PROSODIC DIFFERENCES IN MANDARIN SPEAKERS WITH ALZHEIMER'S DISEASE

Yadong Liu[*1], Arian Shamei[†1], Linda Wu[‡1] and Bryan Gick [♦1,2]
[1]Department of Linguistics, University of British Columbia, Canada
[2]Haskins Laboratory, New Haven, United States of America

1 Introduction

Alzheimer's disease (AD) is a progressive neurodegenerative disorder, and the most common cause of dementia [1]. Previous research has documented changes to the speech of AD patients, including altered voice quality [2] and reduced pitch modulation in speakers of English [3]. Martínez-Sánchez et al. [4] compared the prosodic profile of Spanish-speaking AD patients and neurotypical controls and found AD speech was characterized by a flattened prosodic profile, including reduced variability of F0 and flattened prosodic trajectories within and across syllables.

To our knowledge, no work has investigated whether a flattened prosodic profile can be observed in AD patients speaking a tonal language such as Mandarin. The present study seeks to substantiate the utility of prosodic change as an acoustic biomarker for AD in Mandarin speech. Our study provides insights into how prosodic impairment caused by AD affects users of tonal languages.

2 Methods

2.1 Participants

Speech from ten AD patients (5 male, 5 female) were extracted from DementiaBank Lu corpus [5]. Patients performed picture naming tasks and the Cookie theft picture description task [6]. As no control data was provided in the Lu corpus, speech from ten gender-matched neurotypical older controls (mean age: 65.6, range: 54-74) were extracted from YouTube interviews and talks. Speech samples in both groups were selected to provide naturalistic speech in Taiwan Mandarin. Approximately 50 seconds of continuous speech was extracted for each speaker in both groups. No demographic information was provided for AD patients in the Lu corpus.

2.2 Data processing and analysis

Each audio file was trimmed to remove speech from additional interlocutors. Trimmed files were then run through Prosogram [7] to extract prosodic features. For all files, manual pitch ranges were specified based on gender (males: 70-200 Hz, females: 100-300 Hz) and automatic syllable segmentation was employed. The glissando threshold was specified to 0.16T², DG=20, dmin = 0.035) and frame period to 0.005.

We investigated eight prosodic features, including pitch range (Pitch range), average F0 value (Mean F0), standard deviation of F0 (SD F0), percentage of nuclei with pitch change greater than 4 semitones (% dynamic nuclei), percentage of nuclei with pitch rising greater than 4 semitones (% rises), percentage of nuclei with pitch falling greater than 4 semitones (% falls), intra-syllabic pitch change per second (Intrasyll traj) and inter-syllabic pitch change per second (Intersyll traj). Using the statistical software suite R [8], a one-way analysis of variance (ANOVA) was conducted to evaluate the effect of condition for each feature.

3 Results

We present results for all eight metrics that are summarized for each group in Table 1. All measurements are taken in semitones to allow comparisons between genders. Except for mean pitch, the AD group had reduced means for all metrics. ANOVA test results demonstrated a significant effect of condition for pitch range, SD pitch, % dynamic nuclei, % rises, % falls, intrasyllabic and intersyllabic trajectories observed in the control group compared to the AD group. However, no significant difference was observed for mean pitch between the AD and the control group.

Table 1: Mean and standard deviation of eight prosodic features among AD and control speakers, and ANOVA test results between two groups for each feature.

	AD	Control	ANOVA results
	Mean (SD)	Mean (SD)	
Pitch range	11.7(2.1)	14.8(1.3)	Df =1, F=15.4, p = 0.001
Mean pitch (ST)	87.1(4.6)	86.6(3.7)	Df =1, F=0.09, p = 0.772
SD pitch (ST)	2.6(0.5)	3.6(0.4)	Df =1, F=24.3, p <0.001
% dynamic nuclei	7.3(4.2)	17.2(5)	Df =1, F=23.2, p < 0.001
% rises	0.3(0.3)	1.3(0.7)	Df =1, F=19.6, p < 0.001
% falls	7.1(4.1)	16(5.5)	Df =1, F=17.1, p < 0.001
Intrasyll traj.	11.4(4.3)	19.6(3.2)	Df =1, F=23.7, p < 0.001
Intersyll traj.	17(3.2)	27(4.5)	Df =1, F=32.5, p < 0.001

Figure 1 provides comparative prosograms illustrating prosodic trajectories in the AD (top) and the control group (bottom). Within each prosogram, the y-axis reflects pitch

[*] yadong.liu@ubc.ca
[†] arian.shamei@ubc.ca
[‡] lindaw0207@gmail.com
[♦] gick@mail.ubc.ca

range, and the x-axis reflects time. Intrasyllabic trajectories are denoted by black bars. Green and magenta contours represent absolute and band-passed intensity. Note that F0 trajectories in the AD group are observably flatter than those in the control group within individual syllables, reflecting reduced intrasyllabic trajectory measurements. Variation in F0 between syllables is also reduced, reflecting reduced intersyllabic trajectories.

