Prosody of speech production in latent aphasia

Cong Zhang[1*], Tong Li[2], Gayle DeDe[3], and Christos Salis[1]

[1] Speech & Language Sciences, Newcastle University, Newcastle upon Tyne, UK
[2] School of Foreign Languages, Tianjin University, Tianjin, China
[3] Department of Communication Sciences & Disorders, Temple University, Philadelphia (PA), USA

Introduction

Latent aphasia is a mild aphasia type in which affected individuals perform within normal limits on clinical aphasia batteries or other language tests. Identification of latent aphasia is challenging and little is known about its speech output. Previous work examined language production [1, 2]. To our knowledge, this is the first study to date on prosody of connected speech production in latent aphasia.

Aims

1) To examine how latent aphasia affects features of expressive prosody according to different word positions in utterances (utterance-initial, utterance-final) [3].
2) To investigate whether prosody markers can identify people with latent aphasia versus neurotypical controls.
3) To explore the potential of expressive prosody and demographic information in correctly classifying people with latent aphasia versus neurotypical controls.

Methods

Participants: Ten speakers with latent aphasia, 10 neurotypical controls, statistically similar in age, education, and sex.

Data: Audio-recordings of Cinderella story narrations from AphasiaBank [4] in which utterance-initial words and utterance-final words were manually annotated using Praat.

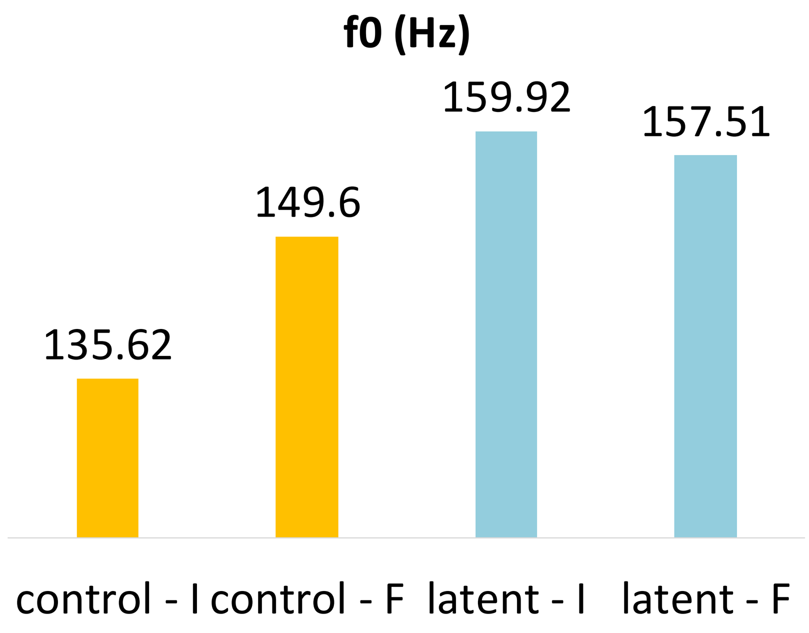
Acoustic feature extraction & Analyses:

- Mean f0 (fundamental frequency), mean intensity, mean duration of words in utterance-initial, utterance-final positions. Linear mixed-effect models (LLMs): DVs were mean f0 (Hz), mean intensity (dB), mean duration (ms). Fixed effects were groups (latent aphasia, control) and word position (utterance-initial, utterance-final). Random effects were speaker and item.

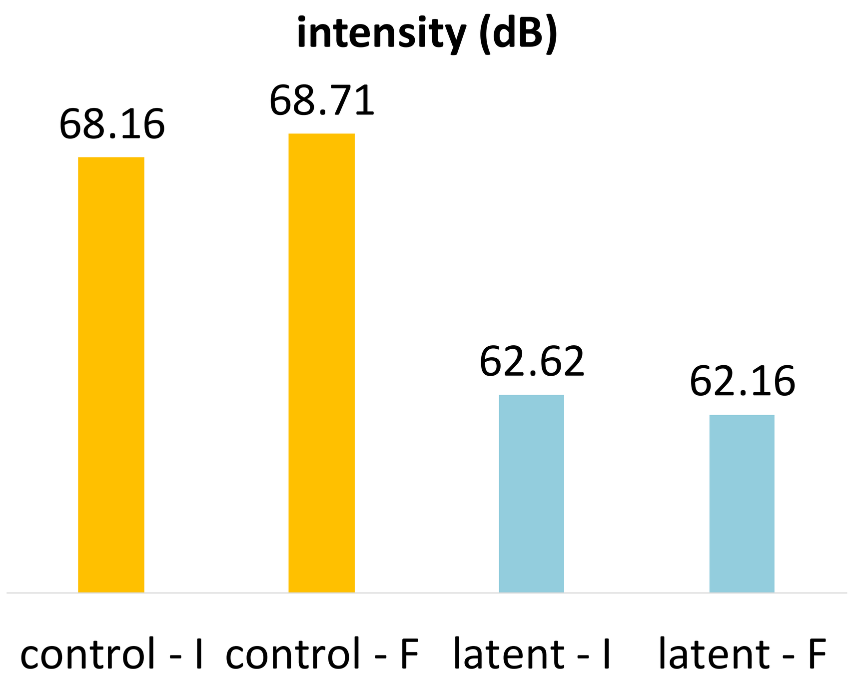
- Auto-classification was based on: **a)** 23 common acoustic correlates of prosody and temporal measurements, e.g., minimum, maximum f0 (and their time points), jitter, shimmer. **b)** 988 features from the *emobase* feature set in *openSMILE* [5]. We used random forest analyses with acoustic correlates, temporal measurements, *emobase* features, and demographic data. Collinearity was not present in the random forest models.

