Title

Aphasia classification: the relationship between objective acoustic measurement of spontaneous speaking samples and naïve listener judgments of similarity and fluency.

For an objective measure to be useful in a clinical setting it must correlate with perceptions of the speaker and what is considered “normal” (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993). Perceptions of the functional communicative abilities of people with aphasia do not correlate well with changes on decontextualised linguistic measures (Wertz, 1999). In addition, the current practice of reliance on specific perceptual ratings of fluency of speech has been shown to be unreliable (Gordon, 1998; Kent, 1996,). Using a multidimensional scaling technique Kreiman, Gerratt & Precoda (1990) found that naïve and expert listeners attended to different aspects of voice when making similarity judgments. They recommended that naïve listeners become the “gold standard” for perceptual judgment tasks as their listening experiences were more homogenous.

Listeners learn, through the computational analysis of connected speech, that pause duration information is an important factor in determining how spoken discourse is segmented, analyzed and interpreted, as pauses allow the listener to identify discourse structures and links between related materials (Fox Tree & Schrock, 1999). Perception of speech is achieved holistically, and values obtained for any one dimension may be highly influenced by co-occurring dimensions and the range of individual past listening experiences (Kent, 1996, 1997; Lehar, 2003). The development of a classification method derived from acoustic measures offers an objective approach to the correlates of speaking fluency. Fluency in this instance is considered a system measure rather than an isolable characteristic of spontaneous speech.

The aim of this study was to explore (a) the relationship between objective pause data and naïve listeners’ similarity judgments of aphasic connected speech samples using MDS (b) the relationship between objective pause data and naïve listeners’ direct magnitude fluency estimates of aphasic connected speech samples and (c) the way listeners employ descriptors and concepts when describing aphasic speech samples.

Participants

Six male and 16 female adults participated as listeners. They had no known history of neurological, cognitive, hearing or communication impairment. Four participants were fluent in Italian as well as English. They reported little experience with communicatively disordered speakers.

Materials*:*

Aphasic speech sample stimuli. The 8 aphasic speech samples used in this study were collected by Ciccone (2003). Samples consisted of a verbal response to the question “What did you do yesterday”? They were 60 seconds or less in length and finished at the nearest syntactic boundary (Rose & Duncan, 1995).

The objective pause data, included mean short pause duration, mean long pause duration, mean speech segment duration (time period between long pauses) and d’ (measure of discriminability between the short pause duration and long pause duration log-normal distributions). Data from each aphasic sample was described in terms of critical z values of greater than 2 or less than -2 in relation to the normal distribution for each parameter.

Speech Sample Interface Tool (SSIT; Woods, 2004) The software automatically presented the 8 aphasic speech samples in all possible triad combinations, in a different random order for each participant.

Procedure

Each participant completed three tasks in the same order. The tasks included: (1) similarity judgments of speech sample triads; (2) written description of the eight individual speech samples; (3) direct estimation of fluency without modulus.

Task 1 required that participants make similarity judgments from a selection of three speech samples until they had listened to all possible triads. They were instructed to consider the samples as a "whole". In the second task each participant was required to "write a short descriptive paragraph" for each of the eight samples. Participants received verbal instructions for the direct magnitude estimation task once they had completed the first two tasks to avoid confounding Tasks 1 and 2 through the focus on fluency. Participants were instructed to assign any positive number to sample A, and rate the following samples proportionately.

Results

1. MDS Analysis: The results of the MDS analysis can be seen in figure 1. Kruskal's stress value for the 1-dimensional solution is 0.11, with the squared correlation of fit equal to 0.95. There was no significant relationship between the one-dimensional MDS solution coordinates and short pause duration data ($r^2 = 0.40$, $p > 0.05$), long pause duration data ($r^2 = 0.06$, $p > 0.05$), d'data ($r^2 = 0.33$, $p > 0.05$) and speech segment duration data ($r^2 = 0.07$, $p > 0.05$).

2. Direct magnitude Estimation: Initially direct magnitude estimates were modulus-normalised according to procedures described by Engen (1971). The average DME value for each speech sample was found by calculating the mean of participant's responses for each sample. (See figure 2).

3. Sample descriptions: All participants' descriptions of samples A to H were entered into the Semantic Analysis (LSA) from the LSA @ CU Boulder website (http://lsa.colorado.edu.au) which provided a similarity matrix based on the semantic relationship between the words used to describe each sample. The results for this analysis revealed that similarity scores (cosines) for sample descriptions ranged from 0.72 to 0.89. The average similarity score was 0.83. These results indicate that the descriptions were highly semantically related and therefore similar concepts were used by participants to describe the 8 speech samples.

