Music Performance Analysis Overview

At the most basic level, music performance analysis involves extracting parameters and analysing attributes from a musical performance. The difference between the two can often be vague, but can be defined as follows:

- Parameters: Details of a performance that can be immediately extracted directly from the audio file of a musical performance. E.g. Fundamental Frequencies, Sampling Rate
- Attributes: Details of a performance that requires some form of analysis before one can draw any useful information on the performance. E.g. Pitch, Spectrogram

Outside of these two, there are also concepts that can be considered more abstract, such as the feel, structure, melody and orchestration of a performance. These can be called characteristics and can be sorted into 3 more categories – Tonal, Timbral and Temporal Analysis.

Tonal Analysis

Tonal analysis involves extracting data relevant to pitch from a recording of a musical performance. Often this will involve applying a Fast Fourier transform to an audio signal and identifying the fundamental frequency from the principle frequencies [1].

Timbral Analysis

Timbral analysis involves analysing the frequency content of an audio recording in order to scientifically define the sound "quality" or "colour". This is significant due to there being a correlation between timbre and the emotion perceived [2].

Temporal Analysis

Temporal analysis deals with the analysis of an audio recording in the time domain i.e. The timing of notes, tempo, rhythm etc. This will often involve identifying the note onsets (often through a novelty and peak picking function) and then analysing attributes from this information [3].

Singing Power Ratio

Background

In 2002, a survey was undertaken where 1000 professional vocal teachers were asked which trait was most important when determining whether or not an untrained singer possessed a talent for singing [4]. Vocal timbre was amongst the top three most important traits, and understandably, improving one's vocal timbre is what a lot of singers strive to do.

One aspect of vocal timbre is what is known as the Singer's Formant (also known as Squillo in classical music). This is a frequency range between 2kHz to 4kHz, and isn't very prominent in a lot of musical

instruments. As such, a lot of professional singers will train their voice to produce more energy in this frequency range in order to sing over loud accompaniments more effortlessly. This is especially prevalent amongst opera singers in order to sing over large orchestras [5].

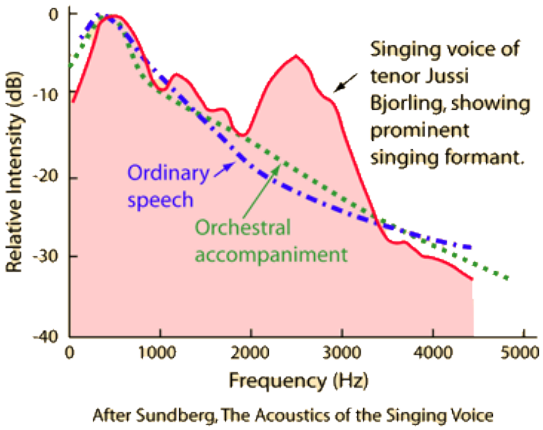

Figure 1 - Frequency energy of a trained singer compared to orchestral accompaniment and speech [6]

A measure of analysing the Singer's Formant is the Singing Power Ratio.

Analysing Singer's Formant and Singing Power Ratio

Calculating Singing Power Ratio (SPR) of an audio sample first involves performing a Fourier Transform in order to get the frequency energy of the sample. You then take the highest peak in the 2-4kHz frequency band (the Singer's Formant) and the highest peak in the 0-2kHz frequency band. These two values are then converted in decibels. The SPR is the difference between the two values. A smaller SPR indicates more of the singer's formant being utilized [7][8].

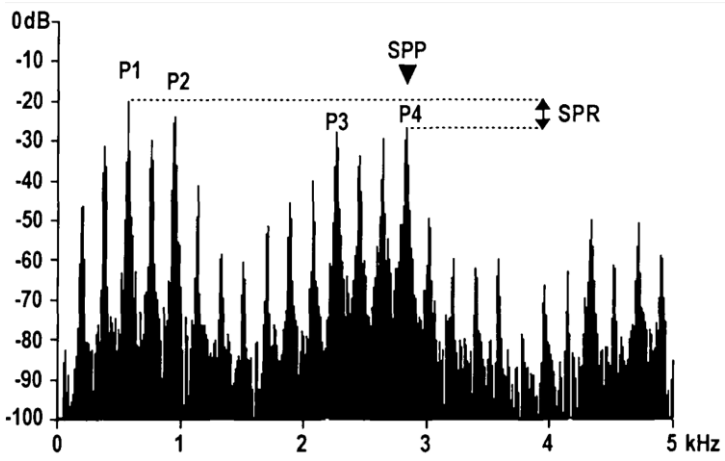

Figure 2- Power spectrum of a voice sample and the SPR [7]

The SPR Meter Plugin

The SPR Meter plugin aims to display the SPR value in a visually informative manner. This involves performing the SPR processing on incoming audio buffer data lines and displaying the value in the form of a meter.

