POLITECNICO DI TORINO

Master degree course in Biomedical Engineering

Master Degree Thesis

Singing voice quality assessment in professional singers through acoustic parameters obtained with different microphones

Supervisor:
prof. Alessio Carullo

Co-Supervisors:
eng. Antonella Castellana
prof. Arianna Astolfi
prof. Alberto Vallan

Candidate:
Ilaria Leocata

Academic Year 2017 – 2018

Ai miei genitori.

Abstract

The quality of singing voice is an important feature for a good singer. If the singer's voice quality is high, the success will be great for singers. The voice signal can be analysed both in time and frequency domain and different parameters and features can be extracted to obtain suitable information for clinical reasons (for example to find vocal disorders) and also for a quantitative evaluation of voice quality.
The main goal of this work consists in analysing vocal parameters that are able to evaluate the singers' voice quality. 39 volunteer singers participated to this study: 14 of them were professional and semi-professional singers and the other 25 were untrained singers. All of them performed different tasks in the anechoic chamber of Politecnico di Torino and filled in an informative questionnaire. The vocal tasks were recorded by means of three different devices: a sound level meter (SLM) and two contact microphones, namely a Piezoelectric microphone (PIEZO) and a Electret Contact Microphone (ECM), that sense the vibration induced by vocal folds activity in two different neck positions. After an initial vocal warm-up, the participants performed different tasks: an arpeggio sung in two conditions different in tempo (slow and fast) and articulation (legato and staccato), two songs with two different extensions (the first of one octave and the second of one octave and half) and reading a phonetic-balanced text.
Different aspects of voice quality were investigated. First, the pitch accuracy of singers with the three microphones was evaluated during arpeggi. Then, the Singer's Formant was evaluated through the *Long Time Average Spectrum* (LTAS) of the singing tracks. In addition, the voice quality of trained and untrained singers, was compared by means of the *Singing Power Ratio* (SPR), which allows the presence of the Singer's Formant to be highlighted. Eventually, a cepstral parameter widely used to discriminate the vocal health status of subjects while speaking, namely the *Cepstral Peak Prominence Smoothed* (CPPS) was evaluated in both reading and singing tasks. In particular, CPPS distributions obtained for each subject in the two tasks and CPPS distributions in trained and untrained singers were compared. Regarding pitch accuracy, the results show that the three microphones are comparable: the overall mean of the differences of the fundamental frequencies estimated with the contact microphones and sound level meter is equal to (1.9 ± 0.4) Hz which

is not discernible by the human ear. This outcome indicates the possibility to replace the sound level meter with contact microphone, which is cheaper and allows singers to be monitored in noisy environments. As the Singer's Formant analysis and SPR are concerned, the obtained results show significant differences among the three microphones as expected, since signals from contact sensor are not affected by the vocal tract filtering. The best discrimination from trained and untrained singers was obtained for ECM results; the worst results were obtained from PIEZO because this microphone presents a middle-frequency boost in the spectrum that affects the obtained values. Moreover, it was found that the energy into the spectrum does not depend on the performed singing track. Another interesting result was obtained during the comparison between the value of SPR in the two categories of singers. The biggest difference was found for SPR obtained from ECM between trained and untrained singers: this difference is greater than 5 dB. Such an outcome could be related to a better capability of ECM to detect a proper preparation of the vocal apparatus to sing than the sound level meter. The results related to CPPS show different behaviours in the analysis of reading and singing tasks and between trained and untrained singers. While the shape of CPPS distributions looks like a normal distribution for the singing tasks, it becomes bimodal for reading task. This could be related to a better control of vocal folds during singing activity. For this reason, the CPPS distributions of untrained singers are not clearly different between reading and singing tasks. Interesting results were obtained in this work, even though further investigations are needed in order to better understand these preliminary outcomes. The future developments could be: increase the dataset, characterize the contact microphones, analyse songs with a larger frequency extension and identify CPPS-related parameters that could be suitable to distinguish between trained and untrained singers.

Contents

1 Introduction **1**

1.1 Vocal Apparatus . . . 2
1.2 Artistic Voice . . . 2
1.3 State of the art . . . 6

2 Materials and Methods **8**

2.1 Subjects . . . 9
2.2 Acquisition Method and Microphones . . . 12
2.2.1 Contact Microphones . . . 12
2.2.2 Microphone in air . . . 14
2.3 Tasks . . . 14
2.4 Preprocessing . . . 17
2.5 Method . . . 18
2.5.1 Evaluation of Pitch Inaccuracy . . . 18
2.5.2 Cepstral Peak Prominence Smoothed . . . 19
2.5.3 Long Time Average Spectrum . . . 20
2.5.4 Singing Power Ratio . . . 21

3 Results and Discussions **23**

3.1 Evaluation of Pitch Inaccuracy . . . 23
3.2 Cepstral Peak Prominence Smoothed . . . 50
3.3 Long Time Average Spectrum . . . 65
3.4 Singing Power Ratio . . . 94

