Analyzing speech movement variability from audio recordings

Frits van Brenk & Anja Lowit

Division of Speech and Language Therapy, University of Strathclyde, Glasgow, UK.

Contact: frits.brenk@strath.ac.uk   a.lowit@strath.ac.uk

Introduction

Measuring variability in speech motor movements

- Direct: kinematics movements by EMA, EPG, MRI and ultrasound → invasive and/or expensive
- Indirect: audio data → non-invasive and cheap

Speech parameters:

- Intensity and F0: laryngeal activity
- F1 and F2: tongue position

Variability measures:

- Spatiotemporal Index (STI), linear [1]
- Functional Data Analysis (FDA), non-linear [2]

Research question in this study:

1. What is the effect of changing speaking rate, sentence length and complexity, and performing dual motor tasks on variability in speech motor control?
2. Are variability measures of acoustic data comparable with variability results obtained from kinematic data in earlier studies?

Methodology

Participants

- Seventeen native Scottish speakers, 13 females and 4 males, age range 18 to 45 years (mean = 27.2 years, SD = 8,6 years).

Experimental task

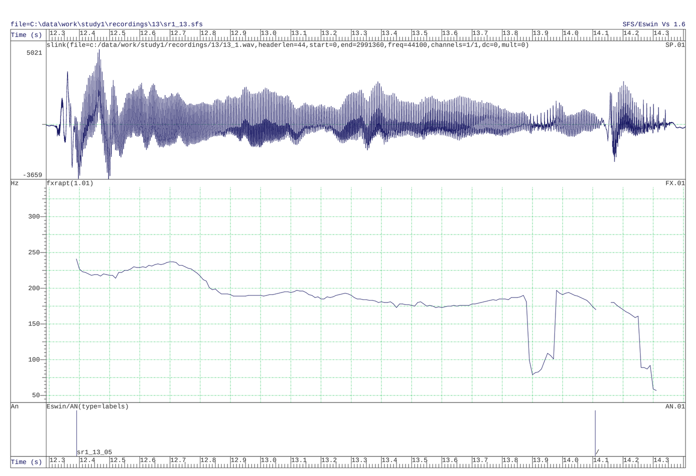
- Repeat the phrase "*Tony knew you were lying in bed*" around 20 times.

Speaking conditions:

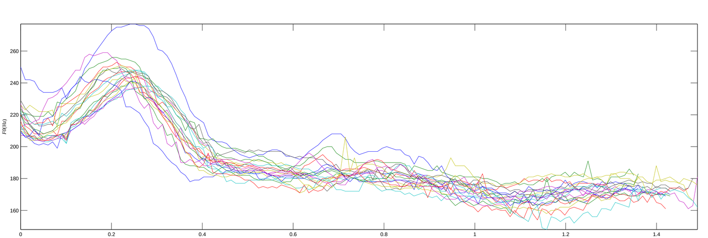
- Habitual speech rate (baseline condition)
- Slow rate
- Fast rate
- Habitual rate, increased sentence length: "*One two three Tony knew you were lying in bed five six seven*"
- Habitual rate, increased sentence length and complexity: "*I heard that Tony knew you were lying in bed this Sunday morning*"
- Habitual rate with simultaneous spiral drawing

Instrumentation and analysis

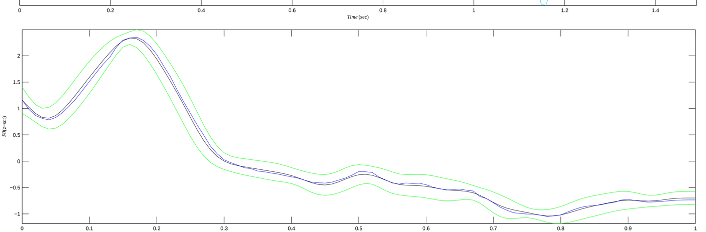
- Audio data collected with portable wave-recorder and head–mounted microphone.
- Annotation and extraction of amplitude envelope, F0-, F1- and F2 tracks in Speech Filing System.



- Variability analysis of sentence repetitions with custom Matlab software [3].
- STI: linear stretching of tracks, cumulative summing of standard deviations across tracks.

