



The Recorder's Paradox: high fidelity audio from natural environments

Paul De Decker

Department of Linguistics, Memorial University of Newfoundland



Introduction

Vernacular speech styles [1] appear in informal environments which are often too noisy to obtain the high quality audio needed for acoustic phonetic analysis. As a result sociolinguistic interviews are rarely conducted in "natural environments". This poses a problem for the analysis of vernacular speech. Studies show the effects of noise on voice quality properties [2] but are noisy environments actually detrimental to analysis of formant frequencies? Does noise affect all formants equally? What types of noise should be avoided?

Method

- Speech from 1 male/1 female; early 20's; NL, Canada
- iPad2; iSLR Field Recorder [3]; 16 bit, 44.1 kHz WAV
- Carrier phrase "The man said X again" x 260
- 3 noise conditions: 60Hz Hum, White Noise, Overlapping Speech
- Noise (dB) scaled [4] to +0, +10 and +20 dB of avg. intensity of "clean" recordings [2]
- Mixed with "clean" recordings; FFMPEG [5]
- Vowel temporal boundaries aligned using Prosody-lab Aligner [6]
- F1, F2 at temporal midpoint: MAN, SAID, "X", AGAIN; automatically (Praat)

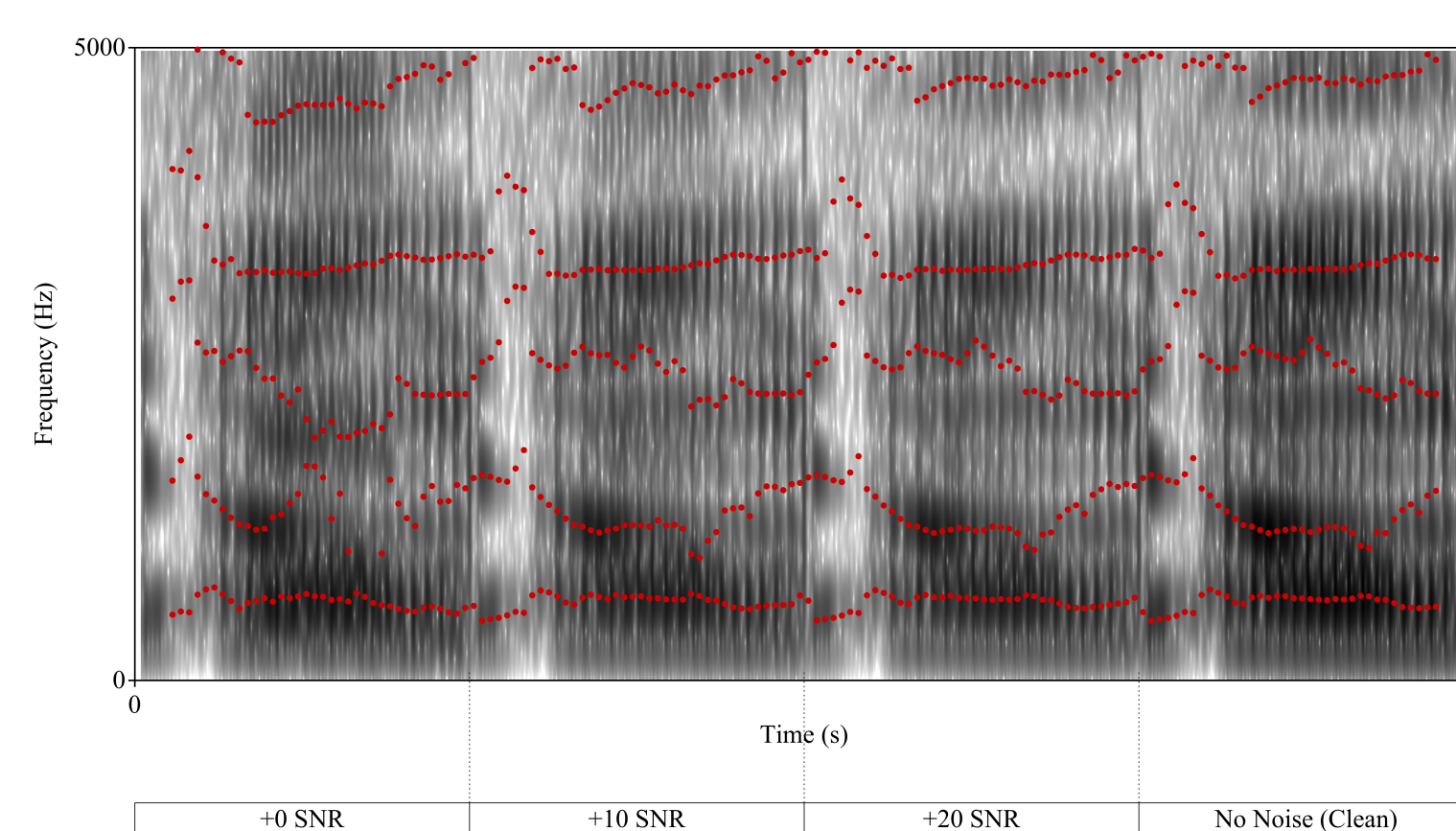


Figure 1: "OWN" mixed with OLSp and with no noise added

+0 SNR Results

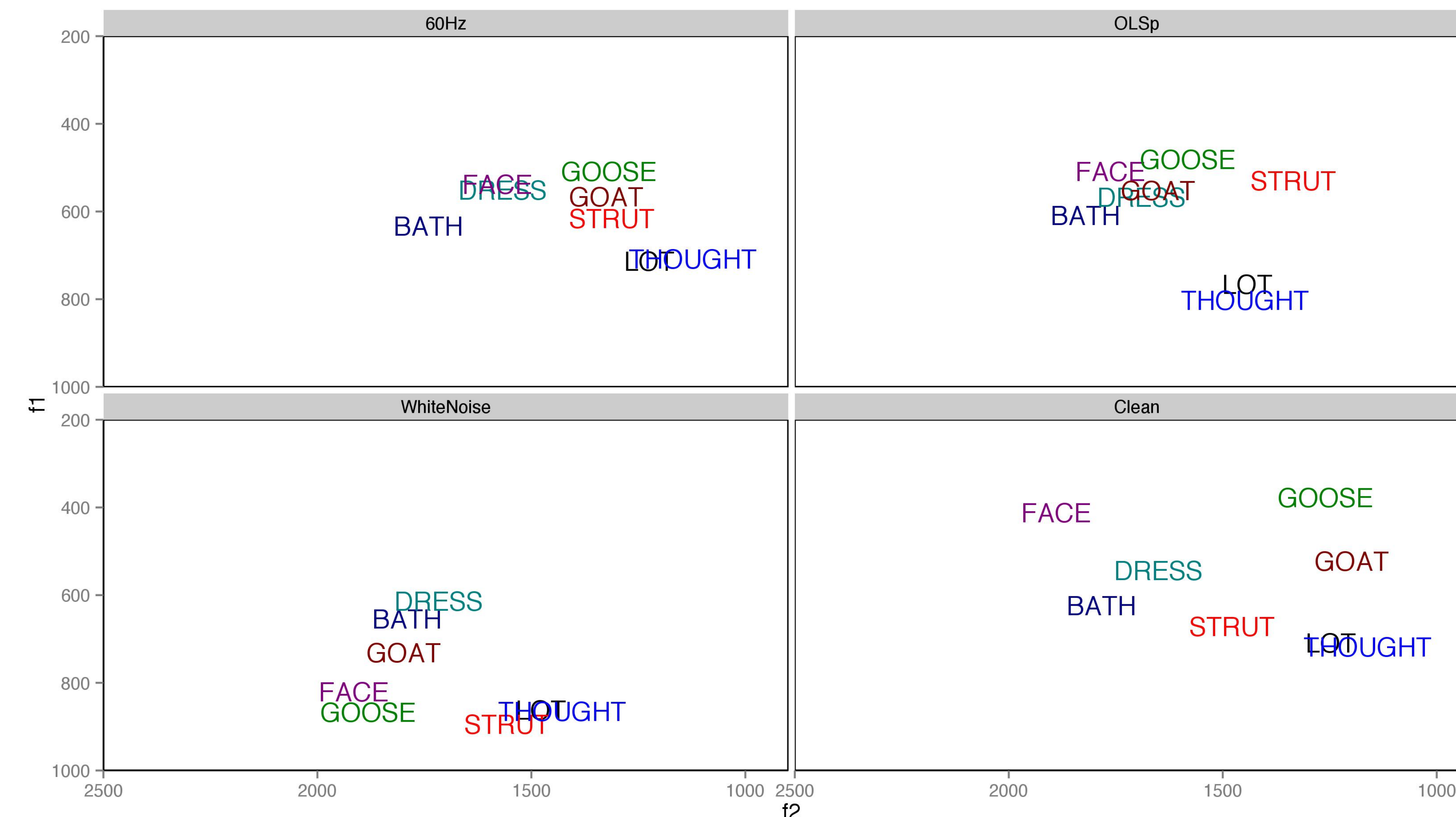


Figure 2: Noise Effects on Male Speaker Vowel Space

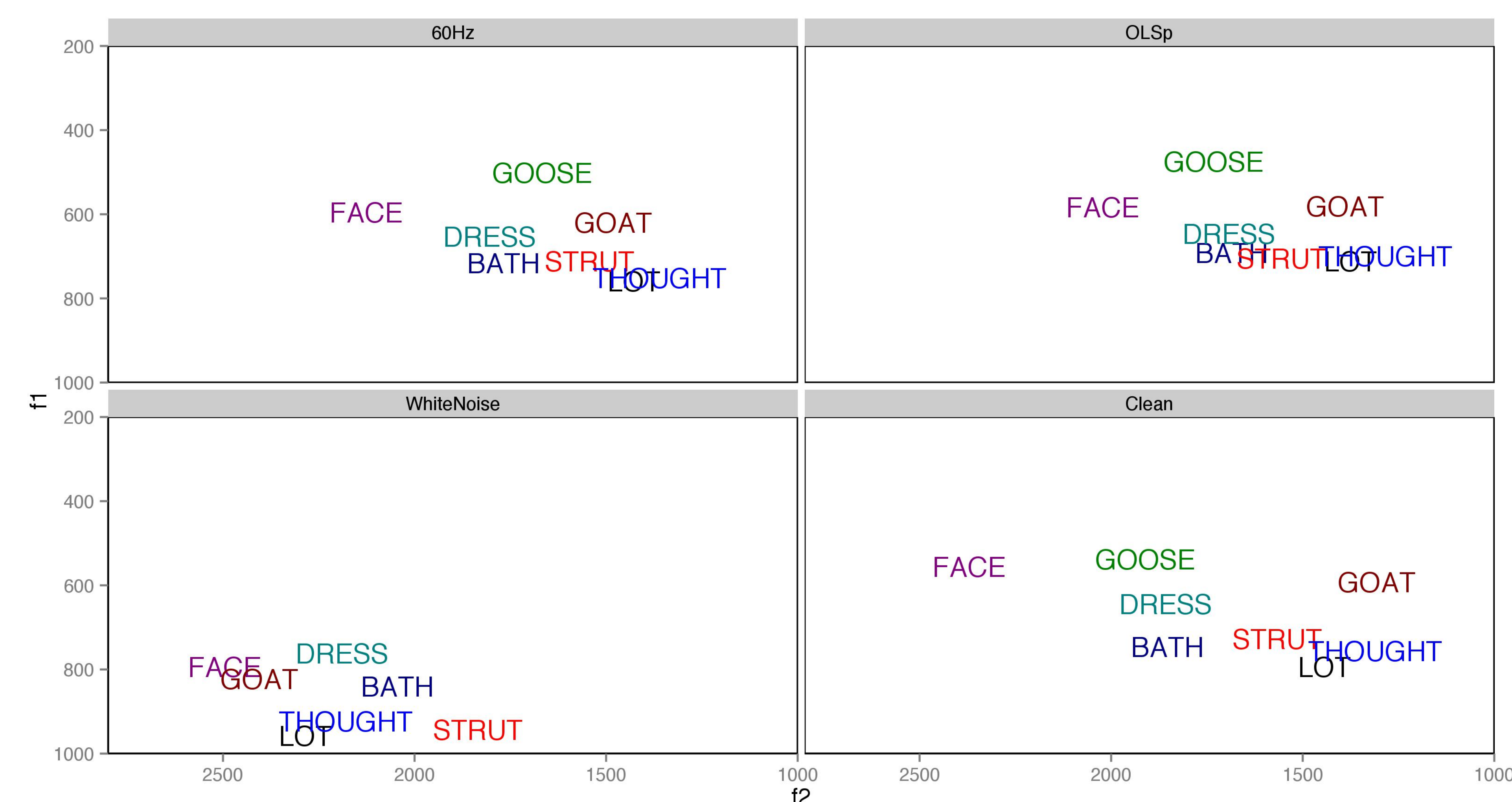


Figure 3: Noise Effects on Female Speaker Vowel Space

Main Findings

- All noises produced measurement error
- Most detrimental noise: White Noise
- All vowels were affected, except FACE (male) and STRUT (female)
- F2 frequencies altered more often than F1

Conclusion

When signal level = noise level automatic measurements of F1 and F2 are significantly affected under each noise condition. White Noise has the largest detrimental effect, probably due to its broad band of frequencies, followed by OLSp and the 60Hz Hum. Some formants and vowel categories are affected more often than others. Interaction between Gender and noise - different noises yield different types of error, depending on speaker gender.

References

- [1] W. Labov, "Field methods of the project on linguistic change and variation. language in use: Readings in sociolinguistics, ed. by john baugh and joel sherzer, 28-53," 1984.
- [2] D. R.-S. Ingrisano, C. K. Perry, and K. R. Jepson, "Environmental noise: A threat to automatic voice analysis," *American Journal of Speech - Language Pathology*, vol. 7, p. 91, Feb. 1998.
- [3] D. Alexandra and C. Coey, "islr field recorder, version 1.1," 2013.
- [4] P. Boersma and D. Weenink, "Praat, a system for doing phonetics by computer," 2001.
- [5] "Ffmpeg (<http://www.ffmpeg.org>),"
- [6] K. Gorman, J. Howell, and M. Wagner, "Prosodylab-aligner: A tool for forced alignment of laboratory speech," *Canadian Acoustics*, vol. 39, pp. 192-193, Sept. 2011.

Contact Information

- Twitter: @pmdedecker
- Email: pauldd@mun.ca