<thead>
<tr>
<th>Location in article</th>
<th>Query / remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2</td>
<td>References McLachlan &amp; Peel, 2000; Crawford, Howell, &amp; Garthwaite, 1998; Ciccone, Hird, and Kirsner (2004) were cited in text but not provided. Please provide these references or delete the citations.</td>
</tr>
<tr>
<td>Q3</td>
<td>Please check the Dochead</td>
</tr>
</tbody>
</table>

**Uncited references:** References that occur in the reference list but not in the text – please position each reference in the text or delete it from the list.

**Missing references:** References listed below were noted in the text but are missing from the reference list – please make the list complete or remove the references from the text.

Electronic file usage

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

- [ ] Scanning (parts of) your article
- [ ] Rekeying (parts of) your article
- [ ] Scanning the artwork

Thank you for your assistance.
Theoretical article

Objective measurement of fluency in natural language production: A dynamic systems approach

Kathryn Hird a,b,*, Kim Kirsner c

a School of Medicine (Fremantle), University of Notre Dame Australia, 35 Henry Street, Fremantle 6959, Western Australia
b Department of Pediatrics, School of Medicine, University of Western Australia, Crawley 6009, Western Australia
c Centre of Experimental and Regenerative Neuroscience, School of Animal Biology, University of Western Australia, Crawley 6009, Western Australia

Article history:
Received 21 December 2009
Received in revised form 1 March 2010
Accepted 1 March 2010

Keywords:
Fluency
Language
Aphasia
Objective Measurement

Abstract

Language research is dominated by the concept of modularity. The basic assumptions involve neural localization of function, and adoption of tasks that tap into specialized functions, involving words or phonemes for example. The tasks that emerge to support this research are generally de-contextualized. Recent work in neuroscience has identified large-scale self organizing neural networks. It is our contention that the advanced neuro-imaging procedures demand an equivalent refinement in the language sampling domain. The collection of natural speaking samples, and an objective approach to fluency, are critical to the understanding of language production. This paper describes a measurement system designed to quantify fluency in natural spoken language. The system classifies environmental and breathing noise, and estimates means and standard deviations for the three lognormal distributions associated with spontaneous speaking: short pauses, long pauses and speech segment duration. The analysis of natural samples produced by three diverse aphasic speakers demonstrates the sensitivity of the fluency measure as well as the profile of independent or correlated changes across the parameters. The system yields objective and sensitive measures of communicative efficiency for individuals across a variety of speaking contexts.

© 2010 Published by Elsevier Ltd.

* Corresponding author. School of Medicine (Fremantle), University of Notre Dame Australia, 35 Henry Street, Fremantle 6959, Western Australia.
E-mail addresses: khird@nd.edu.au (K. Hird), pkirsner@bigpond.net.au (K. Kirsner).

0911-6044/$ – see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.jneuroling.2010.03.001

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001
1. Introduction

Recent advances in neuro-imaging have provided biological confirmation of the presence of large-scale networks underpinning language processing (Thompson & den Ouden, 2008). These dynamic assemblies span both cerebral hemispheres in the normal speaker. Compensatory or alternative patterns of language processing networks have also been observed in speakers with aphasia before and in response to, rehabilitation (Thompson & Small, 2000; Thompson & den Ouden, 2008).

Neuro-imaging evidence is generally derived from studies of normal and brain damaged participants displaying both functional and incidental neuro-plasticity as an adaptive response to novel contexts. These contexts cover opportunities for new learning in healthy brains as well as in response to rehabilitation in damaged brains (Weiller, 2000). Studies targeting motor skills have shown that learning is a significant motivator for actively produced neural reorganisation. This level of neuro-plasticity is dynamically responsive to the continuously variable neurochemical and neurophysiological environment of the brain. Learning is described as a refinement or strengthening of connections between assemblies within a network or the establishment of new connections with other networks (Weiller, 2000).