4 Conclusions **100**

A Insight of acoustic parameters **102**

Bibliography **110**

Chapter 1

Introduction

The voice is one of the most important means to communicate with other people to understand each other. Each voice is unique for all of the people: in fact each person differentiated from the others from a value called *Fundamental Frequency*. This value indicates the personal vibration of vocal folds during normal speech. Some jobs required the costant use of voice: teachers, singers, actors. The extended use of voice during the years can cause some problems to the vocal folds. For these categories, it is important control usually the status of their vocal folds and do accurate analysis. For example, singers do periodically medical visites from a specialist called *Phoniatrist*: he does a quantitative analysis of voice and does the laringostroboscopy to control the healtness of vocal folds. To evaluate the quality and the healtness of the voice, some parameters obtained through the analysis in the time and frequency domain are studied and observed starting from the vocal signal. In this thesis' work, many parameters (health and qualitative) are calculated and evaluated. These parameters are considered for three type of microphones; this type of analysis is not invasive. Through the use of these type of microphones, it wants to try to give a numerical value at different parameters to classificate the goodness of voice. This is the greatest difficulty meets in this work: give a objective value to a parameter to say if a voice has high quality or not it is very hard. This thesis work has the aim to give an objective method to evaluate the quality of singers' voice. Previous studies show some evidences of a connection between the goodness of singing voice and the numerical value of some parameters. For example, the assessment of *Singing Power Ratio*: it consists into the valuation of the height of the peaks into the spectrum of the singers' voice. Other studies take information about the analysis of the *Singers' Formant*. The singer's formant is an high peak in the spectrum energy near 3 kHz;

this is the features of the professional singers like basses, baritones, tenors and altos. This is not a soprano's features instead; the behaviour of the signal spectrum is different than the others. One of the aim of this work is to give a better definition of the Singers' Formant also for sopranos and not only to male singers. Another aim of this thesis, is to evaluate an health paramter called *CPPS* in a different way from the one existed in literature [1]. The starting point of this study is to evaluate the quality of singers' voice analyzing some parameters existing in literature with different microphones. Some comparisons between trained and untrained singers were done. The following sections explain a general view about artistic voice and the state of art about this topic. Then, will be explain the main caratheristic of the used microphones and the tasks required to the singers. Then, the main steps of the method are explained in detail and, at the end, there are the results and future conclusions. More detailed information about the observed parameters are described into the appendix *Insight of acoustic parameters.*

1.1 Vocal Apparatus

The definition of vocal apparatus is inappropriate because it is not a single biological structure but it includes several parts of different apparatus. All of these parts are indispensable to produce voice and sounds. The first element of this biological system are lungs: the air that go out of lungs goes to trachea and larynx. Here, the air meets an obstacle, the vocal folds determining an increase in subglottic pressure. The vocal folds are formed by internal beams of tiroarotenoid muscles and inferior tiroarotenoid ligaments; they are also lined with laryngeal mucosa. The space delimited by vocal folds is called glottis; this one dilates and shrinks during vocal emission like consequence of vibrations of vocal folds. The vibrations of vocal folds create an air current that passes through the pharynx and then reaches the upper cavities (mouth and nose). When air current reaches mouth, and also teeth, tongue, lips, soft and hard palate intervene to produce vowels and consonant [2, 3].

1.2 Artistic Voice

The voice source, the raw material of all voice sounds, is a complex tone composed of a number of harmonic partials: the frequency of first partial is called fundamental frequency (F0). This F0 corresponds to the vibration frequency of the vocal folds.

Every note has an own frequency; if the fundamental frequency sounded alone, people perceived always the same sound. The spectral information varies during the movement of air current from the vocal folds to the lips (called *Vocal tract*). The reason of this is that the ability of the vocal tract to transfer the sound dependent on the frequency of the sound being transferred. The resonances of the vocal tract are called *Formants*. These formants depend on the position of the differents parts of mouth (lips, tongue, ecc) and not just the sound; they manifest with a peak in the spectrum in a precise frequency range. This peak is different for each person. For example, the formants are higher in children and women than in adult men; this is because the shape of vocal tract is very different between these three types of people; futhermore the formants are different if it is analyze the power spectrum of normal speech or singing. So, the important aspects of vowel sounds are explained following: the amplitude of the foundamental which is dependent upon glottal adduction and the amplitude of the overtones which dependent by subglottal pressure.

These features have different charateristics between trained and untrained people. For singers, one of the first study of vocal signal was conducted in the 1934 from W. T. Bartholomew [4]. Here was studied some charateristic of the quality of voice of professional singers: vibrato, intensity of the pitches and low and high formants. Singers, usually, sing at an F0 higher than F0 using in normal speech. The F0 in normal speech for women is around 200 Hz and for men is 110 Hz. For example, during singing for sopranos, the frequencies can reach 1050 Hz and for tenors reach 520 Hz. During singing, the strategies used to be audible during a performance are different. The concept of formant frequency is totally different for male and female opera singers; this concept is closely linked to audibility of singer during an exhibition. For male singers the pitch-dependent choice of the two lowest formant frequencies is not to be expected except in vowels with a low formant frequency sung at high pitch. This seems has the consequence of a production a speech spectra similar to a orchestra spectra; this behaviour is different instead for the singing spectra. The great audibility of the mens voice during a concert is permitted because their voice is higher when the accompaniment is loud. The singing voice contains more energy in the frequency region of 2.5/3 kHz: infact the spectra shows an high peak. This region corresponds to the falling frequency region of the orchestra: there are not too much frequency infomations in the high region so the mens voice can stand above the sound of the orchestra (the highest orchestra's frequency reached is around 500 Hz). This high peak is called *SINGERS FORMANT* and it is a feature of all of the type of mens singers (tenors, baritons, basses). The difference between this type of singers is the pitch range avaiable into the Figure 1.1. Also the formant frequencies are different in these types of singers: the reason of this difference is linked to the different shape and length of vocal tract of singers. On the other hand, the lie is different for the women. They also have different type of singers who have different pitch range shown in Figure 1.1.

