126

Controlling for Speaker Variability within a Speech Stimuli Database

Molly Beier, Z. Ellen Peng, Sara Misurelli, Ruth Litovsky
University of Wisconsin-Madison, USA

Undergraduate Research Symposium
Union South
April 13th, 2018

Binaural Hearing and Speech Laboratory

INTRODUCTION

Testing how well people understand speech is important for documenting typical function and for diagnosing clinical conditions. Historically, standardized "speech corpuses" were developed by different groups and recorded with different speakers; hence introducing confounds to generalizing findings across studies.

The **Binaural Hearing and Speech Lab – Speech Database (BHSL-SD)** controls for speaker variability by:

- Recording the same speaker for each corpus
- Identifying each speaker's fundamental frequency
- Identifying each speaker's speech rate
- Identifying each speaker's vowel space area

These materials are unique in their suitability for studying speech perception in a wide age range and in clinical populations. This project involves careful characterization of the speakers and will contribute to the release of the BHSL-SD resource nation-wide.

CORPUSES

A corpus is a collection of recorded utterances used as the basis for language analysis. The BHSL-SD uses the following corpuses:

Words:

- PeRFeCT
  - Monosyllabic rhyming words
- CRISP[1] and CRISP Jr.[2]
  - Monosyllabic and bisyllabic

Sentences (Closed sets): Limited number of sentence options and organization (small set size).

- Oldenburg Matrix Sentences[3]
- Kidd Matrix Sentences

Sentences (Open sets): Many different sentence options and organization (large set size).

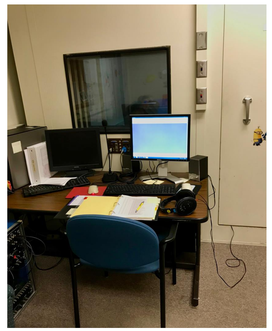
- AuSTIN [4]
  - Ex: *The man is playing the drums.*
- AzBio [5]
  - Ex: *The dog ate the snow.*
- Harvard IEEE [6]
  - Ex: *Steam hissed from the broken valve.*

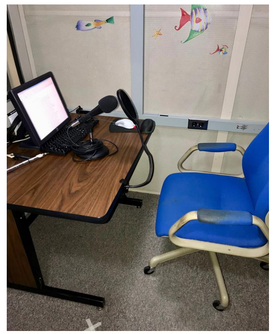
Continuous Discourse:

- Non-fiction science stories from Time for Kids magazine
  - Ex: *"Cephalopods live in all of the world's oceans…"* [7]

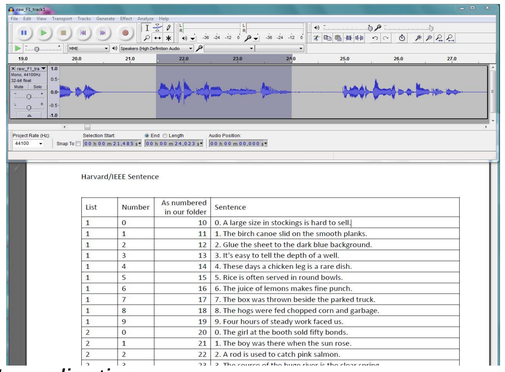
METHODS

- *Recording*
  - Sound booth
    - One person outside booth records and monitors sentence accuracy and sound levels
    - One person inside booth reads scripts into a microphone
  - Audacity
  - RME Babyface Sound Card





- *Editing & Cutting*
  - Select and listen to the desired waveform
  - Compare it to the script listed in the corpus



- *Normalization*
  - The edited recordings are normalized to the same intensity using (1) RMS-based or (2) loudness-based (EBU R128) algorithms

SPEAKERS

Criteria: Speakers must:

- Have a similar accent
- Put equal and normal stress on words and sentences
- Have a clear voice
- Have a fundamental frequency that fell within the typical range for their sex
  - Males (M): 85-155 Hz
  - Females (F): 165-255 Hz

Speaker	F1	F2	M4	M6
IEEE Speech Rate (syllable/sec)	3.4	3.7	3.1	3.6
Story Speech Rate (syllable/sec)	3.9	3.8	4.1	4.2
Fundamental Frequency (Hz)	177	219	130	112
Region	Wisconsin	Arizona	Wisconsin	Missouri

- Praat Speech Analysis program was used to find the speakers' rates of speech and fundamental frequencies. [8]
- Each speaker has similar speech characteristics and falls within the normal range for their gender.