According to Weiller, efficiencies in neural activity can also be thought of in terms of synergistic changes in the excitability and inhibition of relevant brain regions. The strength of these connections is sensitive to the stage of learning. In the case of brain damage for example, the establishment of new connections is reflected as increased excitability as the remaining connections are sought and maximised. As these connections are rehearsed and become more established, normalisation of brain activity returns as fewer neurons are required.

The relationship between neuronal efficiency and speaking fluency is difficult to specify but it is often assumed. Speaking fluency is used as an index of language processing in a variety of contexts such as; developmental language acquisition, second language learning and recovery from brain damage and carries with it the implication that there is a positive correlation between speaking fluency and the efficiency of neural activity supporting language processing. As networks are re-established, speech becomes more fluent and less effortful. The significance of this proposition concerning diagnosis and rehabilitation of acquired neurogenic communication disorders has received little attention to date despite the surge in studies correlating FMRI images with classification of communication impairment.

The relationship between neuro-imaging data and acquired neurogenic communication disorders is complex, and depends on a number of factors. One question concerns the specificity and sensitivity of neuro-imaging techniques. This domain has changed dramatically in the last 20 years, and it is by no means clear whether or not we have reached the critical level of description. A second question concerns the relationship between specific but rapidly changing patterns of brain activity on the one hand, and the dynamical responsiveness to the continuously variable neurochemical and neurophysiological environment of the brain. Learning is described as a refinement or strengthening of connections between assemblies within a network or the establishment of new connections with other networks (Weiller, 2000).

Is neuro-imaging ready for natural language measures in regard to spatial and temporal resolution? FMRI, for example, provides excellent spatial resolution on Blood-Oxygen-Level-Dependent (BOLD) signal patterns but it cannot detect changes in blood perfusion over time, and it does not therefore provide reliable temporal resolution (Steyn-Ross, Steyn-Ross, Wilson, & Sleigh, 2009). However where spontaneous language is concerned, temporal resolution is critical. MEG and EEG measures detect changes of neural activity with good temporal resolution but have not yet been correlated with hemodynamic changes. A computational solution that links data from neuro-imaging studies with both temporal and spatial resolution remains a task for the future.

Assumptions concerning the dichotomy of resting and active neural states are also under investigation. The subtraction of active states from resting states has been used extensively to localise function in several neuro-imaging studies targeting language processing. For example, Steyn-Ross et al. (2009) suggest that central orchestration of fixed anatomical structures are not responsible for the propagation of information in the brain, rather “changes in effective connectivity arise naturally as a result of dynamic, self organizing context spatial interactions between cortical populations of excitatory and inhibitory neurons” (page, 299). If this explanation is accepted, context and time must be used to define the relationship between neural activity and language processing.
Thompson and den Ouden (2008, p 479) have provided a comprehensive review of treatment induced changes in brain activity captured using fMRI. They report significant heterogeneity in neural activation patterns associated with the presence, type and treatment dose for acquired communication impairments. Of particular interest is that the language tasks employed across all of the studies described by them depend on modular assumptions. The tasks were designed on the assumption that language can be comprehensively characterized by paradigms that pre-suppose specialization of function, neural localization of those functions, and the further assumption that context is not critical. They typically measure for example sentence, word and phoneme performance as if language comprehension and production are the sum of these and possibly other component processes.

According to Coltheart (1999, p 115) “a cognitive system is modular when and only when it is domain-specific”. That is, a module only responds to stimuli of a particular class. In addition, processing efficiency of a module is not influenced by a person’s beliefs, presumptions or desires. Jackson and Coltheart (2002) describe modules as pertaining to proximal functions that are not influenced by distal factors such as age, education, or other environmental characteristics. Whether or not a cognitive module has neurological representation is, according to Coltheart (1999), relatively unimportant for the definition of modularity.