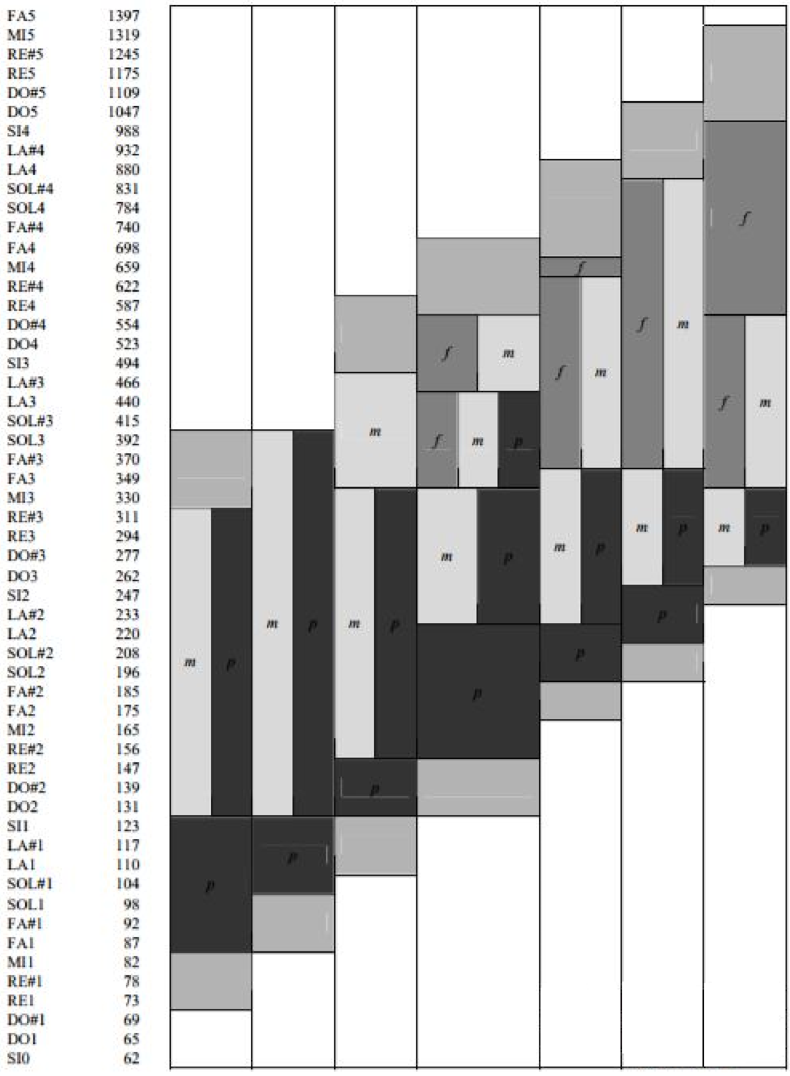


Figure 1.1. *Extension of Singers. In order (from left to right): Basses, Baritones, Tenors, Contralto, Altos, Mezzo Sopranos, Sopranos [5].*

Here is available the table of color codify.

p	Chest Register
f	Head Register
	Execeptional Cases
m	Medium Register

Table 1.1. *Table of colors.*

During normal speech, for sopranos, the F0 is far above from the frequency corespond to the first formant (F1); they avoid this situation during singing and move F1 to F0. To obtain this result, singers reduce the maximum constriction of the vocal tract and then to extend the jaw. During this process, the amplitude of the fundamental increases considerably. The position and the amplitude of the other formants do not change very much.

The consequence of this behaviour is an increasing of the sound pressure level (SPL); this phenomenon occurs as frequency increases. The SPL of an orchestra (playing loudly) is around 100 dB: the singers SPL must be higher to allow to the singer to be heard by the audience also if it is strange that SPL can reach this value for a person. As opposed to men, the women do not have a precise peak into the range of 2.5/3 kHz but the energy of the spectra is distributed into a larger range that reach the frequency of 4/5 kHz.

Another important skill is the *Register*. It is an aspect of phonation with no precious definition. It is referred to the ability of singer to produce different sounds in a similar way and to produce sound equal in timbre. Different registers have different manner of vibrations of vocal folds. For men exist three types of register: vocal fry, modal and falsetto. For women instead there are four types of register: chest, middle, head and whistle. The Figure 1.1 contain some different type of registers.

For professional singers is also important the concept of *Vibrato*. It is manifests automatically during voice training. It is an oscillation of F0. It is important to know the rate of vibrato (number of oscillations per second) and the extent (the depth of modulation expressed in cents of semitone). Usually, the rate of vibrato is around 6 Hz; it depends by singers and their repertoires.

At the end, for professional singers, is also important the aspect of *Expressivity*. This concept represents the ability of the subject to transmit emotions and feelings to public [6].

1.3 State of the art

This section summarizes what is current state of art of the voice parameters and the study of quality of voice of professional singers during the years.

There are some parameters used to analyze the vocal signal: *Jitter* measures the variation of the frequency of the signal in time and *Shimmer* is relative to the amplitude of the signal. The high variability of these parameters do not allow to conduct precious analysis. Other parameters are: *Harmonics to Noise Ratio* (HNR) is an assessment of the ratio between periodic components and not periodic components of a segment of voiced speech, *Normalized Noise Energy* (NNE) is the measurement of noise present in voice respect to the total energy and *Glottal to Noise Excitation Ratio* (GNE) gives information about the origin of the voice signal that is from the vibration of vocal folds or turbolent noise generated in the vocal tract.

