Voice Type Discovery

Mário Amado Alves
amadoalves@fe.up.pt

Vítor Almeida
valmeida@fe.up.pt

Ricardo Sousa
sousa.ricardo@fe.up.pt

Aníbal J. S. Ferreira
ajf@fe.up.pt

Sérgio I. Lopes
sil@fe.up.pt

Departamento de Engenharia Eletrotécnica e de Computadores
Faculdade de Engenharia da Universidade do Porto

Abstract

We address the problem of automatically recognizing the voice type, or tessitura, of singers. Such procedures have applications to singing learning, training, and entertainment. The main non-trivial component of our approach is the robust extraction of the fundamental frequencies present in a sample of the singing voice.

1 Introduction

We address the problem of automatically recognizing the voice type, or tessitura, of singers. This is an exploratory study. We consider the classical definition of voice types: bass, baritone, tenor, alto, mezzo-soprano, soprano [1,2].

The automatic recognition of the voice type of a singer has applications to singing learning, training and entertainment. For example, upon voice type discovery, a karaoke-style game may automatically adapt its pitch-matching algorithm to better suit the voice type of the singers, thus allowing players with different voice types to compete fairly on songs that would otherwise favour one pitch range over the other. In fact, such a game, called SingingBattle, designed for deployment at Casa da Música in Porto, is currently under development by ourselves in the context of the ARTTS project (gnomo.fe.up.pt/~voicestudies).

2 Algorithms

We have developed software libraries and computer applications to let the users discover their voice type, by singing, to the connected microphone, the passage described on Figure 1, which is our representation of the procedure as prescribed in [2], namely:

> *Begin singing a note that is somewhere in your lower middle range. Sing a chromatic scale downwards in pitch. Write down the lowest note that you are able to vocally produce. Then, beginning at a comfortable upper-middle note, begin singing a chromatic scale upwards in pitch. Write down the highest note that you are able to sing.*

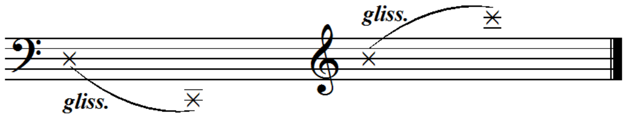


Figure 1: Singing passage designed to discover the voice type. The cross note-heads represent approximate, imprecise pitches.

The overall algorithm, depicted in Figure 2, takes the sound of the singer performing the voice discovery passage (Figure 1), extracts the fundamental frequencies involved, then finally classifies the voice type based on an analysis of these frequencies. The details of each component are described in the next subsections.

2.1 Pitch and frequency

We use the mathematical notion of musical pitch, and the widely accepted pitch notation described in [3]. Namely, we call *pitch* to the musical note related to acoustic pitch and frequency as follows. Pitch *classes* are the notes represented by the letters and/or respective number of *semitones* in Table 1.

C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	1	2	3	4	5	6	7	8	9	10	11

Table 1: Pitch class symbols and semitones.

The pitch classes divide the *octave* into twelve equal logarithmic steps of $2^{1/12}$, called *semitones.* The octave is the interval between a frequency *f* and its double 2*f*. The names *octave, semitone* and the note letters have merely historical significance.

By international convention, octaves are numbered, and each octave starts at a `C`. A pitch class *K* at octave *N* is notated K_N or $\mathtt{\mathit{KN}}$. To clarify: a pitch is a frequency. Also, by convention we have:

$$\mathtt{A4} = 440\text{Hz}. \qquad (1)$$

To find the frequency of any note we work from (1) and the above rules. To simplify, we work with *semitone indices* across all octaves defined as

$$i(K_N) = 12\ (N - N_{\mathrm{Ref}}) + i(K) - i(K_{\mathrm{Ref}}) \qquad (2)$$

where $i(K)$ is the function defined by Table 2. Then the frequency in Hertz of a pitch index is given by

$$f(i) = f_{\mathrm{Ref}}\ s^{(i - i_{\mathrm{Ref}})} \qquad (3)$$

where *s* is the semitone interval $s = 2^{(1/12)} \approx 1{,}06$. The reference values must be consistent across (2) and (3). Normally we choose reference values tuned (no pun intended) for MIDI note numbers, for interability with MIDI components, namely $N_{\mathrm{Ref}} = -1$, $K_{\mathrm{Ref}} = \mathtt{C}$, $f_{\mathrm{Ref}} = 440\text{Hz}$, $i_{\mathrm{Ref}} = 69$.