The results of the Leximancer (http://www.leximancer.com.au) analysis provided the relative count in percentages for use of specific concepts in participant's descriptions. They were: speech (100%), words 96.9), slow (50%), and sentences (40.9), voice (40.9) and sounds (40.9).

Summary of Initial Results

Short pause data and d' data respectively explain 40% and 33% of the variance in the 1-dimensional similarity judgments solution. These variables also respectively explain 26% and 16% of the variance in participants' mean direct magnitude estimates of fluency. Participants as a group employed similar concepts to describe the speech samples. However when describing individual speech samples, participants varied in the concepts they employed. When describing samples, participants' focused on speech itself, not on the content. "Pauses" were among the main concepts that were salient to participants.

Summary of Post Hoc Results

Syllables per minute data explain 63% of the variance in the 1-dimensional similarity judgments solution, and 66% of the variance in participants' mean direct magnitude estimates of fluency. Participants' mean direct magnitude estimates of fluency explain 89% of the variance in the 1-dimensional similarity judgments solution.

Conclusions

The results suggest that listeners are perceptually attuned to timing and fluency aspects of spontaneous speech. The importance of these temporal speech characteristics support further investigation language processing models that places emphasis on the dynamic interactions between variables. Further investigation concerning the multidimensional acoustic characteristics of spontaneous speech and the way in which listeners respond is warranted. The characteristics of short pauses and their relationship with speech segment duration are of particular interest.

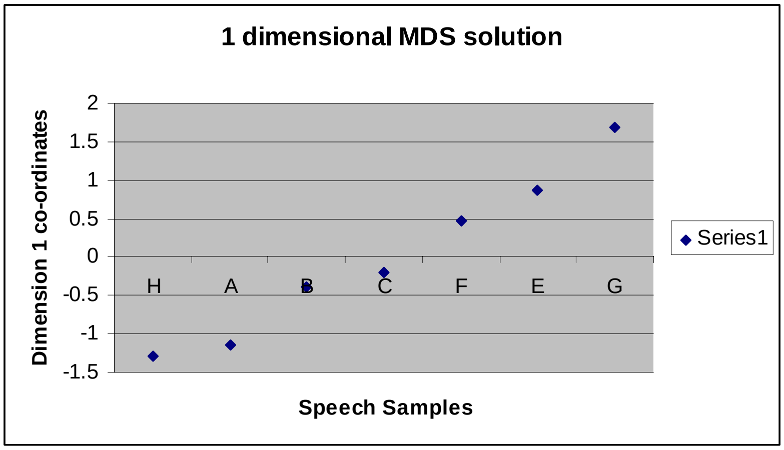

Figure 1. One-dimensional Multi-dimensional Scaling solution.

Least fluent									Most Fluent	
1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2
G			E		F D C		A	BH		

Figure *2*. Relative fluency of the speech samples based on participants' direct magnitude estimates (after data transformation).

References

Ciccone, N., (2003). Stability of acquired neurogenic communication disorders. *Unpublished Doctoral Thesis, Curtin University of Technology*. Perth. Western Australia.

Fox Tree, J., Schrock, J. (1999). Discourse markers in spontaneous speech: Oh what a difference an Oh makes. *Journal of Memory and Language, 40*, 280-295.

Gorden, J. (1998). The fluency dimension in aphasia. Aphasiology, 12 (7/8), 673-688.

Kent, R. (1996). Hearing and believing: Some limits to the auditory perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5(3*), 7-23.

Kent, R. (1997). The speech sciences. San Diego: *Singular Publishing Group, Inc*.

Kirsner, K., Dunn, J., Hird, K., Parking, T., & Clark, C. (2002). Time for a pause….*Proceedings of the 2002 Speech Science Technology Conference.* Sydney. Australia.

Kreiman, J., Gerrat, B., Kempster, G., Erman, A. & Berke, G. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *Journal of Speech and Hearing Research, 36*, 21-40.

Kreiman, J., Gerrat, B.,& Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research, 33*, 103-115.

Lehar, S. (2003). The world in your head: A gestalt view of the mechanism of conscious experience. *New Jersey: Lawrence Erlbaum Associates, Publishers*.

Leximancer manual (Version 2.0). (2004). Retrieved September 3, 2004, from http://www.leximancer.com.au

Rose, P., & Duncan, S. (1995). Naïve Auditory identification and discrimination of similar voices by familiar listeners. *Forensic Linguistics, 2(*1), 1-17.

Woods, A.(2004). Speech Sample Interface Tool (Version 1.7). Perth Western Australia.