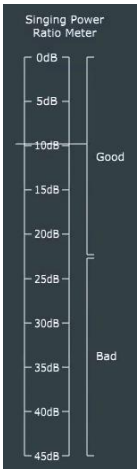

Figure 3 - The SPR Meter Audio Plugin

Threshold

The threshold between a trained and untrained singer (or the good/bad threshold on the plugin UI) was derived from several papers researching the SPR differences between trained and untrained singers. However, the averages varied substantially between different papers. One paper found a mean SPR of 13.1 dB amongst professional singers and 21.1 dB amongst non-singers [7]. Another found a mean SPR of 22.6 dB amongst trained singers and 30.7 dB amongst untrained singers [8]. The SPR meter uses a rough average between the papers, but the ideal solution would likely involve performing another experiment with a large number of participants and obtaining a threshold specifically for the plugin.

Voiced/unvoiced detection

Unvoiced audio signals can skew the SPR readings and output unwanted and incorrect data. As such, there's also a built-in unvoiced signal detection algorithm, which utilizes a zero-crossing threshold (since unvoiced signals are more likely to cross 0 than voiced signals) and a maximum auto-correlation function threshold (a measure of similarity of an audio signal with a delayed version of itself). This filters out the unvoiced signals. However, these aren't completely perfect. As such, a spectral flux threshold (a measure of change in frequency content) and a ramp was also implemented so that unvoiced signals don't suddenly cause a spike in the output.

Shortcomings

It has been noted that the 2-4kHz range of the Singer's Formant is merely an approximation. A potentially more ideal ratio that has been explored is taking the energy ratio between formants. One proposed ratio is

$$Quality\ Ratio = \frac{Area_{1,2}}{Area_{3,4,5}}$$

Where $Area_{1,2}$ is the frequency energy in formants 1 and 2 in decibels, and $Area_{3,4,5}$ is the frequency energy under formants 3, 4 and 5 in decibels [9]. This could lead to a more accurate reading for a wider range of voice types.

Conclusion

Whilst not completely perfect, the Singing Power Ratio provides a good measure into how capable a singer is with utilizing the Singer's Formant and singing over accompanying instruments. The SPR meter plugin provides a tool for singers to use when training their use of this formant, as well as being a mixing tool for music producers to use to get a rough gauge on the quality of a vocal performance.

References

[1] P. McLeod and G. Wyvill, "Visualization of Musical Pitch", *Department of Computer Science, University of Otago*.

[2] J. C. Hailstone, R. Omar, S. M. D. Henley, C. Frost, M. G. Kenward, and J. D. Warren, "It's not what you play, it's how you play it: Timbre affects perception of emotion in music," *Q. J. Exp. Psycho*l., vol. 62, no. 11, pp. 2141–2155, 11/2009.

[3] A. Lerch, *An introduction to audio content analysis*. Hoboken, NJ: Wiley, 2012.

[4] C. Watts, K. Barnes-Burroughs, M. Andrianopoulos and M. Carr, "Potential factors related to untrained singing talent: a survey of singing pedagogues", *Journal of Voice*, vol. 17, no. 3, pp. 298-307, 2003. Available: 10.1067/s0892-1997(03)00068-7.

[5] S. Lee, H. Kwon, H. Choi, N. Lee, S. Lee and S. Jin, "The Singer's Formant and Speaker's Ring Resonance: A Long-Term Average Spectrum Analysis", *Clinical and Experimental Otorhinolaryngology*, vol. 1, no. 2, p. 92, 2008. Available: 10.3342/ceo.2008.1.2.92.

[6] J. Sundberg, "The Acoustics of the Singing Voice", *Scientific American*, vol. 236, no. 3, pp. 82-91, 1977. Available: 10.1038/scientificamerican0377-82.

[7] K. Omori, A. Kacker, L. Carroll, W. Riley and S. Blaugrund, "Singing power ratio: Quantitative evaluation of singing voice quality", *Journal of Voice*, vol. 10, no. 3, pp. 228-235, 1996. Available: 10.1016/s0892-1997(96)80003-8.

[8] C. Watts, K. Barnes-Burroughs, J. Estis and D. Blanton, "The Singing Power Ratio as an Objective Measure of Singing Voice Quality in Untrained Talented and Nontalented Singers", *Journal of Voice*, vol. 20, no. 1, pp. 82-88, 2006. Available: 10.1016/j.jvoice.2004.12.003.

[9] C. Manfredi, D. Barbagallo, G. Baracca, S. Orlandi, A. Bandini and P. Dejonckere, "Automatic Assessment of Acoustic Parameters of the Singing Voice: Application to Professional Western Operatic and Jazz Singers", *Journal of Voice*, vol. 29, no. 4, pp. 517.e1-517.e9, 2015. Available: 10.1016/j.jvoice.2014.09.014.