The first study about voice of professional singers was carried out from Bartholemew [4], he spoke about the concept of "good voice" in male singers. He found a peak near to 3 kHz. This peak, was called by Sundberg in 1974 *Singer's formant* [7]; it is different from each singer and its level depends on vocal intensity, fundamental frequency and voice classification. During the years, lots of studies were carried out to analyze the spectral information of singers' voice. In 1995 Omori et al. [8] studied for the first time a new parameter to valuate the quality of voice in professional singers analyzing the peaks into the spectrum of sung /a/ and the spoken /a/: it called *Singing Power Ratio* (SPR). This parameter consists into the difference between the peak found into the range 2/4 kHz and the peak into the range 0/2 kHz into the spectrum of signal. This stategy is much used into the studies of voice; there are studies about children's voice [9] and studies to evaluate the quality of voice in trained singer and untrained singers [10]. The last study in lecterature of this parameter is the study conducted by McHenry et al. in 2017 [11]. Another important method used for voice analysis is *Long-time Average Spectrum* (LTAS): is a frequency analysis that allows to analyze the frequency information of a signal. It is used for analysis of both speech and singing. It allows to analyze the important components of voice: low frequency for speech and high frequencies for singing. It allows to found the frequency differences between the voice of women and men. One of the first study of vocal disorders was published in 1977 using LTAS [12]. During the years, lots of studies were published about the analysis of simple speech [13], singing choirs [14], singing vowels and commercial recording like Sundberg [15]. The

LTAS is useful also to compare the spectrum of professional singers and untrained normal speakers [16]. One of the more recent study using LTAS was proposed by Johnson et al. [17] to obtain a good voice classification in male singers. There is a better parameter that is used more frequently to study the vocal disorder instead of LTAS: the *Cepstrum*. The first use of this parameter was in 1963 from Bogert et al. [18]. The cepstrum is a log power spectrum of a log power spectrum how said Hillenbrand et al. [19]. From this parameter born the *CPPS* [1], the most frequently parameter used today: the *Cepstral Peak Prominence Smoothing* parameter is used to discriminate healty from unhealty people analyzing the shape of its distribution. This parameter is useful at low frequency (50/300 Hz) because this is the range of speech and this parameter gives more detailed information of good health of voice. Today does not exist a similar parameter for the analysis of good health in singing instead of speech: CPPS infact does not give good results into the analysis of singing at high frequencies. Recently are also found some threshold values for specific microphone useful to classificate healty from unhealty people [20]. These types of analysis were accompanied during years by the new technologies of medical imaging. X-Ray analysis were used in 1985 to analyzed the movement of the vocal tract and the vocal folds to correlate the sound and the movement [21]. More recent studies used magnetic resonance instead X-Ray for the analysis of movement of vocal tract [22]. These studies are relatively recent: there are not much studies and parameters about the use of contact microphone in lecterature. The advantages of using this contact microphones is that they are not sensible to background noise, they are not very expensive and they are simple to use; they are also not invasive and confortable. Anyway, it is possible to catch a good signal to analyze.

Chapter 2

Materials and Methods

This chapter describes how the experiment has been carried out. Here is described the protocol from acquisition to the elaboration. In the first time are described the subjects and their features, then the microphones and their frequency behaviour.
Before starting the experiments, all of the subjects completed a questionnaire with the basic infomation summarized into the Table 2.1 and an audiometric test using the app "uHear" avaiable for free on Apple Store.
The tasks required to the singers are also indicated. In the end there is the most important section, the elaboration of the signal to obtain important parameters and values about the singers' voice. All of the signals were preprocessed before starting the main part of analysis.

2.1 Subjects

Fourteen volunteers singers (professional and semi-professional) took part into the experiments. The features of the singers are summarized in Table 2.1[1].

Gender	**Group**	**Voice Type**	**Age**	**Repertoire**	**Years of Exp.**	**Diseas**
Female	P	Alto	50	A, C	29	No
Female	P	Soprano	47	A	20	No
Male	P	Tenor	36	A, E	10	No
Male	P	Bass	51	A, B, C	29	No
Female	P	Soprano	56	B	25	No
Female	P	M-Soprano	37	A, B	4	No
Female	P	Soprano	48	A,B	25	No
Female	P	Soprano	26	C, D, H	3	No
Male	P	Baritone	32	A	4	No
Female	S-P	Soprano	25	B	2	No
Female	S-P	M-Soprano	21	B	-	No
Female	S-P	Soprano	26	B, G, H, I, E	4	No
Female	S-P	Soprano	24	B	2	Yes
Female	S-P	Soprano	24	C, E	3	Yes

Table 2.1. *Features of the Professional Singers.*

It is available the Table 2.2 conteinig the repertoires' coding.
The Table 2.3[2] contains the features of the not professional singers. All of the untrained singers have, al least, five years of choral sing.

[1]Into the column "Group", P is *Professionist* and S-P is *Semi-Professionist*. Into the column "Repertoire", the letters are codified into the Table 2.2.

[2]Into the column "Group", U is *Untrained*.