There are two issues that arise in the application of the theory of modularity to the assessment of aphasia. The first concerns the validity of the experimental methodology typically used to produce evidence which supports the identification of domain-specific function. Dunn and Kirsner (1988) discuss limitations in the use of double dissociation to isolate domain-specific functions. Issues raised concern task purity, magnitude of detectable differences in function, the definition of functional differences as opposed to tasks performance.

There is therefore a mismatch between the underlying assumptions associated with each measure: dynamic neural networks and modularized linguistic processing respectively. FMRI studies reflect the dynamic and distributed function of cell assemblies across the brain whereas the language tasks target specific linguistic functions that, presumably, reflect the function of isolable cognitive modules. Assessment of specific linguistic categories prohibit the inspection of the mechanisms supporting natural language processing and, as Hird and Kirsner (2004) assert, there is no real justification to generalise the results of studies employing these tasks to an understanding of language processing or to the characteristics of aphasia.

It is our contention that neuro-imaging procedures demand an equivalent refinement in the sampling domain, and that the collection and analysis of natural language samples is essential. The aim of this paper is to introduce a language paradigm that provides a better fit between the methods and models of language production research on the one hand, and the complex systems assumed by FMRI research on the other. We will use results from a clinical trial involving a small sample of aphasics to assess the viability of the new paradigm.

1.1. Fluency

Perhaps the most important contrast in the language sciences is that between competence and performance (Chomsky, 1965). According to Chomsky, competence involves our internalised and tacit knowledge of language, whereas performance, the external evidence of competence, is sensitive to memory and other limitations that compromise the extent to which it reflects intrinsic competence. Chomsky argued for example that corpus evidence could never be a useful tool for linguists, and that linguists should therefore model competence rather than performance.

Port and Leary (2005, p. 956) provided a compelling analysis of the limitations of the competence-driven approach, however. Thus,

“In a linguistics committed to the physical world (……………), language needs to be naturalized so as to fit it into a human body. That implies, first of all, casting it into the realm of space and time [italics supplied]. It requires changing our focus of attention from our preconceived views of the form of linguistic knowledge toward the study of linguistic behavior and performance.”

With consideration restricted to performance, and the focus on fluency, one further contrast merits consideration. Spontaneous speech and therefore language disorders involve two potential data streams. These streams involve information and the temporal distribution of acoustic energy respectively.
1.1. The informational stream

The first or informational stream involves the amount and the quality of information conveyed by a speaker per unit of time. This type of knowledge is usually subject to psycho-linguistic assay, and indexed by specialized lexical, syntactic and semantic tests. In our approach however, with the emphasis on spontaneous speech, we have adopted an established procedure to measure the rate at which information is being transmitted. Following Ciccone, Hird, and Kirsner (2004), we have adopted seconds per Correct Information Unit (s/CIU) to meet this requirement. Our analysis of CIU follows that of Nicholas and Brookshire (1993), and the temporal designation (s/CIU) was introduced to make it comparable to lexical decision. It is of course inter-changeable with minutes per CIU as used by Hula, McNeil, Doyle, Rubinsky, and Fossett (2003).

1.1.2. The temporal distribution of acoustic energy

The second stream concerns the way in which acoustic energy is distributed over time. In broad terms this stream is concerned with the frequency and duration of pauses and speech segments. The Boston Diagnostic Aphasia Examination (BDAE) includes provision for analysis of this type of data in speakers with acquired aphasia. It includes provision for an objective count of phrase length, and subjective rating scales for melodic line, articulatory agility and paraphasia in running speech. This level of description, the poor inter-judge reliability employing subjective criteria has been described by Gordon (1998), Hird, Silvestri, Dunn, and Kirsner (2005).

The proposition that the temporal distribution of acoustic energy, or, more specifically, the pause and speech duration distributions, offer an important source of information for the science of communication disorders is not new. It is inherent in regular use descriptions of language and explicit in reports by Butterworth (1979) and Niemi (1988).

The purpose of this paper is to critically review the assumptions underlying work on pause and speech duration, and to describe the utility of an objective approach to analysis as a better fit to the assumptions underlying neural processing.