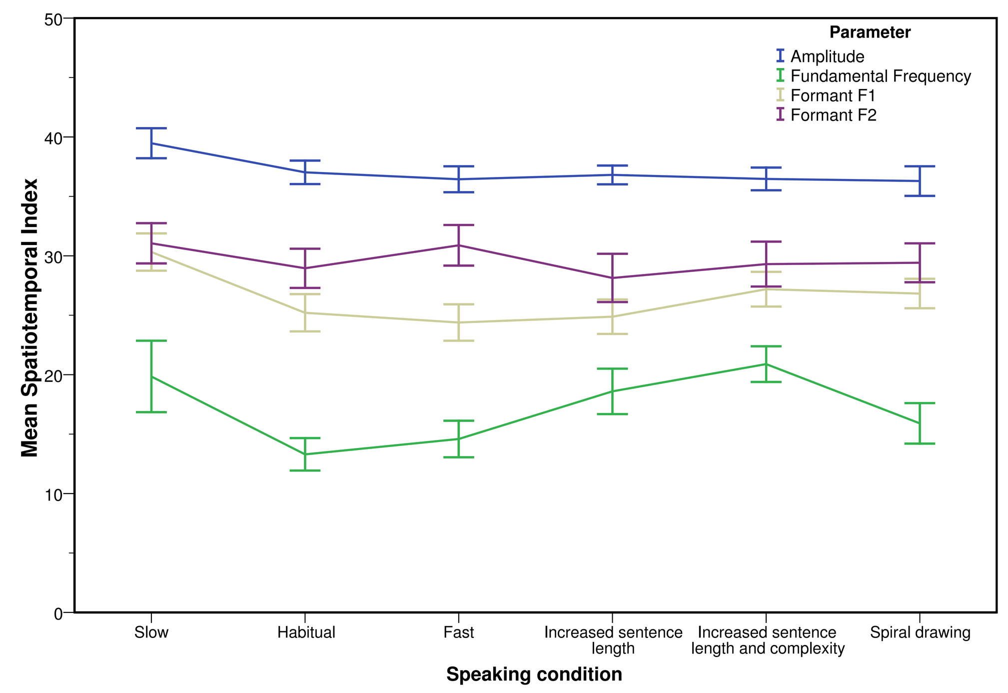


- FDA: non-linear stretching: spatial and temporal variability separately



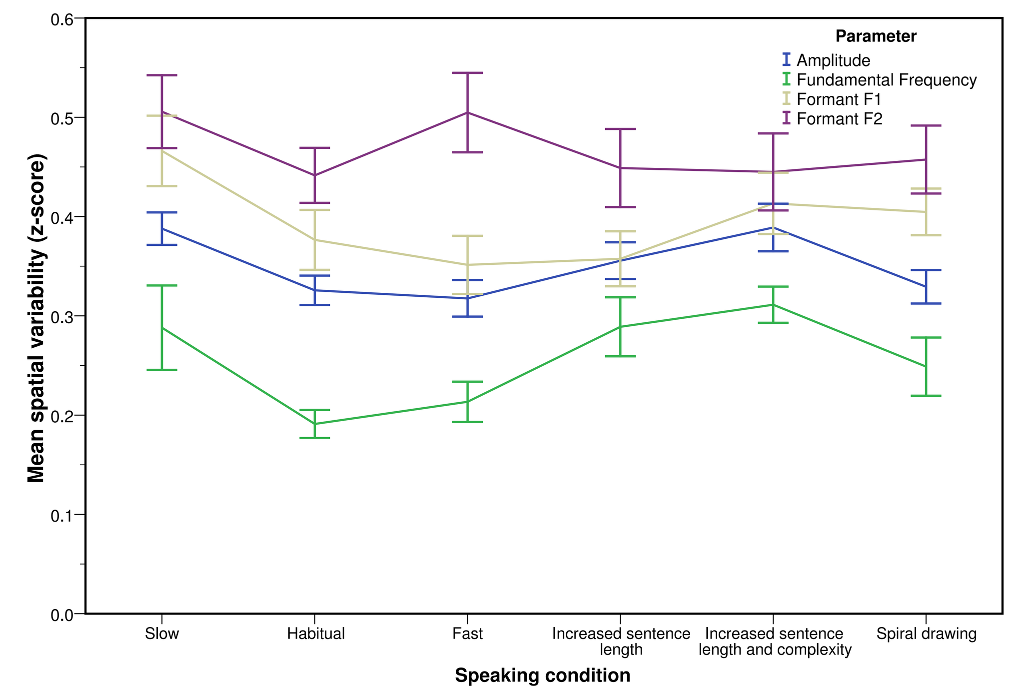
Results

Spatiotemporal Index



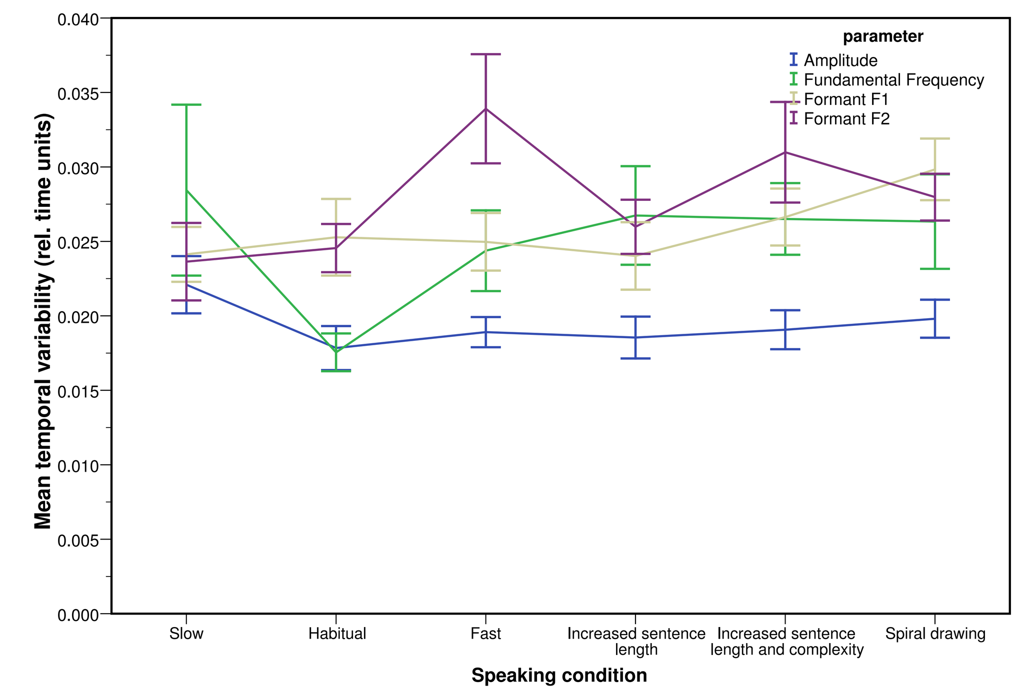
Speech Rate: STIs for F0 and F1 were significantly higher in slow compared to habitual speech rate.
Sentence complexity: STIs for F0 were significantly higher in increased length and increased length and complexity conditions, compared to habitual speech rate.
Concurrent motor task: STI in the dual motor task did not significantly differ from habitual speech rate.

Spatial Variability



Speech Rate: Spatial variability for Amplitude and F0 was significantly higher in slow speech rate compared to habitual and fast speech rate. For F1, spatial variability was higher at slow speech rate compared to fast speech rate.
Sentence complexity: Spatial variability of Amplitude was significantly higher in sentences with increased length and complexity compared to the baseline sentence. For F0, variability in sentences with increase length and increased length and complexity was higher compared to the baseline sentence.
Concurrent motor task: Spatial variability in the dual motor task did not significantly differ from habitual speech rate.

Temporal variability



Speech Rate: Temporal variability for F0 was higher at slow speech rate compared to habitual speech rate. For F2, temporal variability was higher in fast speech rate compared to slow and habitual speech rate.
Sentence complexity: Temporal variability of F0 was higher at slow and fast speech rate, compared to sentences at habitual speech rate.
Concurrent motor task: Temporal variability of F0 was higher in the dual motor task compared to the baseline condition.