Repertoire	**Coding**
Opera	A
Lyric	B
Baroque	C
Contemporary Music	D
Belcanto	E
Reinassance Music	F
Musical	G
Pop	H
Jazz	I
Soul	L

Table 2.2. *Coding of Repertoires.*

Gender	Group	Voice Type	Age	Diseas
Female	U	Alto	28	No
Female	U	Alto	23	No
Female	U	Alto	24	No
Female	U	Alto	28	No
Female	U	Soprano	23	No
Female	U	Soprano	21	No
Female	U	Soprano	34	No
Female	U	Soprano	22	No
Female	U	Soprano	21	No
Female	U	Soprano	24	No
Male	U	Tenor	19	No
Male	U	Bass	28	No
Male	U	Baritone	22	No
Male	U	Baritone	23	No
Male	U	Bass	26	No
Male	U	Tenor	50	No
Male	U	Tenor	20	No
Male	U	Tenor	23	No
Male	U	Bass	22	No
Male	U	Baritone	19	No
Male	U	Bass	22	No
Male	U	Bass	21	No
Male	U	Bass	25	No
Male	U	Bass	24	No
Male	U	Bass	26	No

Table 2.3. *Features of the Untrained Singers.*

2.2 Acquisition Method and Microphones

The experiment has been made in the anechoic chamber of Polythecnic of Turin. Three microphones were used: two contact microphones and one air microphone.

Figure 2.1. *Anechoic Chamber.*

2.2.1 Contact Microphones

The first microphone is an *Electret Condenser Microphone* (ECM AE38, *Alan Electronics* GmbH, Dreieich, Germany). The position of this microphone is on the jugular notch, where it is possible to feel a good vibration of the vocal folds through the vibration of the external skin [23]. This microphone was connected to the handy recorder ROLAND R05 (*Roland Corp.*, Milano, Italy). The sample frequency is 44100 Hz and resolution is 16 bits. The ouput signal was saved into a dedicated SD card.

The second microphone is *Piezoelectric Contact Microphone* (HX-505-1-1, *HXKK*, 406, Plant 1, Jiading Science Park, Dalang, Longhua New Dist., Shenzhen, Guangdong, China). It is a neck-ring contact microphone that senses the vocal folds vibrations and translates their movements into an electrical signal. The sample frequency is 22050 Hz and resolution is 16 bits. This microphone is connected through an AUX cable to a smartphone (Samsung SM-G310HN). The recordings were made through the "Vocal Holter" app. The output signal was saved into a dedicated folder into the internal memory of the smartphone.

The position of the microphones is shown in the Figure 2.2.

Figure 2.2. *Contact microphones.*

2.2.2 Microphone in air

The microphone in air is the calibrated *Sound Level Meter* (SLM, XL2, NTi Audio, *Schaan*, Liechtenstein), with a class 1 omnidirectional measurement microphone M2210 by NTi Audio. It is a microphone that requires that the subject remains at a fixed distance during the test. For the entire period of the test, each subjects was asked to stand in front of the microphone, on axis, at the fixed distance of 30 cm as provided by a thin spacer. The recommended mouth-to-microphone distance for this kind of measurements is 30 cm and with this suggested distance, when the background noise level is lower than 25 dBA, the low-intensity voice levels can be obtained with a signal-to-noise ratio (SNR) of at least 10 dB [24].
The position of this microphone is shown in the Figure 2.3.

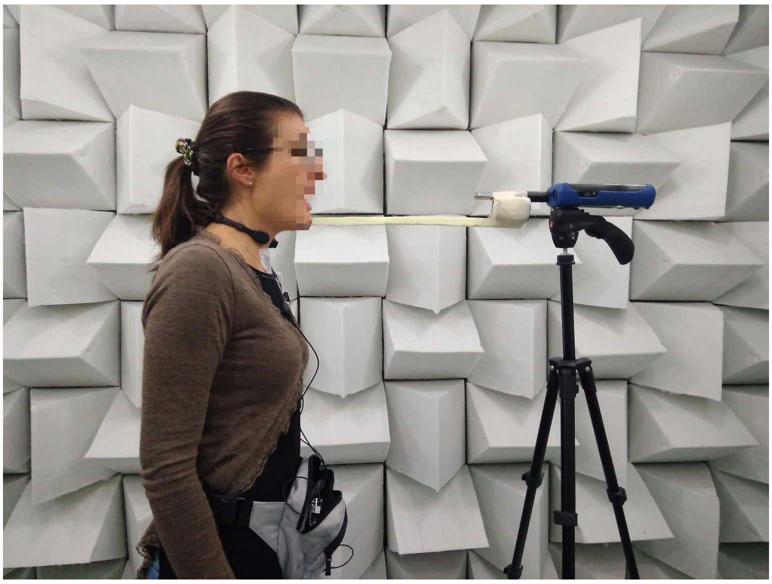


Figure 2.3. *SLM at 30 cm distance from mouth.*

2.3 Tasks

To conduct a study about the quality of voice, were asked to the first group of singers to speak and sing some tasks to obtain the values of different parameters.

The first required task was to read the phonetic-balanced text shown below.

> *Il papà (o il babbo come dice il piccolo Dado) era sul letto. Sotto di lui, accanto al lago, sedeva Gigi detto Ciccio, cocco della mamma e della nonna. Vicino ad un sasso cera una rosa rosso vivo e, lo sciocco, vedendola, la volle per la zia. La zia Lulù cercava zanzare per il suo ramarro, ma dato che era giugno (o luglio non so bene) non ne trovava. Trovò invece una rana, che, saltando dalla strada, finì nel lago con un grande spruzzo. Sai che fifa, la zia! Lo schizzo bagnò il suo completo rosa che divenne giallo come un taxì. Passava di lì un signore cosmopolita di nome Sardanapalo Nabucodonosor che si innamorò della zia e la portò con sé in Afghanistan.*

This task was useful to obtain the value of CPPS for each singers; this parameter was calculated to verify that the subjects were in good healt.