To derive the pitch of a given frequency we use approximation. Namely, the pitch index *i* of a given frequency *f* is the one such that *f(i)* is the closest to *f*, on the logarithmic scale.

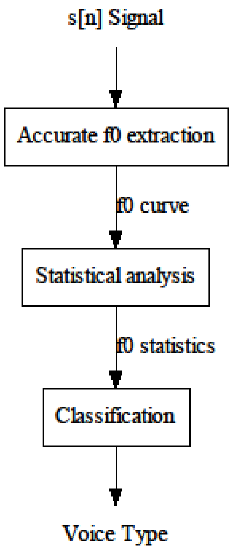


Figure 2: Overall algorithm

2.2 f_0 extraction

Our algorithm is based on an analysis of the fundamental frequencies f_0 produced by the singer. The extraction of the f_0 values of the singing voice is a non-trivial matter [4,5]. We use a combination of techniques including peak-picking for harmonic location [5], interpolation of the spectral magnitude for f_0 estimation precision [4], and harmonic spectral magnitude based weigthing [5]. See these references for detailed descriptions.

This same method of f_0 extraction has proven successful also in other applications, such as real-time audio compression [5], real-time pitch modification, and real-time pitch tracking for karaoke-like gaming, also under development at our laboratory (gnomo.fe.up.pt/~voicestudies).

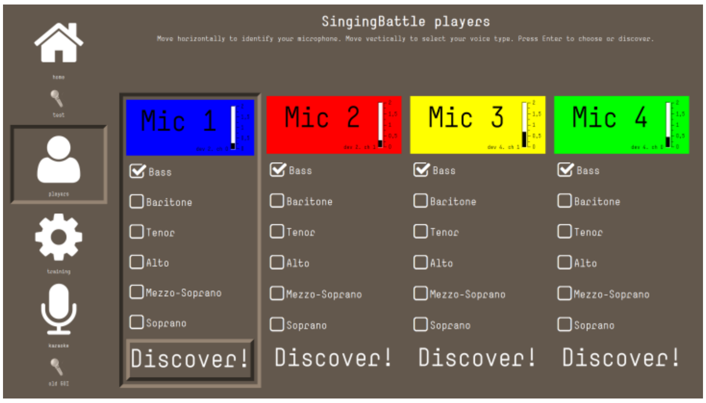
2.3 Voice type classification

For the final analysis of the extracted pitches, we assume the production has respected the score in Figure 1. With this assumption a simple approach based on the mean and median of the pitches is attainable. Basically, we match the sample's mean/median pitch with the reference's mean/median pitch derived from the ranges described in Table 2. Namely, we select the voice type which mean/median pitch is the closest to the sample's mean/median pitch (in any direction).

Voice type designation	Pitch range	
	From	To
Bass	D2	E4
Baritone	F2	G4
Tenor	A2	D5
Alto	E3	E5
Mezzo-soprano	G3	A5
Soprano	B3	C6

Table 2: Common designations and pitch ranges of the six main classical voice types, adapted from [1]. For the pitch notation see section 2.1

Figures 3-5 contain screenshots of the voice discovery function at play in the SingingBattle application under development. When a player presses the “Discover!” button on their “Mic” panel (Figure 3), the voice discovery widget shows up (Figure 4), consisting of a pitch ruler of sensible range (C2-C7). Upon performance, by the singer, of the voice discovery passage, the acquired pitches are plotted in real-time upon the ruler as white dots. Eventually the voice type classification takes place and is manifested by a label in the widget and by the selection of the corresponding checkbox on the players “Mic” control widget (Figure 5). This design is still under construction.


SingingBattle players

Move horizontally to identify your microphone. Move vertically to select your voice type. Press Enter to choose or discover.

Figure 3: Invoking the voice type discovery function.

3 Exploratory results and things to try

Systematic evaluation has not been done yet. This is an exploratory study. Extensive ad hoc testing has shown that the above method works well more often than not. But there are occasional misclassifications. A number of possible improvements is envisaged.

