Acoustic Correlates of Intelligibility in Dysarthria: Findings from Between-Speaker Hybridization

Frits van Brenk[1], Alexander Kain[2], Kris Tjaden[1]

[1]Department of Communicative Disorders and Sciences, University at Buffalo
[2]Department of Pediatrics, Oregon Health & Science University

INTRODUCTION

In Search of Acoustic Correlates of Intelligibility Variation

- Understanding speech production characteristics underlying intelligibility variation is an important goal of dysarthria research:
  - Advances conceptual understanding of intelligibility
  - Helps shaping targeted and patient-tailored treatments that address specific and predefined speech production variables contributing to reduced intelligibility

Hybridization

- Analysis – hybridization - resynthesis paradigm allows for selection (and interpolation) of specific acoustic parameters from two or more distinct versions of the same sentence (Kain et al., 2008)
- Directly investigate the link between acoustics and intelligibility, both in neurotypical speakers and speakers with speech disorders
- Previous research involving hybridization successfully identified acoustic causes of:
  - improved intelligibility during a clear speaking style in hypokinetic dysarthria due to Parkinson's Disease (Tjaden et al., 2014)
  - intelligibility gains and losses during a slow speaking style in ataxic-spastic dysarthria due to Multiple Sclerosis (van Brenk et al., 2021)

PURPOSE

- Explore feasibility of between-speaker hybridization and resynthesis
- Investigate the acoustic basis of intelligibility variation of male speakers with Parkinson's Disease by blending acoustic properties of sentences produced by speakers with known differences in baseline intelligibility.

METHODOLOGY

Speaker Participants

- 5 male American English speakers
- Speakers were selected from an existing database to represent a range of previously established *relative* intelligibility scores amongst a pool of 28 male speakers
- Speakers were selected approximating 0, 25, 50, 75, and 100 percentiles of sentence transcription scores, as reported in Stipancic et al., 2016
- Base Speaker (sentence transcription score):
  - PD_50percentile (68.8)
- Donor Speakers (sentence transcription score):
  - PD_0percentile (18.4)
  - PD_25percentile (60.4)
  - PD_75percentile (83.6)
  - CON_100percentile (92.0)

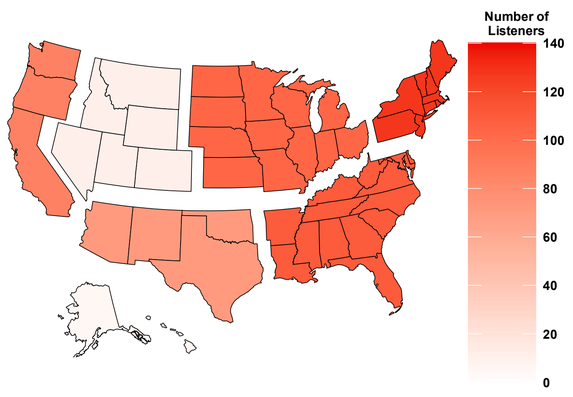
Speech Production Task

- 25 Harvard Psychoacoustic Sentences
  - Seven to nine words; five keywords each
  - Produced at habitual rate and loudness
  - Random selection of 10 sentences

Listener Participants

- 521 crowdsourced adults (252 F, 260 M, 9 O),
- Age range 18-74 y/o (*M* = 31.0, *SD* = 10.6)
  - Recruited with Prolific.co
  - On-line experiment hosted at Pavlovia.org
  - Approval rate ≥ 80%
  - Confirmed status of U.S. residence
  - Self-reported native speakers of American English
  - No Hx of speech, language, or hearing problems (self-reported)

Fig 1. Distribution of listeners by US region



Perceptual Task

- Online sentence transcription task
- For each stimulus at least 8 valid transcription scores were obtained