Results

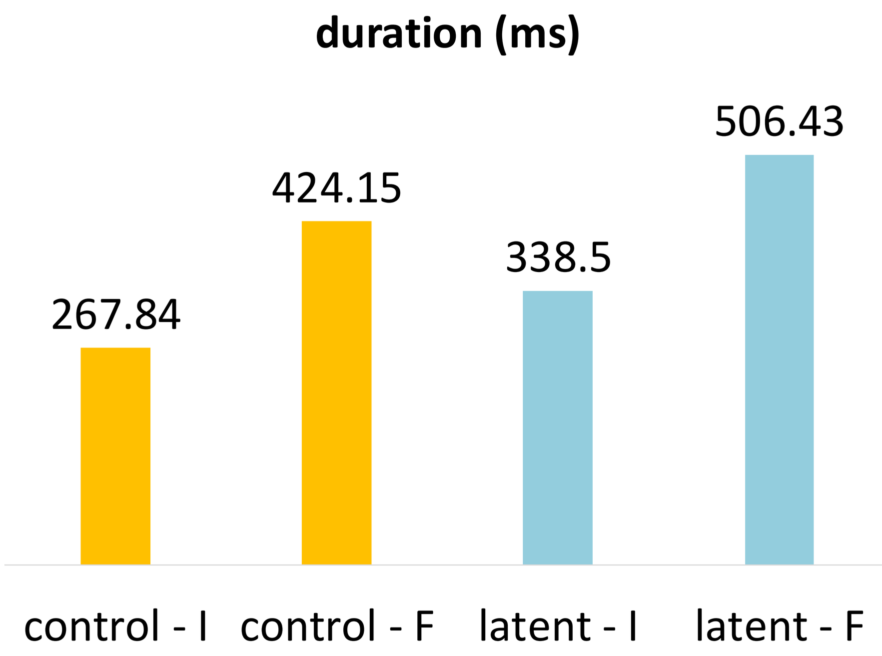
The bar charts below show the means by group and word position within utterances (I = initial, F = final). The textboxes show only statistically significant results (*p < .05) of the LMMs.