The procedure described below involves an objective approach to segmentation and pause analysis. Three problems merit consideration; first, the extent of intra- and inter-speaker variability; second, the presence of lognormal distributions in both pause and speech segment duration samples; and, third, the possible presence of two lognormal distributions in natural language.

Fig. 1 depicts a pause duration distribution for a single participant from the study reported in this article. The speech samples were obtained from picture descriptions and procedural narratives for a pilot study. The data was accumulated over eight speech samples of approximately 2 min each. The test sessions were implemented during a single two week period. The speech samples were segmented by hand, using P The Detection Threshold – the lowest admissible pause duration value – was 20 ms. The figure shows the number of pauses at each of 100 consecutive 25 ms bands. The distribution is based on more than 700 pause duration observations. Fig. 1a and b depicts the distribution in milliseconds and log units respectively.

The dominant feature is that the distribution is massively skewed. The vast majority of the pauses fall below 100 ms. One implication of this figure is that the arithmetic mean does not provide a defensible measure of the central tendency of the distribution. Fig. 1b involves the same set of observations depicted in Fig. 1a. The difference is that each and every observation has been subject to a natural log transformation prior to creation of the figure. The bin size is 0.1 log units. The real character of the pause duration distribution is now revealed (see also Kirsner, Dunn, Hird, Parkin, & Clark, 2002). It is now apparent that there are two pause duration distributions rather than one.

1.2.1. The classification question

Calculation of the natural log means and standard deviations for these distributions depends on a classification decision about the boundary between the two distributions. As depicted in Fig. 1b, the
distributions overlap, and it is therefore necessary to define a Classification Threshold – to classify pauses into short and long bins – before estimating the statistical characteristics of each bin or pause duration distribution.

Previous studies had implicitly classified different pause types by adopting an arbitrary Detection Threshold, and rejecting pauses with durations less than this value. One solution therefore is not to discard such short duration pauses, but to classify them as the short type, with all pauses with durations greater than the threshold being classified as the long type. Because of individual differences however, and the impact of brain damage on such differences, there is no guarantee that this strategy would be optimal for any speaker or for all speakers. A more sensitive approach would be to determine an optimal classification threshold for each speech sample based upon the estimated statistical characteristics of the short and long pauses. This approach is outlined below.

1.2.2. The fluency profiling system

Our solution to the taxonomy, description, and classification questions involves three main steps. First, pauses with durations greater than a detection threshold of 20 ms are identified in a given speech sample by examining the amplitude contour using Praat (Boersma, 2001). Second, the set of pause durations is fit by a single lognormal distribution and by a mixture of two (or potentially more) lognormal distributions. Third, an optimal classification threshold is calculated based on the means and variances of the two fitted distributions and the expected misclassification rate determined. The optimal classification threshold is defined as the value that minimizes the expected misclassification rate.

This approach is based on two main assumptions. First that there are two distributions of pause duration corresponding to short and long pause types, respectively. Second, that each of these distributions is at least approximately lognormal. Given the validity of the assumptions, it is possible to estimate the mean, standard deviation, and relative proportion of each component distribution from the data. This is achieved by fitting a mixture of two normal distributions to log pause duration using the Expectation Maximization (EM) Algorithm (McLachlan & Peel, 2000). Once the component distributions have been fit to a sample of pause durations, it is possible to estimate an optimal classification threshold (Kirsner, Hird, & Dunn, 2005).

The Classification Threshold can be used to optimally classify individual pauses as being of either the short or long type. Following the work of Goldman-Eisler (1968), the set of long pauses was used to mark segments of continuous speech that may contain one or more short pauses.

The patterns of distribution have been shown to be stable for normal speakers. Control data derived from 13 non-brain damaged speakers revealed a mean short pause duration of $\log_{10} \text{ms} \ 4.21$ (SD $\log_{10} \text{ms} \ 0.59$), a mean long pause duration of $\log_{10} \text{ms} \ 6.61$ (SD $\log_{10} \text{ms} \ 0.68$) and a speech segment duration well-described by a single lognormal distribution with a mean of $\log_{10} \text{ms} \ 7.22$ (SD 0.77).