It was asked also to sing two glissatos (ascendent and discendent) to know what is the vocal extension of each singer.

The second task was to do an arpeggio using the vowel /a/ in a selected tonality done in four different ways:

- legato and slow;
- legato and fast;
- staccato and slow;
- staccato and fast;

The A+ key was chosen for Altos, Baritones and Basses; the C+ key for Tenors, Mezzo-Sopranos and Sopranos. This task was required to quantify the accuracy of every subjects to do a precious pitch in a different ways as said in a previous paper [25]. Each single pitch was compared with its reference frequence. Before each arpeggio, singers listened with headphones the corresponding tonal scale played on a virtual keyboard using a tablet. The app "Perfect Piano" of virtual keyboard was avaiable for free on Google PlayStore.

To find and analyze the singer's formant, was used the method of LTAS.
Previous studies analyzed the spectrum of different vowels and commercial recordings [15]; here were analyzed the LTAS of two songs that differ for their extension.

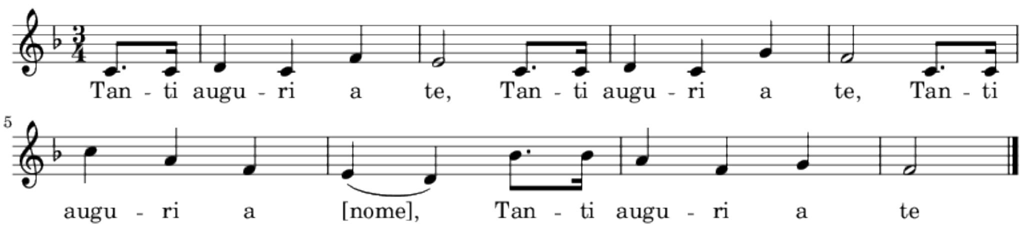


Figure 2.4. *Example of musical score of "Tanti auguri".*

The first was "Tanti Auguri" song (with an extension of one octave); the Figure 2.4 show an example of a musical score.
The second task is a strophe of "Inno di Mameli" song (with and extension of one octave and half). It is available an example score in Figure 2.5.

Figure 2.5. *Example of musical score of "Inno di Mameli".*

Both of songs was sung in the italian language. Every singers choose by theirself the most confortable key for this task.

To one group of not-professional singers were asked only to sing "Tanti Auguri" song and a strophe of "Inno di Mameli" song in the same way used for trained singers. At the second group, were asked to do the arpeggio in the legato and slow mode.

2.4 Preprocessing

All of the tasks are recorded in a single recording. The first step was to resampling at the frequency of 22050 Hz the signals of the ECM microphone and the SLM microphone and cut the original file in three signals using the software *Adobe Audition CC 2015*.

The ECM microphone acquires at sample frequency of 44100 Hz and the SLM microphone at frequency of 48000 Hz; the signal of piezoelectric microphone was acquired directly at the sample frequency of 22050 Hz.

Original signals were cat into two parts: one containing the phonetic-balanced text and the other containing the singing tasks. For every singing tasks, the time of start and finish was written.

For the first part of the analysis, the "evaluation of pitch inaccuracy" of singer, the signal was resample at a frequency of 352800 Hz to increase the resolution in frequency before process the signal.

The second part, before starting the analysis of CPPS, the reading text was elaborated to delete the silences. In according with A. Castellana [20] and J. Hillenbrand [1], was used a window of 1024 sample (coresponding at a temporal window of 46 ms) to delete silences into these texts. This value of the window was chose because is confrontable with the length of intersillabic pause. The obtaneid texts were elaborated to obtain the distribution of CPPS.

In a second time, to remove silences into the songs before using them for LTAS analysis, was used a window of 256 samples instead of 1024. Here there is no limitations about the length of the window: this window was chose to be more precious to delete the silences into the recording. The resulted song was used to obtain the spectrum using LTAS.

In both of elaborations, the resampling and the removing silences were made with Matlab R2017a.

2.5 Method

This section contains the most important themes and the protocol of experiment conducted into this thesis. The first analysis was to evaluate the pitch inaccuracy of singers; the second one was to analyze the distribution of CPPS; the third one was about the frequency content of signals analyzable with LTAS and the last one analyzed the value of SPR.

2.5.1 Evaluation of Pitch Inaccuracy

Here was valuated the pitch inaccuracy of singers during the arpeggio in four ways like Bottalico et al. [25]. For each arpeggio done by singers, the F0 of each pitch was calculated and this result was compared with the teoretical F0. First of all, the right piece of signal was selected to analyze (approximately one second per pitch) and the autocorrelation of signal was effectuated on Matlab R2017a. First the analysis was limitated in frequency between 50-550 Hz for male singers and between 50-1000 Hz for female singers. Then, the autocorrelation was computed from the first to the last sample of the interested pitch. Then the maximum value of the vector of autocorrelation was found and the F0 for each pitch was obtained. To obtain the accuracy for all of arpeggios, the operation of mean was carried out. The same operation was used to achieve the accuracy of the ascending and descending semi-phrases. This procedure was used for the arpeggios for the three microphones. Then there were evaluated the results obtained from the microphone in air compared to the fundamental frequency for each pitch of the arpeggio. It was calculated the difference between the output results of microphone in air and the value of the fundamental frequency for each note. After, it was calculated the difference between the values of the same note during the ascending and descending semi-phrase of the arpeggio. Then it was calculated a new parameter, called H with the following formula:

$$H = \frac{\sum_{k=1}^{N} |a_i|}{N} \tag{2.1}$$

Here, N is the number of the considered pitches and $|a_i|$ are the absolute value of the previous differences. If this value is high, it will mean that the singer is influenzated from the type of semi-phrase (so ascending or descending). At the end, it was evaluted the behaviour of the contact microphones compared to the microphone in air. The microphone in air was considered as reference microphone. The obtained results were evaluated for contact microphone in term of resolution.