Transcription Analysis

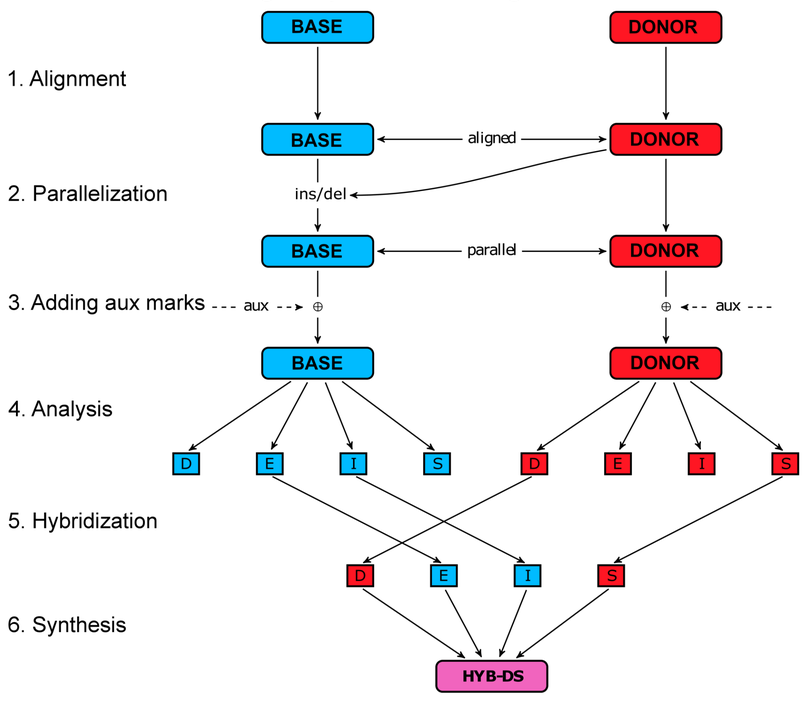
- Outcome measure: rationalized arcsine unit (RAU) transformed percentages of correctly transcribed keywords
- Statistics: Linear Mixed Models for each Base-Donor pair, with Sentence Variant as main factor of interest

METHODOLOGY CONTINUED

Hybridization and Stimuli Preparation

- All sentences were normalized for energy
- Residual-excited LPC waveform resynthesis of sentences produced by the base speaker (Kain et al., 2008)
- Imposing acoustics from sentences produced by donor speakers:
  - I. Energy Envelope (E)
  - II. F0 Envelope (I)
  - III. Segment Durations (D)
  - IV. Short-term Spectra (S)
  - V. Duration and Spectrum (D+S)
  - VI. Prosody (D + E + I)
- Total of 320 stimuli: 4 donor-base pairs x 10 sentences x 8 variants (6 hybrids, 1 base, 1 donor)
- Stimuli were mixed with 10-talker babble at SNR of 0 dB to avoid ceiling effects

Fig 2. Hybridization diagram

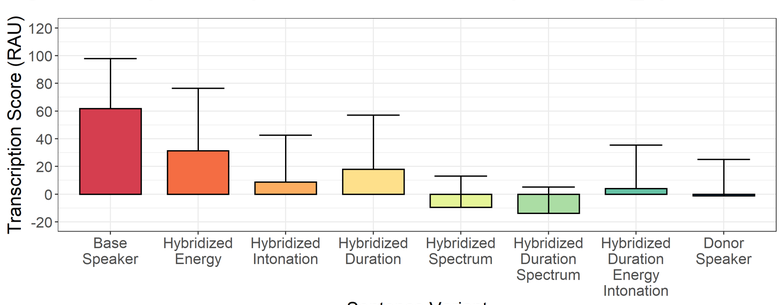


RESULTS

Intelligibility: Base more intelligible than Donors

Which acoustic variables contributed to the decrease in intelligibility when hybridizing using Donors less intelligible than the Base speaker?