LMM: *position, *group x position



LMM: *group x position



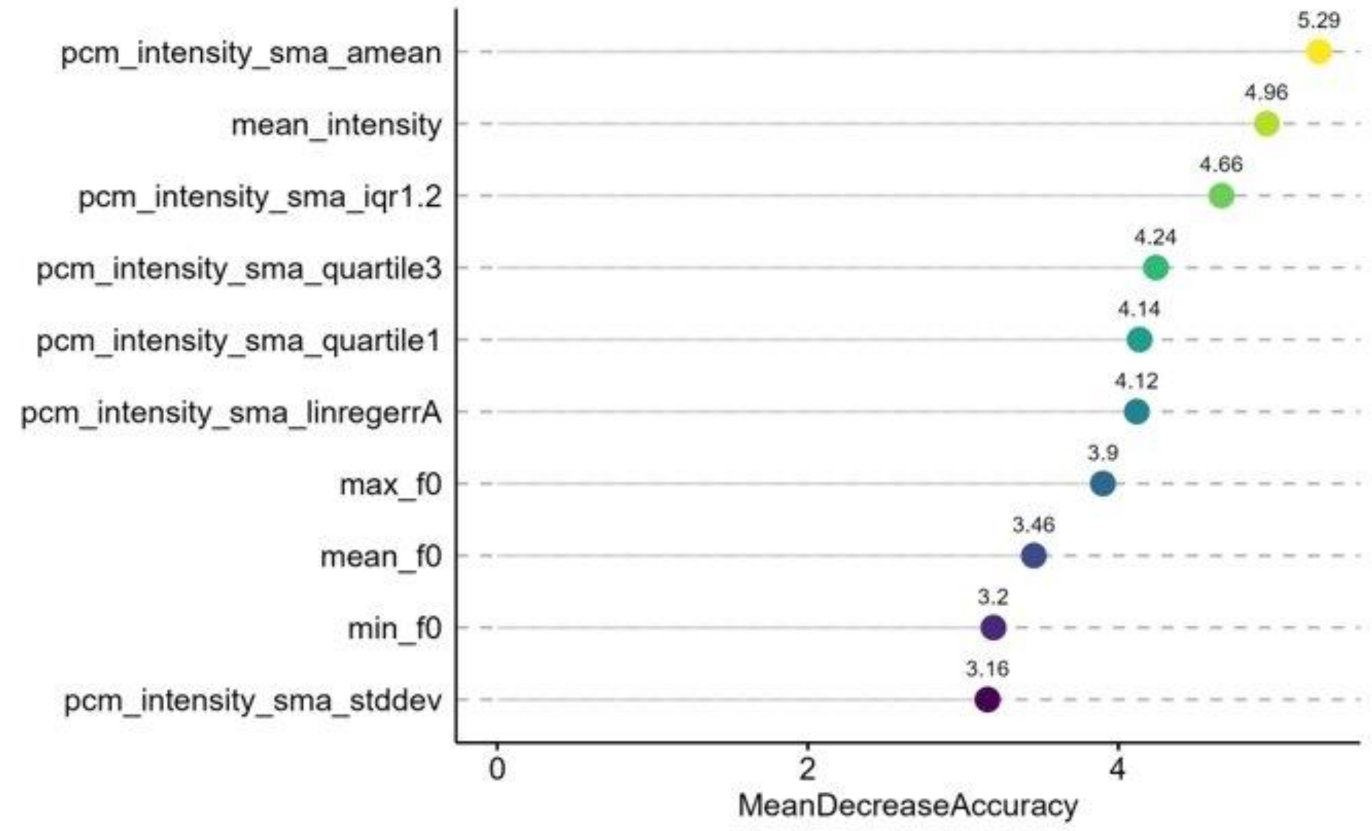
LMM: *group, * position

The table below shows the three auto-classification analyses (%).

		Accuracy	Sensitivity	Specificity
Models				
1	Random split (70% train; 30% test)	97.8	98.6	97.1
	10-fold cross-validation	99.9	97.5	98.8
2	Leave-one-subject-out	74.2	84.1	64.3
3	Balanced test set	78.4	61.9	100

Results (cont.)

The following figure displays the top-10 most important features in the data for predicting the classification of speaker group. The higher the Mean Decrease Accuracy value, the more important it is in predicting the classification.



Discussion

Aim 1: From a cognitive-prosodic perspective [6], our findings could reflect different manners in the way the two groups executed speech plans at the beginning and end of utterances.

Aims 2 and 3: The combination of features extracted with Praat and *openSMILE*, along with demographic information, are effective and representative of all the features needed for an automatic classification tool. This has been shown in other clinical groups such as depression and ASD.

Our findings highlight the merits of prosodic research in identifying subtle pathological differences, paving the way for future research in subclinical and hidden cognitive-linguistic problems.

Selected references

[1] B. Neto and M. Emília Santos, “Language after aphasia: Only a matter of speed processing?,” *Aphasiology*, vol. 26, no. 11, pp. 1352–1361, Nov. 2012. **[2]** G. DeDe and C. Salis, “Temporal and episodic analyses of the Story of Cinderella in latent aphasia,” *American Journal of Speech-Language Pathology*, vol. 29, no. 1S, pp. 449–462, Feb. 2020. **[3]** W. E. Cooper and J. M. Sorensen, “Fundamental frequency contours at syntactic boundaries,” *Journal of the Acoustical Society of America*, vol. 62, no. 3, pp. 683–692, Sep. 1977. **[4]** B. MacWhinney, D. Fromm, M. Forbes, and A. Holland, “AphasiaBank: Methods for studying discourse,” *Aphasiology*, vol. 25, no. 11, pp. 1286–1307, Nov. 2011. **[5]** F. Eyben, M. Wöllmer, and B. Schuller, “OpenSMILE: The Munich versatile and fast open-source audio feature extractor,” in *Proc. MM '10 – the 18th ACM international conference on Multimedia*, Firenze, Italy, Oct. 2010, pp. 1459–1462. **[6]** K. L. Dahl and C. E. Stepp, “Changes in relative fundamental frequency under increased cognitive load in individuals with healthy voices,” *Journal of Speech, Language, and Hearing Research*, vol. 64, no. 4, pp. 1189–1196, Apr. 2021.