It is further assumed that the long pauses reflect cognitive processes, for conceptualization, formulation and lexical selection. Short pause is integral to speech production, and it is proposed...
that they differentiate the articulatory gestures into recognizable combinatorial units. But how does brain damage impact on these distributions and the relationships between the underlying control mechanisms?

If the assumption that language processing and production is supported by distributed dynamic networks that have been linked via their duality of patterning in response to context requirements is given, then damage impacting the control mechanisms required for speech production will also impact on those required for language production and vice versa. Hird and Kirsner (2008) reported preliminary fluency analysis data involving a small number of participants with aphasia that supports this proposition. In this study two participants experienced selective impairment to speech production and not language production; three participants showed impairment to both speech production and language production, and two participants without impairment to either system. Hird and Kirsner suggested that these results could be consistent with either of two hypotheses. Firstly that the control processes for language production and speech production are independent, and that double or reversed associations are therefore possible (Dunn & Kirsner, 1988). Secondly that the underlying distribution reflects a positive association where impairment of speech production is invariably accompanied by an impairment involving language production, but that the data analysis is not sufficiently sensitive to detect the relevant change. Scrutiny of the individual pause duration distributions suggested that this might be the case. The results could be explained by disturbance of diffuse distributed cortical representation. An alternative explanation however involves the disruption of a symbiotic link between the processes required for speech and language production. The notion is derived from the concept of duality of patterning because both speech and language are critical for communication and it is possible that the timing mechanisms for these functions cannot operate independently.

If language processing is considered in these terms then significant changes to models of both normal and disordered speech and language processing are required.

2. Method

2.1. Participants

To illustrate the utility of the chronometric analysis technique and as objective measurement of fluency, three speakers with acquired aphasia following stroke were selected from participants involved in a larger study (Ciccone, Hird, & Kirsner, 2000) as they exhibited diverse fluency patterns as classified by the Boston Diagnostic Aphasia examination.

Participant GJ was judged as non-fluent and classified with Broca’s aphasia with a severity of 5 (mild). GJ was a 74 year old female with a Hemorrhagic Left Middle Cerebral Artery. The test session was conducted 11 months post-onset and she had had eight years of education. GJ had had eight years of education, and the test session was conducted 11 months post-onset.

Participant GN was judged as fluent and classified with Wernicke’s Aphasia with a severity of 1 (severe). GN was a 75-year old female with Ischaemic stroke to the Left Middle Cerebral Artery. GN had had nine years of education, and the test session was conducted 24 months post-onset.

Participant JB was judged as non-fluent and classified with Trans0cortical Motor Aphasia with severity of 3. JB was a 72-year old male with a fronto-parietal Haemorrhagic stroke. JB had had seven years of education, and the test session was conducted 44 months post-onset.

An experienced speech pathologist judged each participant as highly intelligible speech without evidence of motor speech disorder.

2.2. Speech sampling

In the larger study natural discourse was sampled across three speaking contexts from 8 participants. Speech samples were collected as described in Ciccone et al. (2000) and involved procedural narratives, picture descriptions and natural responses to questions such as “what did you do yesterday?” Each participant produced eight samples across the four speaking conditions within a two week period. Each participant was required to respond to the same elicitation stimuli and the time...
taken to complete these tasks was also recorded. These samples were then collated and pauses were identified according to the procedure described in Kirsner et al. (2002). The EM algorithm described previously was implemented and short and long pauses distributions for each speaker were defined. Long pauses were then used to identify the boundaries for resulting speech segments. Correct Information Units (Nicholas & Brookshire, 1993) were calculated on the language segments and the scores were converted to measures of second per correct information unit to obtain a measure of communicative efficiency. All parameter scores were converted to Crawford's $t$ (for small samples) (Crawford, Howell, & Garthwaite, 1998) and were individually compared to the summary reference distribution derived from speaking samples produced by a group of 13 non-brain damaged normal controls participants between 40 and 70 years of age.

