Identification of Singing Style Techniques Via Spectral Analysis

Camille Noufi
Independent Researcher

ABSTRACT

In this abstract we use spectrograms to identify vocal techniques singers employ to differentiate between different styles. We identify techniques that would be useful to detect when improving or modifying the construction and/or analysis of a singing voice. This study examines the spectrograms of audio recordings sung in both contemporary and classical style. These vocal styles demand stark variations in the way a singer approaches vowels and consonants and they utilize different register placements, ultimately producing different spectral profiles. We examine the spectrograms of contemporary and classical style by the same singer to quantify 1) how a belted pitch differs in frequency content from a mixed/head tone, 2) how a voiced "r" consonant differs from a flipped "r" consonant, and 3) how vowel purity affects the frequency content of a sung syllable. This information can tell us what data patterns an analysis program may need to be aware of in processing a voice and/or what spectral content a synthesis program could modify to construct a more authentic singing voice.

1. INTRODUCTION

Computer-based vocal synthesis, analysis and transcription related to singing is challenging because of the variety of anatomical, environmental and *stylistic* factors influencing spectral content. Current singing synthesizers [1, 5] construct voices sample-by-sample, unaware of the humanistic or musical elements they implement and therefore unable to modify or improve the end product. Understanding exactly what spectral data corresponds with what singing technique allows for improvement or modification of a voice. Previous studies in this domain have sought to measure singing style based on phonation, glottal characteristics and other acoustic measurements of the voice [3]. This study differs by examining measurements related to singing technique. It focuses on using spectral analysis of digital sound samples via Fourier transform algorithms to identify commonplace techniques vocalists use when singing in one style or another.

2. METHODS

We use five recordings of a female singer trained in both classical and contemporary pop technique to eliminate environmental and anatomical variations between recordings. Antonio Caldera's 18th century classical Italian aria "Alma Del Core," Adele's 21st century pop recording "Rolling in the Deep," and a standard *solfege* vocal warmup are recorded by this singer. The chosen pieces allow for natural singing in each style while the vocal warmups provide a control variable for each style's prominent vocal techniques.

Contemporary Pop	Classical
Diphthongs	Monophthongs
Approximant "r" consonant	Plosive "r" consonant
Belt timbre	Classical mix timbre

Table 1. Vocal techniques used in classical and pop style

Song	Style	Range
Solfege Scale	Contemporary	A3 - A5
"Rolling in the Deep"	Contemporary	B3 - C#4
"Alma Del Core"	Contemporary	A4 - F#5
"Alma Del Core"	Classical	A4 - F#5
Solfege Scale	Classical	A3 - A5

Table 2. Audio recordings of classical and pop style, Key: A major

3. PRELIMINARY FINDINGS

3.1 Vowel Purity

Vowel shape can be determined through spectral analysis by examining the distance of the second formant F2 from the first formant F1. The recordings all show that pure monophthong vowels, a necessity of classical singing, results in a flat second formant throughout intonation of a syllable whereas diphthongs, accepted in contemporary style, can fluctuate during voicing of a single syllable. The spectrograms in Figure 1 shows an example of a pure [ɛ] vowel sung in the classical warmups and the corresponding [e:I] diphthong represented by the slide of F2's frequency components.

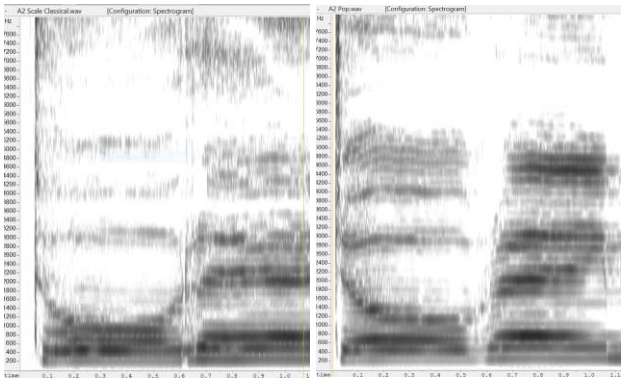

Figure 1. Beginning of A3 Solfege scale sung on *Do Re* in classical (left) and contemporary pop (right)

3.2 "R" Consonant

The "r" consonant is primarily sung with a flip or roll of the tongue in the classical style. The consonant is plosive in nature. In contemporary singing, the "r" consonant is generally employed as an approximant, meaning it is voiced more similarly to a vowel. Identifying the "r" consonant as an approximant would likely indicate singing in the contemporary style. Figure 2 shows the frequency content of the solfege spectrograms for the "d" in *Do* and the r" in *Re*. As expected, the contemporarily styled "r" contains almost all of its energy around F1 and F2 of the neighboring vowels. The classical flipped "r", like the plosive "d," contains noise energy more evenly spread across a high bandwidth.