Discussion

Variability in audio recordings

- For most speech parameters, STI and spatial and temporal variability are lowest in baseline condition.
- Modifying speaking rate from habitual to fast or slow: increase in variability ⇒ reflecting earlier results on lower lip movement variability by Smith et.al. (1995) [1].
- Increasing sentence length and complexity also resulted in an increase in variability, contradicting earlier results on lower lip variability by Kleinow et.al. (2000) [4].
- An increase in variability was found during the dual task condition (although limited to temporal variability of F0) ⇒ reflects the findings of Dromey et.al. (2003) [5].

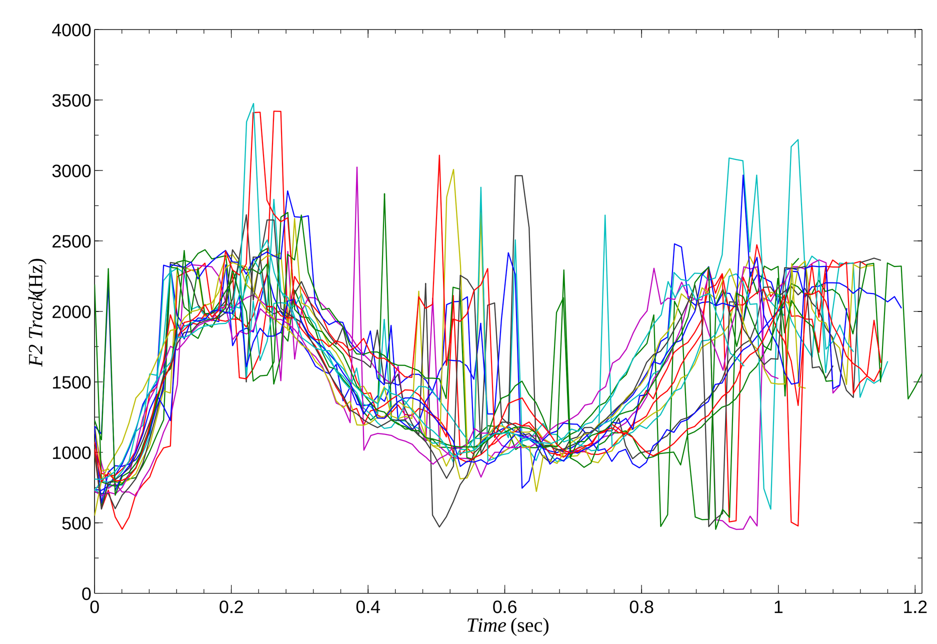
Methodological issues

Audio recordings

- Continuous voicing is necessary ⇒ speech materials contain preferably only sonorants.
- Excellent audio quality necessary. Background sounds results in interrupted voicing.
- Solution for voicing interruptions: interpolation algorithms for F0 track.

Analysis

- Accuracy of formant estimation in SFS is sometimes problematic.
- Solution: remove outliers by iterative re-assignment of peaks based on mean trajectories.



Conclusions

- Measuring variability in audio data is a promising and easy applicable method to analyse speech motor control.
- Results are similar to direct measures of variability.
- However, possible problems with data collection and processing may lead to a decrease in sensitivity compared to direct kinematic measures.

References

[1] Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., McGillem, C., "Spatiotemporal stability and patterning of speech movement sequences". Experimental Brain Research, 104:493 - 501, 1995.

[2] Ramsay, J. O., Munhall, K. G., Gracco, V. L., Ostry, D. J. "Functional data analyses of lip motion. Journal of the Acoustical Society of America, 99(6):3718 - 3727, 1996.

[3] Anderson, A., Lowit, A., Howell, P., "Temporal and spatial variability in speakers with Parkinson's Disease and Friedreich's Ataxia", Journal of Medical Speech - Language Pathology, 16(4):173 - 180, 2009.

[4] Kleinow, J., Smith, A., "Influences of length and syntactic complexity on the speech motor stability of the fluent speech of adults who stutter". Journal of Speech, Language, and Hearing Research, 43:548 - 559, 2000.

[5] Dromey, C., Benson, A., "Effects of concurrent motor, linguistic, or cognitive tasks on speech motor performance", Journal of Speech, Language and Hearing Research, 46(5):1234 - 1246, 2003.