Machine learning techniques applied to singer classification

Antonio C. S. Pessotti

Laboratory of Phonetics and Psycholinguistics (Lafape), State University of Campinas

antoniopessotti@gmail.com

Singers have different degrees of training, frequently related to individual experience, such as practice and scholarship. Several studies report comparisons between soloists and choristers through the analysis of acoustic measurements and ensuing articulatory inferences. Among them, Rossing, Sundberg, and Ternström (1986) is noteworthy for having found that the singer's formant is more prominent in soloists than in choristers. In addition, soloists sing lyrics in a more distinct way than choristers. Another work Rossing, Sundberg, and Ternström (1987) showed prominence between 2-4 kHz in women soloists, as well as articulation closer to speech in choristers. Other studies (Garnier et al., 2008;2010) point to singing strategies that differentiate singer groups. These works suggest that there are patterns to be explored to distinguish such groups.

Pattern Recognition is a general class of methods to extract patterns from signals, and is applicable to a number of different areas. Machine learning techniques are used in Pattern Recognition. They have been applied in many areas, in many signals such as images, written text, time series of prices or weather, sounds, etc. The aim of this study is to apply some machine learning techniques to singer classification, based on segmented and tagged acoustic data.

The subjects were ten sopranos: five soloists (SOL) and five choristers (CHOR). They were recorded singing the Brazilian chamber song "*Conselhos*", composed by Carlos Gomes, whose lyrics contain 195 syllables. The recording session took place in a professional studio with 96 kHz stereo sampling rate, converted to 16 kHz mono for analysis. Sung performances were recorded ten times: five with digitized accompaniment and five without. The lyrics were read five times by each singer.

Digitized Musical accompaniment was heard through headsets. It was created from musical score with *MuseScore* 1.0. Segmentation, tagging and statistical analysis were conducted through specific scripts, created with *Praat* (Boersma and Weenink, 2014) and R (R Development Core Team, 2009).

The independent variable was singer group (SOL or CHOR). The dependent variables were: a) formants (F1, F2 and F3), in Hertz; b) the same formants converted to Bark, c) intonation (pitch) in semitones, with reference to 440Hz; d) segment duration in milliseconds.

Machine Learning techniques were applied with *Waikato Environment for Knowledge Analysis* (WEKA)(Hall et al., 2006). Employed Models were: Multilayer Perceptron (MP), Classification via Clustering, Classification via Regression, Inference and Rules-based Learner (JRip), Naive Bayes. Here we will report results on MP and JRip only.

Figure 1 shows F3 Bark (x-axis) *versus* F2 Bark (y-axis) measurements of spoken and sung speech (with and without accompaniment) as grouped by the MP model. These patterns are similar to those obtained with F2 x F1, intonation x formants, and intensity x formants. JRip produces similar patterns. The MP approach is based upon Perceptron training; the JRip approach is based upon statistical inference rules. Both models were evaluated with Bayesian Criteria described in Witten and Franck (2005), and showed good classification results, with $p < 0.0001$. Rules resulting from both techniques suggest minimal paths through which it is possible to create a direct signal classification.

Figure *1: F3 Bark (x-axis) versus F2 Bark (y-axis) measures of data; Soloists (crosses) and Choristers (balls).*

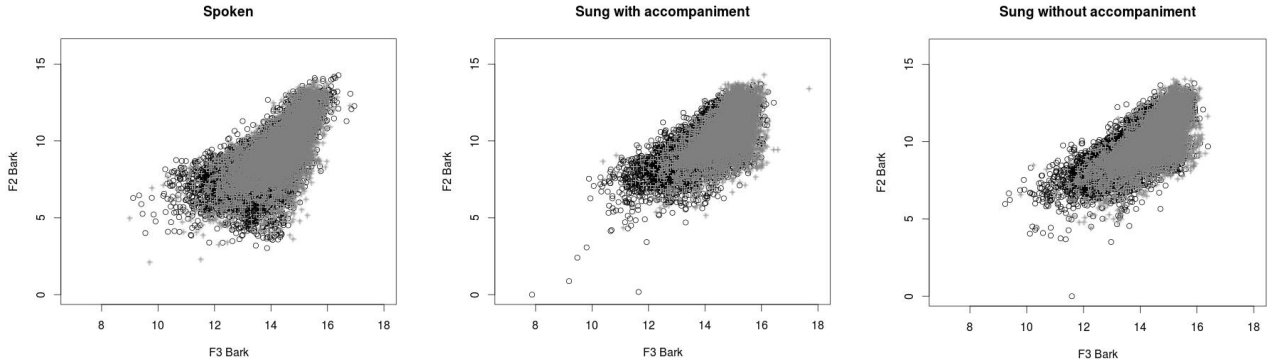

Figure 1 clearly indicates that soloists reach higher F2 and F3 than choristers, possibly in an attempt to attain the high frequency prominence described by Rossing, Sundberg, and Ternström (1986, 1987). This fact suggests that specific articulatory strategies may be employed by soloists, possibly derived from biomechanical adaptation from their practice as singers.

Results point to a precise distinction between soloists and choristers, in sung and spoken productions. Pattern recognition techniques applied here (JRip and MP) allow for future development of an automatic classification system.

References

Boersma, Paul & Weenink, David (2014). *Praat*: doing phonetics by computer [Computer program]. Current Version 5.3.76, retrieved 8 May 2014 from http://www.praat.org/.

Garnier, M., Wolfe, J., Henrich, N., Smith, J. (2008) Interrelationship between vocal effort and vocal tract acoustics: a pilot study, In *Interspeech*-2008, 2302-2305.

Garnier, M.; Henrich, N.; Smith, J.; Wolfe, J. (2010) Vocal tract adjustments in the high soprano range. *Journal of Acoustic Society of America*, 127-6.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. (2009); *The WEKA Data Mining Software*: An Update; SIGKDD Explorations, Volume 11, Issue 1.

R Development Core Team. *R*: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.r-project.org/; 2009.

Rossing, T. D., Sundberg, J. and Ternström, S. (1986) Acoustic comparison of voice use in solo and choir singing. *Journal of the Acoustical Society of America*, 79 (6), 1975-1981.

Rossing, T. D., Sundberg, J. and Ternström, S. (1987) Acoustic comparison of soprano solo and choir singing. *Journal of the Acoustical Society of America*, 82 (3), 830- 836.

Witten. I. H. and Franck, E. (2005) *Data Mining: Practical machine learning tools and techniques*. 2nd edition; Morgan Kaufmann. San Francisco.

***