Speaking samples for control participants were collected and analysed using the same procedures described for the participants with aphasia. Each parameter was described statistically and the mean and 95% confidence intervals were calculated for comparative purposes. Both the long and short pauses were described in terms of proportion, rate (per minute), mean (log ms) and standard deviation. Other parameters include duration of the speech sample (minutes), pause classification threshold (log ms), proportion of pause misclassification and seconds per CIU. Speech segment duration is described in terms of mean (log ms) and standard deviation.

3. Results

The results for the non-Brain Damaged control group are depicted in Fig. 2.

The short pause, long pause and speech segment duration distributions have mean durations of $4.21 \pm 0.59 \text{ log units (67 ms)}$, $6.61 \pm 0.68 \text{ log units (742 ms)}$, and $7.22 \pm 0.77 \text{ log units (1356 ms)}$ respectively. The figure also shows the relative proportions of the short and long pause distributions ($p = 0.77$ and $p = 0.23$), mean sample duration (1.83 s), mean EM-Threshold (5.59 log units or 268 ms), the mean proportion of misclassifications between the pause distributions (0.024), and the mean rate of transmission of Correct Information Units (0.5 s/unit).

**Control Group - non-brain damaged (N=13)**

![Graph showing pause and speech segment duration distributions for 13 non-brain damaged individuals.](image)

**Fig. 2.** Summary depiction of the pause and speech segment duration distributions for 13 non-brain damaged individuals. The pause and speech duration distributions are shown above and below the abscissa respectively.

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001
3.1. Case GJ – Broca’s aphasic

Fig. 3 depicts the results for Case GJ, a mild Broca’s aphasic. The figure and the associated statistics show that mean Time per CIU increased relative to the Non-Brain Damaged control, at 1.4 s per CIU ($t = 8.6$), and that mean Long Pause Duration also increased relative to the mean for the control group, to 7.51 log units ($t > 5.61$).

3.2. Case GN – fluent Wernicke’s aphasic

Fig. 4 depicts the results for Case GN, a fluent Wernicke’s aphasic. Case GN had an exceptionally long speech sample, at 3.49 min ($t = 2.39$). The Short Pause, Long Pause and Speech Segment Duration statistics are in the normal range for GN but Time per CIU increased significantly, at 2.78 s per CIU compared with the control mean of 0.50 ($t = 21.7$).

3.3. Case JB – non-fluent trans-cortical motor aphasic

Fig. 5 depicts the results for Case JB, Non-Fluent Trans-Cortical Motor Aphasic. Case JB differed from the control group on six out of 14 parameters. The major changes involved the proportions of each pause type, 0.44 Short Pause compared with 0.22 for the control group, and 0.56 Long Pause instead of 0.78 for the control group; mean short pause duration, at 4.59 log units compared with the mean for the control group of 4.21 log units ($t = -5.27$); and mean Speech Segment Duration, with 6.41 compared with the mean for the control group of 7.22 ($t = -3.0$). Time per CIU also increased, at 2.08 s compared with 0.50 for the control group ($t = 15.08$).

4. Discussion

The results show that natural speaking can be characterized in terms of a series of parameters that quantify fluency, that are stable in normal speakers, and can be measured objectively. The fluency

**Case: 18 GJ: Non-Fluent - Broca (LHD) (Severity = 5)**

![Graph](image)

Fig. 3. Includes the data from one participant diagnosed as a mild Broca’s aphasic. The patient’s distributions are shown as a thin black line against the thick grey functions for the non-brain damaged control group. The pause and speech duration distributions are shown above and below the abscissa respectively.

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001
Case: 17 GN: Fluent-Wernicke (LHD) (Severity = 3)

Fig. 4. Includes data from one participant diagnosed as a Fluent Wernicke’s aphasic. The patient’s distributions are shown as a thin black line against the thick grey functions for the non-brain damaged control group. The pause and speech duration distributions are shown above and below the abscissa respectively.