FUTURE USE

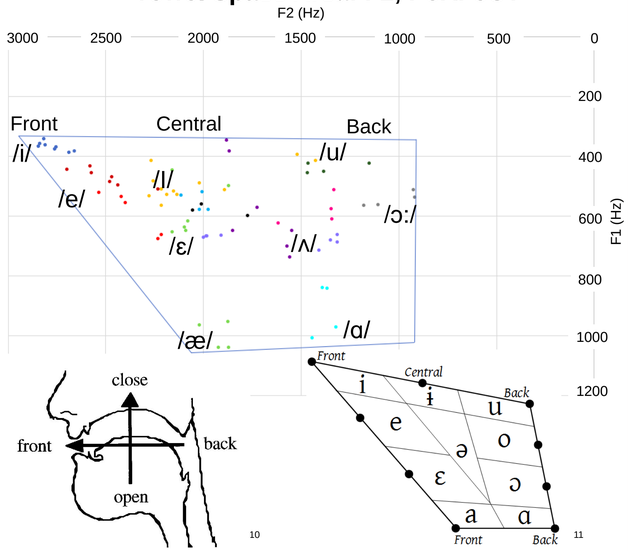
Using this speech database with detailed and controlled speaker characteristics in laboratory environments, we can study a wide range of topics related to how adults and children understand speech in everyday, complex acoustic environments, e.g., restaurants and classrooms:

- The effect of increasing informational masking on speech in noise recognition
- How does pitch difference promote spatial release from masking?
- How does this affect individuals using assistive hearing devices, e.g., cochlear implants and hearing aids.?

Vowel Space Area

- Refers to the two-dimensional area bounded by lines connecting the first and second formant frequency coordinates (F1/F2) of vowels.[8]
- Used for studying speech production deficits and reductions in intelligibility and vowel distinctiveness.[8]
- Vowels are defined by their position in space (the highest point of the tongue).[8]

Vowel Space Area: F2, PeRFeCT



CONCLUSION

The BHSL-SD is a valuable tool that controls for speaker variability in speech perception studies. The vowel space area plots for each speaker will be used to map the word list corpuses and compare regional dialect differences in speakers. Ultimately, this tool is designed to be available to the public to reduce confounds in future speech perception studies, with detailed and precise analytical information in quantifying talker variability.

REFERENCES

[1] Litovsky, R. Y. (2005). Speech intelligibility and spatial release from masking in young children. *The Journal of the Acoustical Society of America, 117*(5), 3091-3099.
[2] Garadat, S. N., & Litovsky, R. Y. (2007). Speech intelligibility in free field: Spatial unmasking in preschool children. *The Journal of the Acoustical Society of America., 121*(2), 1047-1055.
[3] Kollmeier, B., Warzybok, A., Hochmuth, S., Zokoll, M. A., Uslar, V., Brand, T., & Wagener, K. C. (2015). The multilingual matrix test: Principles, applications, and comparison across languages: A review. *International Journal of Audiology, 54*(sup2), 3-16.
[4] Dawson, P. W., Hersbach, A. A., & Swanson, B. A. (2013). An adaptive Australian sentence test in noise (AuSTIN). *Ear and Hearing, 34*(5), 592-600.
[5] Spahr, Anthony J., et al. (2012). Development and validation of the AzBio sentence lists. *Ear and hearing 33.1*(112).
[6] Rothauser, E. H., et al. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoustic 17.3* (225-246).
[7] Kletter, M. (2016). Weeds of the Sea. https://www.timeforkids.com/
[8]De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior research methods, 41*(2), 385-390.
[9] Sandoval, S., et al. Automatic assessment of vowel space area. *The Journal of the Acoustical Society of America, 134(5),*
[10] Knútsson, P. Vowel Space. https://notendur.hi.is/peturk/KENNSLA/02/TOP/VowelSpace.html
[11] Donnideau, A. What are the places of articulation of vowels. (2017). https://www.quora.com/What-are-the-places-of-articulation-of-vowels

ACKNOWLEDGEMENTS

I would like to thank Ellen Peng, Sara Misurelli, McKenzie Klein, Shelly Godar, and Ruth Litovsky for helping and encouraging me to follow this project and learn new and exciting aspects of the massive field of Communication Sciences and Disorders. This work was supported in part by a core grant from the NIH-NICHD (U54 HD090256 to Waisman Center), The work was funded by grants from NIH-NIDCD (R01DC003083 and R01DC008365), and in part by a core grant from the NIH-NICHD U54 HD090256 to Waisman Center.

Printing courtesy of the Friends of the Waisman Center