2.5.2 Cepstral Peak Prominence Smoothed

CPPS is a parameter that discriminates healthy from unhealty voices and it shows the peak related to the different frequency components of the interested signal. The aim of this part of thesis, is to calculate the CPPS for each subject of a phonetic-balanced text and the CPPS of two singing texts. After the step of preprocessing explained into the Section 2.4, this texts were processed with the Matlab algorithm to obtain the distribution of CPPS.
For the reading text, it was used a window of observation of 1024 sample (corresponding of a time window of 46 ms) with an overlap window of 44 sample (coresponding of a time window of 2 ms).
The first step was done the FFT of the first window of the signal; then, this window is moltiplicated for an hanning window of the same length. The step after consisted to obtain the spectrum of the signal with the Formula 2.2:

$$S = 20 \log |C| \tag{2.2}$$

Then, the second logarithm was computed to obtain the Cepstrum in the domain of Quefrency. The Formula 2.3 explain this concept:

$$CPP = 20 \log |S| \tag{2.3}$$

The spectrum and the cepstrum, are both normalized to their maximum value.
This process was repeated for each window of the signal with that overlap.
Then, the process of smoothing was computed, once in the time domain and once in the quefrency domain. After this process, an histogram contained all of the peaks of each frame was obtained.

The worked frequency range is from 60 to 300 Hz (that is the frequency range in which it is possibile to found the F0 of speech of a person). The same process was repeated in a different way. The new frequency range was considered between 50 and 600 Hz. Then, the distribution of CPPS for the speech text and the singing tracks was obtained for each subject and each microphone. Then a comparison between the results was made.

2.5.3 Long Time Average Spectrum

The Long-Time Average Spectrum (LTAS) is useful to have information about the spectral distribution of a signal in a specific period of time. LTAS consists of an average of all the spectra obtained from the analysis of all the windows in which the signal was divided as shown in Formula 2.4:

$$LTAS = \frac{1}{N} \cdot 20 \log_{10} |FFT(a) \cdot hamming(w)| \tag{2.4}$$

where a is the observation signal, *hamming(w)* is the window of hamming of the same length of a and N is the number of elements of the output FFT. The observation window was of 1024 samples and there was no overlap between two consecutive windows into the analysis. The window of 1024 samples was chose to have a better frequency resolution into the spectrum (about 20 Hz of frequency resolution). For each frame was obtained the spectrum and each spectrum was normalized from his maximum value. At the end, the obtained overall spectrum was averaged. In Figure 2.6 there is an example of the LTAS of a song.

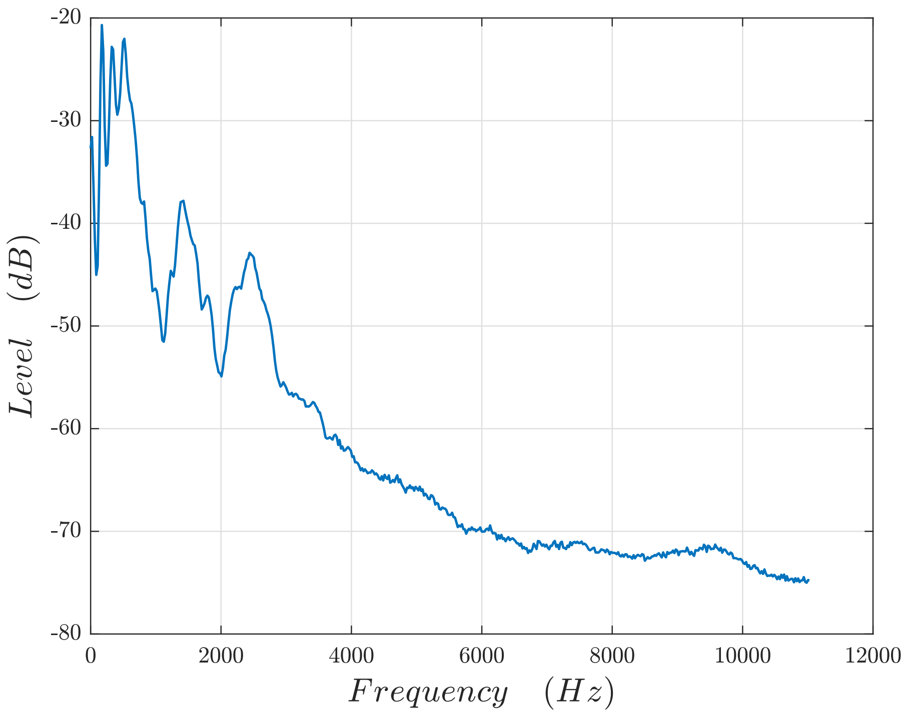


Figure 2.6. *Example of LTAS of a song.*

Here, LTAS was used to have information of the frequency behaviour of the analyzed