The current method requires the disciplined action of the singer to perform the prescribed passage. A possible improvement in the direction of a more natural, less contrived interaction with the singer would be to allow any singing to be produced. Actually the simple mean/median pitch method already used seems applicable to such unconstrained sample, because it does not look at the sequential aspect. That is, all other things being equal, it seems the prescribed passage is not required after all.

Unfortunately, it might not be the case that all other things can be maintained equal. Another possible improvement is to make the procedure more robust to intervening noise. The method as is includes any f_0 that may have its source in extraneous elements to the singer e.g. other voices than the singer, office noise, etc. Such noises are likely to produce outliers in the set of acquired f_0 values. We are currently investigating possible methods of identifying such outliers. The methods we have envisaged so far resort to the sequential aspect of the input, namely the two phase curve form. That is, for these methods, we again require the disciplined input of Figure 1.

Yet another possibility is to use image pattern recognition methods for voice type classification [6]. The main idea is to generate a visual representation, i.e. a 2D image, containing the time-pitch data obtained in the voice type discovery exercise. It is expected that this spectrogram-like image be highly correlated with the image pattern that represents each voice type. This approach can effectively reduce the outliers impact on classification, thus improving the voice type classification rate.

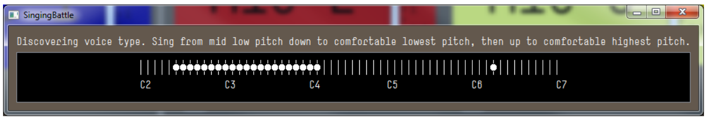

Discovering voice type. Sing from mid low pitch down to comfortable lowest pitch, then up to comfortable highest pitch.

Figure 4: Voice type discovery widget.

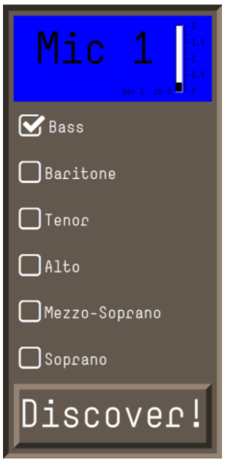


Figure 5: A player’s “Mic” control widget.

4 Conclusion

We have developed f_0 extraction algorithms [4,5] which are adequate for the recognition of the voice type of singers. The algorithms are fast enough to be used in real time applications. We are developing computer software prototypes for interactive, real-time, voice type discovery (Figures 3, 4, 5).

Avenues of future research and development include the improvement in usability and performance of the voice type discovery functionality, and further computer applications that help to learn, train, or simply enjoy singing. A number of such applications are at various stages of development (gnomo.fe.up.pt/~voicestudies).

This work was supported by the Portuguese Foundation for Science and Technology, an agency of the Portuguese Ministry for Education and Science, under research project PTDC/SAU-BEB/14995/2008.

References

[1] J. F. T. S. Ferreira. Tecnologia de Apoio em Tempo-Real ao Canto. Relação entre parâmetros perceptivos da voz cantada com fenómenos acústicos objectivos. Diss. Mestrado em Música, ESMAE, 2012. gnomo.fe.up.pt/~voicestudies, consulted 2013

[2] K. O'Connor. SingWise. An information-Based Resource For Singers By Vocal Technique Instructor, Karyn O'Connor. singwise.com, consulted 2013

[3] R. W. Young. Terminology for Logarithmic Frequency Units. J. Acoust. Soc. Am. Volume 11, Issue 1, pp. 134-139 (1939); (6 pages)

[4] R. Sousa and A. Ferreira. Non-iterative frequency estimation in the DFT magnitude domain. In: Proceedings of the 4th International Symposium on Communications, Control and Signal Processing, ISCCSP 2010, Limassol, cyprus, 3-5 March 2010

[5] A. J. S. Ferreira. Perceptual Coding of Harmonic Signal. In: Proceedings of the 100th Convention of the Audio Engineering Society, 1996 May 11-14 Copenhagen

[6] Khunarsal, P.; Lursinsap, C.; Raicharoen, T., "Singing voice recognition based on matching of spectrogram pattern," Neural Networks, 2009. IJCNN 2009. International Joint Conference on , vol., no., pp.1595,1599, 14-19 June 2009