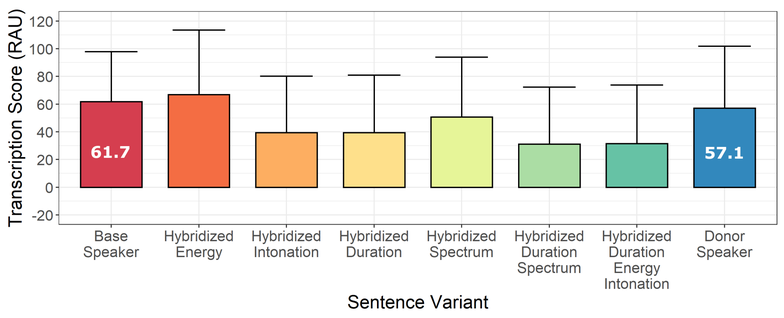
Fig 3. Base speaker hybridized with donor speaker PD_0percentile



Notable findings:

- Lower intelligibility for donor speaker, compared to base speaker ($p < .001$)
- Lower intelligibility for all hybrid variants, compared to the base speaker (all $p < .001$)
- Primary driver of intelligibility decline is Duration+Spectrum hybrid

Fig 4. Base speaker hybridized with donor speaker PD_25percentile



Notable findings:

- No difference in intelligibility between donor speaker and base speaker ($p = .84$)
- All hybrid variants except Energy ($p = .92$) had lower intelligibility, compared to base speaker (all $p < .001$)

LINKS TO EXAMPLE SENTENCES

Go to http://www.acsu.buffalo.edu/~brenk/BSMCS/ to listen to audio examples of the base speaker, donor speakers, and hybridized sentences

REFERENCES

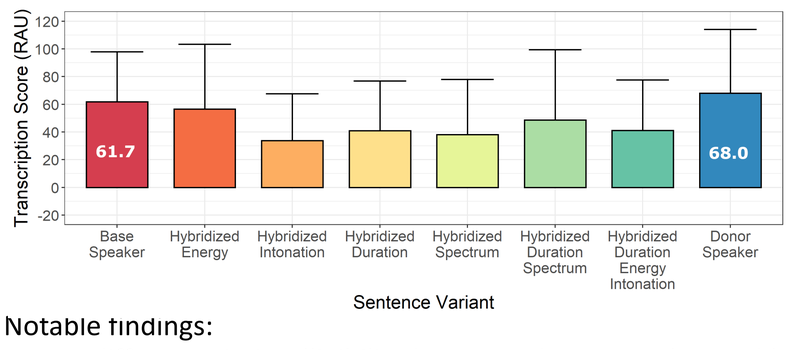
- Kain, A., Amano-Kusumoto, A., & Hosom, J. P. (2008). Hybridizing conversational and clear speech to determine the degree of contribution of acoustic features to intelligibility. *The Journal of the Acoustical Society of America*, *124*(4), 2308-2319.
- Stipancic, K. L., Tjaden, K., & Wilding, G. (2016). Comparison of intelligibility measures for adults with Parkinson's disease, adults with multiple sclerosis, and healthy controls. *Journal of Speech, Language, and Hearing Research*, *59*(2), 230-238.
- Tjaden, K., Kain, A., & Lam, J. (2014). Hybridizing conversational and clear speech to investigate the source of increased intelligibility in speakers with Parkinson's disease. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1191-1205.
- van Brenk, F., Kain, A., & Tjaden, K. (2021). Investigating Acoustic Correlates of Intelligibility Gains and Losses During Slowed Speech: A Hybridization Approach. *American Journal of Speech-Language Pathology*, 1-18.

RESULTS CONTINUED

Intelligibility: Base less intelligible than Donors

Which acoustic variables contributed to increased intelligibility when hybridizing using Donors more intelligible than the Base speaker?

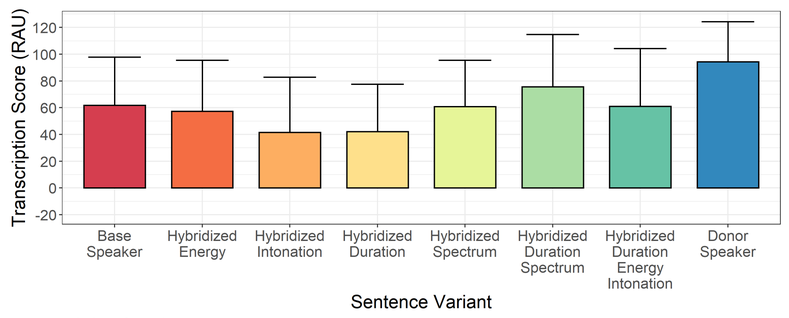
Fig 5. Base speaker hybridized with donor speaker PD_75percentile