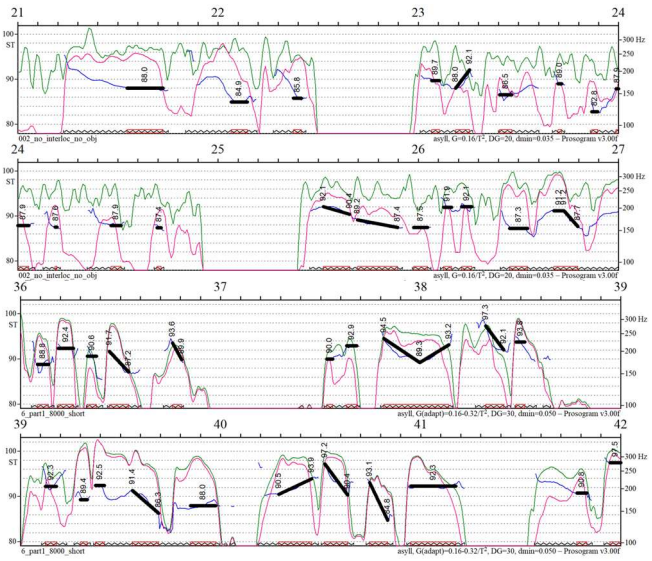

Figure 1: Comparative prosograms of an AD patient (top) and control (bottom).

4 Discussion and conclusion

The results of our prosographic analysis align with previous observations of a flattened prosodic profile for AD patients in Spanish [4] and English [3].

In the present analysis, AD patients exhibited reduced pitch range and reductions to pitch trajectories within and across syllables, comparable to those previously observed in [4]. Our data suggests that prosodic impairments observed in AD affect speakers of languages that make use of phonemic tone.

We acknowledge that differences in the nature of speech samples between AD and control groups (experimental speech task vs interview/lecture) may have contributed to differing prosodic profiles, but consistency between the present findings and previous work grant us confidence in our findings. Future work may benefit from comparing a larger number of speakers and ensuring consistency between speech tasks in the control and AD group.

Acknowledgments

The authors would like to thank Melissa Wang and Kaiwen Yu for performing data collection and transcription. Work funded by the NIH grant number DC-002717 to Haskins Laboratories and NSERC Discovery grant RGPIN-2021-03751 to Bryan Gick.

References

[1] Waldemar G, Phung KT, Burns A, Georges J, Hansen FR, Iliffe S, et al. Access to diagnostic evaluation and treatment for dementia in Europe. International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences. 2007 Jan;22(1):47-54.

[2 Luz S, Haider F, de la Fuente S, Fromm D, MacWhinney B. Alzheimer's dementia recognition through spontaneous speech: the ADReSS Challenge. arXiv preprint arXiv:2004.06833. 2020 Apr 14.

[3] Horley K, Reid A, Burnham D. Emotional prosody perception and production in dementia of the Alzheimer's type. Journal of Speech, Language, and Hearing Research. 2010; 53(5): 1132-1146.

[4] Martínez-Sánchez F, JJ GM, Pérez E, Carro J, Arana JM. Expressive prosodic patterns in individuals with Alzheimer's disease. Psicothema. 2012 Feb 1;24(1):16-21.

[5] MacWhinney, B., Fromm, D., Forbes, M. & Holland, A. (2011). AphasiaBank: Methods for studying discourse. *Aphasiology, 25*,1286-1307.

[6] Cummings L. Describing the cookie theft picture: Sources of breakdown in Alzheimer's dementia. Pragmatics and Society. 2019 Jul 5;10(2):153-76.

[7] Mertens, Piet (2004) The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. in B. Bel & I. Marlien (eds.) Proceedings of Speech Prosody 2004, Nara (Japan), 23-26 March. (ISBN 2-9518233-1-2)

[8] R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.