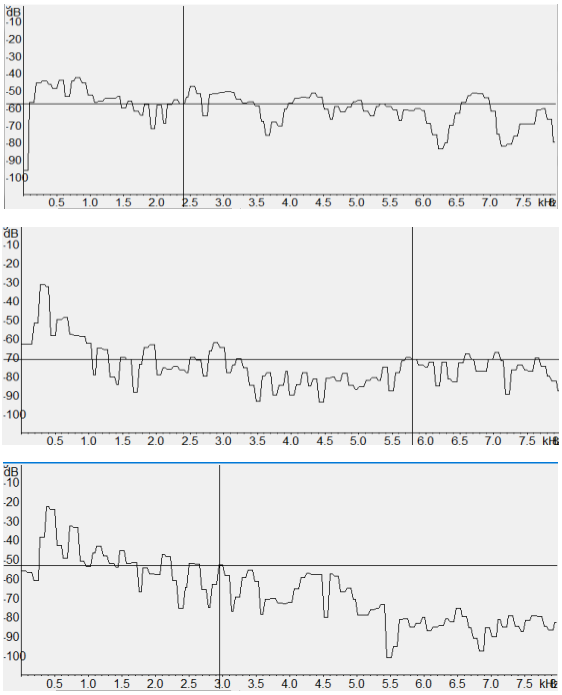

Figure 2. Spectral content of plosive "d" (top), flipped "r" (middle), and voiced "r" consonants (bottom)

3.3 Belt Timbre

The belt technique present in many styles of contemporary music is likely due to the strong presence of upper partials in the spectrum, giving way to a brighter, more forward sound. An energy concentration of F2 resonating at the second harmonic H2 is also expected. Figure 2 quantitatively displays that these anticipated techniques.

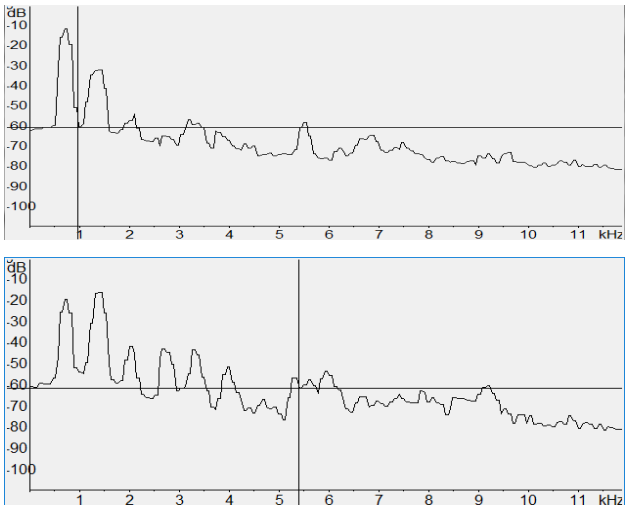

Figure 3. Spectral Average of classical head tone (top) vs. belted pop (bottom) voicing of [o] vowel, "Alma Del Core"

4. CONCLUSION AND FUTURE WORK

This preliminary research aims to quantitatively identify commonplace vocal techniques singers employ to emote classical and contemporary styles. This is a practical way to align computer understanding of singing style with human understanding.

The next steps in this research are 1) to expand the data set, 2) to determine and refine robust quantitative descriptions of the techniques described above, and 3) to develop learning algorithms to identify these techniques and classify the style of the recording.

5. REFERENCES

[1] M. Blaauw, J. Bonada: "A Neural Parametric Singing Synthesizer," *ePrint arXiv:1704.03809*, 2017.

[2] E. Godoy, N. Malyska, T. Quatieri, "Voice Quality: Part 2, Elements of Voice Quality", *MIT Lincoln Laboratory*, 2015.

[3] M. Thalen, J. Sundberg: "Describing different styles of singing: a comparison of a female singer's voice source in 'Classical', 'Pop', 'Jazz' and 'Blues'," *Logopedics Phoniatrics Vocology*, Vol. 26:2, pp.82-93, 2001.

[4] N. Henrich: "Mirroring the voice from Garcia to the present day: Some insights into singing voice registers," *Logopedics Phoniatrics Vocology*, Vol. 31:1, pp.3-14, 2006.

[5] A. van den Oord et al: "WaveNet: A Generative Model for Raw Audio," *ePrint arXiv:1609.03499*, 2017.