Notable findings:

- No difference in intelligibility between donor speaker and base speaker ($p = .92$)
- No hybrids significantly more intelligible compared to the base speaker

Fig 6. Base speaker hybridized with donor speaker CON_100percentile



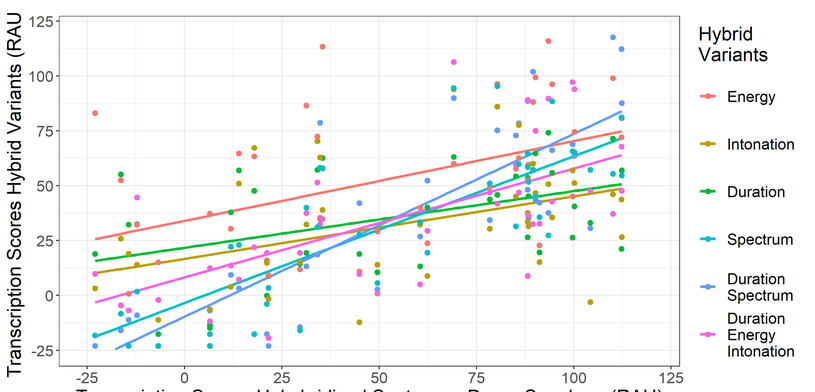
Notable findings:

- Higher intelligibility for donor speaker, compared to base speaker ($p < .001$)
- Higher intelligibility for Duration + Spectrum hybrid, compared to base speaker ($p = .03$)

Correlating unhybridized sentences of donor speakers and base-donor hybrid sentences

Which acoustic variables contributed most to the intelligibility variation found in the four donor speakers?

Fig 7. Correlations between unhybridized sentences and hybrid variants of four donor speakers



Pearson Correlations:

- Energy: $R^2 = .218$, $p = .002$
- Intonation: $R^2 = .182$, $p = .006$
- Duration: $R^2 = .204$, $p = .003$
- Spectrum: $R^2 = .642$, $p < .001$
- D + S: $R^2 = .679$, $p < .001$
- D + E + I: $R^2 = .430$, $p < .001$

DISCUSSION & CONCLUSION

Comparisons between base, donor, and hybrids

- Blending acoustic properties of base speaker with acoustic variables of a less intelligible speaker leads to decrease in intelligibility of most resulting hybrids (Figs 3 and 4)
- Energy properties may contribute as a preserving effect on intelligibility
- Blending acoustic properties of base speaker with acoustic variables of a more intelligible speaker leads to increase in intelligibility of Spectrum+Duration hybrids (Figs 5 and 6)
- Intelligibility differences between speakers less pronounced compared to lab-sourced transcription results (Stipancic et al., 2016)

Correlations between donor and base-donor hybrids

- Correlations indicated that Spectrum, Duration+Spectrum, and Intonation+Energy+Duration hybrids were the strongest predictors of intelligibility variation of the donor speakers
- Overall results indicate that primarily segmental and to a lesser extent suprasegmental properties of the acoustic signal mediate intelligibility variation associated with speakers with Parkinson's Disease

Implications

Hybridization is a powerful technique to select and interpolate segmental and suprasegmental acoustic variables between speakers of varying severity:

- Reconstructed sentences following hybridization using different speakers yield speech products of high quality, and valid targets for further perceptual and acoustic experimentation
- Allows identification of acoustic properties c*ausative for* (as opposed to correlate with) intelligibility variation in speakers with Parkinson's Disease

University at Buffalo The State University of New York

OHSU

brenk@buffalo.edu
kaina@ohsu.edu
tjaden@buffalo.edu

@fvbrenk

Department of Communicative Disorders and Sciences
Motor Speech Disorders Laboratory
ubwp.buffalo.edu/msdlab/

This study was funded by grant NIH-NIDCD R01DC004689 (Kris Tjaden)