Case: 20JB: Non-Fluent-Trans-Cortical Motor (LHD) (Sev = 3)

Fig. 5. Includes data from one participant diagnosed as a Non-Fluent Trans-cortical Motor aphasic. The patient’s distributions are shown as a thin black line against the thick grey functions for the non-brain damaged control group. The pause and speech duration distributions are shown above and below the abscissa respectively.

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001
patterns for the three aphasic speakers are in keeping with the Boston Diagnostic Aphasia classifications but in addition provide a quantitative description of the factors that contribute to fluency disruption. In this sense the Fluency Profiling System provides data that reflect language processing efficiency, and can be used to test hypotheses about dynamic cognitive and language processes. In addition, the level of detail provided by the Fluency Profiling System is sensitive, and it is therefore capable of detecting change over time during rehabilitation, or between contexts.

4.1. Segmentation

There are several additional benefits of the Fluency Profiling System. The pause analysis provides an objective method of segmenting continuous speech samples without reference to linguistic rules. This is important when speech has been disrupted and the use of linguistic rules is not conventional or, in the case of second language users, when their second language grammar is not as proficient as their native grammar. An additional application involves developmental language, where communicative intention is emphatic but language acquisition is incomplete.

It is probable that the speech segments defined by the long pause boundaries reflect cognitive processing chunks that emerge in a step-wise linear sequence. Questions about segmentation pose a challenge to psycho-linguistics. Which source of segmentation information is used during language comprehension? Is it situation-or individual-specific for example, or does it involve a neural race among alternative sources of information? With few exceptions, the established segmentation models depend on competence rather than performance, even to the extent that training can be used to facilitate segmentation by professionals. But what if the entire process in natural language production and reception is based on a performance model?

One obvious alternative involves pause duration. In this and other studies we have demonstrated that pause duration involves two lognormal distributions, distributions that vary substantially from individual to individual and context to context. For example, in brain damaged people the range of observed thresholds – the value that provides the optimal separation between short pause and long pause distributions – can vary from approximately 100 ms to 450 ms. But even in the non-brain damaged population, the observed thresholds vary from 150 ms to 400 ms, with a mean across individuals of approximately 250 ms, the value nominated by Goldman-Eisler (1968). For group purposes her value is valid but when the same value is used for all of the participants in a study it will yield substantialpause misclassification, of short as long, and vice versa. However, with the procedure developed by us, the EM algorithm selects a value for each individual that minimizes misclassification.

We propose therefore that the performance approach based on the mathematical separation of speech into segments or chunks may be particularly appropriate. Furthermore, given automatic classification of the speech stream into ‘silence’ and ‘speech’ (in preparation), the system can be refined to offer a solution that is both objective and automatic, and independent of human classification based on competence models. The System can be refined to offer solutions that are both objective and automatic, and independent of human classification based on competence models.

The Fluency Profiling system offers other advantages. Using the technique described here it is possible to determine how speech errors relate to speech segments and hence the speech analyst has a better notion of speaker intention as the sample emerges. Alternative systems rely on assumptions associated with sophisticated computational linguistic rules involving both syntax and semantics. In addition, once the speech segments have been defined, by further analysis involving lexical, syntactic or phonological information can be implemented.

4.2. Objective measurement of complex dynamic systems

A long-term goal for research into the neurobiology of language should be to develop parallel instruments for the measurement of neural activity and language. The Fluency Profiling System described in this article provides statistics that characterize the overall performance of the language production system for sample periods of two or more minutes.

The Fluency Profiling System is in some respect comparable to the Electrocardiogram. The Electrocardiogram or ECG involves a single or typical cycle, and it is possible to depict a sample cycle for

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001
language production as well. The Fluency Profiling System as described here involves average values for dozens or even hundreds of cycles, and provides estimates of cycle components that correspond loosely to the ECG components. The point of this comparison is to underline the similarities between the steps on the road to quantification in biological and medical science on the one hand, and language on the other.

The next step in this comparison is to provide a provisional estimate of the clinical value of component measures in the Fluency Profiling System. For the ECG, shortening or prolongation of the QT interval implicates Hypercalcaemia and Hypocalcaemia respectively while inverted T waves generally indicate Coronary Ischemia.

Fig. 3 depicts a patient classified with Broca’s Aphasia, and the anticipated changes in the Long Pause Distribution are evident; the mean and standard deviation are increased relative to the non-Brain Damaged control group. The information transmission measure shows too, that less information is being transmitted per unit of time.

Fig. 4 depicts a patient classified with Wernicke’s aphasia. The patient’s mean Long Pause and Speech Segment Durations are actually shorter than the equivalent means for the non-Brain Damaged control group, although not significantly so. However, despite the fact that the patient’s language is clearly ‘fluent’ in acoustic terms, the extent of the decline in the rate of information transmission is in line with traditional characterizations of Wernicke’s aphasia.

Fig. 5 depicts a patient classified with non-Fluent trans-cortical motor aphasia. In such a patient we would usually expect short, halting and effortful utterances. The most striking change in the Fluency Profiling parameters involved Speech Segment duration, reduced significantly compared to the equivalent mean for the non-brain damaged control. It should also be noted that although both the Broca and Trans-cortical motor aphasics show increases in their Pause-to-Speech Ratios, this is achieved by increased mean Long Pause Duration in the patient classified with Broca’s aphasia while it reflects a reduction in mean Speech Segment Duration in the patient classified with Trans-cortical motor aphasia.

Fig. 6. Sketch depicting potential mapping between the spatial and temporal parameters associated with a sample of the tools for mapping brain areas (see Van Horn, 2003) on the one hand, and the measurement of speech segment duration in the fluency domain, on the other.

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001
4.3. Mapping between neuroscience and behavioural science parameters

To identify the potential relationship between the neuroscience and behavioural domains, it is necessary to return to the quotation from Port and Leary (2005). Space and Time are critical. Consider Fig. 6. The figure is a sketch based on Van Horn (2003). The original figure depicted the spatio-temporal regions associated with 11 tools for mapping brain areas. The axes involve log base 10. The fluency parameters originally calculated for Natural Log have been modified to fit base 10. The spatio-temporal areas depicted for fMRI and MEG are from Van Horn. A provisional area is also shown for speech segment duration based on our own work. The area shows only the range; the values for specific segments fall between these extremes. The spatial limits are of course unknown.

Speech segment duration has been used as an example in this figure. It is also possible to map the short and long pause durations into the spatio-temporal region defined by Van Horn (2003). The figure has been designed to illustrate the potential mappings that could be achieved for objective and sensitive measures of continuous and contextualized language tasks on the one hand, and brain mapping on the other.

5. Conclusion

The Fluency Profiling System supports objective and sensitive measurement of fluency for natural speech samples. Fluency in this sense can be thought of as an index of integrated neural function for speech production. In addition, it provides a clear algorithm for segmentation of continuous speech based on the acoustic signal. This is important as speaking fluency requires integrated brain function and could be an important reflection of neural processing efficiency. The system permits the analysis of efficiency, and the identification of factors that could interfere with efficiency. Future studies could focus on the correlation between the measures provided by the Fluency Profiling System and dynamic measures of brain function. Currently the Fluency Profiling System provides a useful baseline measure to capture the efficiency of speaking performance at stages of development, second language acquisition, and recovery from brain damage.

References


Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, Journal of Neurolinguistics (2010), doi:10.1016/j.jneuroling.2010.03.001

Please cite this article in press as: Hird, K., Kirsner, K., Objective measurement of fluency in natural language production: A dynamic systems approach, *Journal of Neurolinguistics* (2010), doi:10.1016/j.jneuroling.2010.03.001