songs. There were analyzed the two songs explained into the Section 2.3 after a step of preprocessing.
From the LTAS of the songs, different parameters were obtained. First, the maximum value of the LTAS and its corrispondent frequency were found. Then the maximum value (the maximum peak) into the range of 2000-4000 Hz and its corrispondent frequency were searched. Then it was found the difference between these two peaks in dB.
Then the energy in three different frequency ranges was calculated and several differences were obtained. The energy in dB in the frequency range between 0 Hz and 1000 Hz (B_0), 1000 Hz and 4000 Hz (B_1) and 4000 Hz and 9000 Hz (B_2) were found. The calculated differences are shown in the Formula 2.5 and 2.6:

$$\Delta B_{low} = B_1 - B_0 \qquad (2.5) \qquad\qquad \Delta B_{high} = B_2 - B_0 \qquad (2.6)$$

The energy in dB in each range was calculated using this Formula 2.7

$$B = 10 \cdot \log E \tag{2.7}$$

where E was obtained with the following Formula 2.8

$$E = \sum_{i=1}^{n} 10^{\frac{LTAS(n)}{10}} \tag{2.8}$$

where n is the length of each obtained vector related to LTAS results. These equations were used for the three considered ranges; then B_0, B_1 and B_2 were obtained.

2.5.4 Singing Power Ratio

Singing Power Ratio is a parameter to valuate the quality of the singers' voice. It consists into the difference between the two maxima peaks into two definite range in the frequency spectrum: the first peak falls into the frequency range of 0-2000 Hz and the second one into the 2000-4000 Hz; the difference is obtained between the second and the first one.
Each signal of 96 ms was analyzed; it was made the FFT of the signal with an hamming window of the same length of the signal. Then the maximum peaks into the two ranges were found and it was calculated that difference. This procedure was follow for each pitch of the arpeggio.
This task is useful to compare the quality of voice of professional singers and untrained singers. Here the arpeggio for the two categories of singers was analyzed and

the value of SPR of each pitch between trained and untrained singer was compared. This analysis was carried out for the results of microphone in air (SLM) and for the contact microphone (ECM).

Chapter 3

Results and Discussions

3.1 Evaluation of Pitch Inaccuracy

In this section are shown the results about the pitch accuracy of singers obtained with different microphones. The first example represents the result of the pitch accuracy of the tenor during the arpeggio in a confortable key. In the figure 3.2 the result for the legato and slow arpeggio in the Do+ key is shown. The center of the blu bars indicates the reference note (for example C) and the higher half bar represents the upper semi-tone (C#) and the lower one represent (C♭). If the result goes out of this range, the singer will be considered out of tone. In this case, the singer is in tone. if the result is above or belove the reference note, the singers will be considered respectively "crescente" or "calante". In the most of results, singers tend to be *calante* or *crescente* in the descending semi-phrase of the arpeggio.
The results about the output of microphones are very good. It is possible to appreciate the high correlation between microphones into the Figure 3.1.

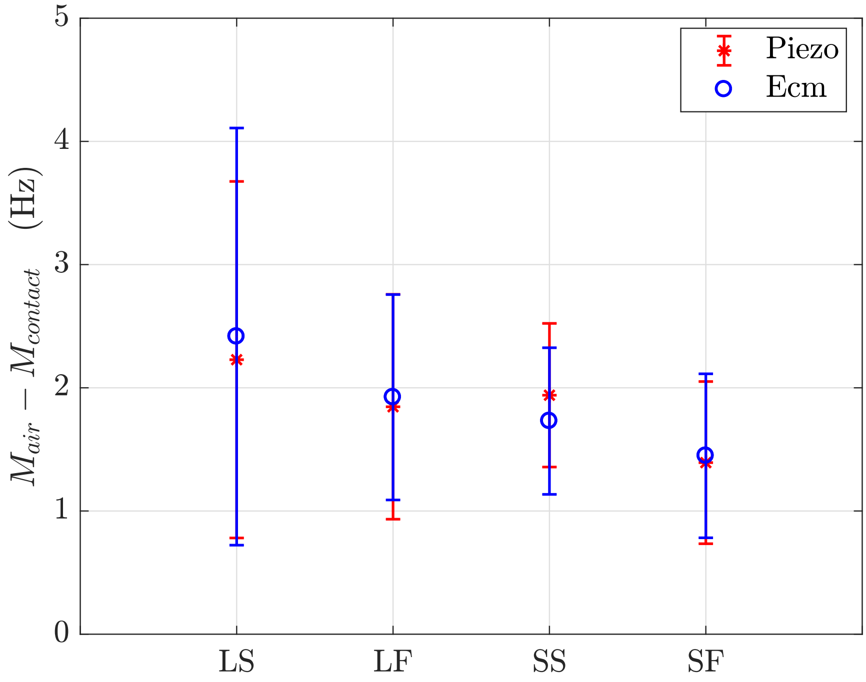


Figure 3.1. *Mean and Standard Deviation of four types of arpeggios.*

In this figure the mean and the standard deviation of the differences between the output results of frequencies of ECM and SLM and then the PIEZO and SLM of all of the singers are obtained.
The overall averages for the two microphones in air are resume in the Table 3.1:

Mean for ECM (Hz)	Mean for PIEZO (Hz)
1.9 ± 0.4	1.9 ± 0.4

Table 3.1. *Overall mean for contact microphones.*

It is possible to appreciate these correlations just looking the following Figure 3.2.

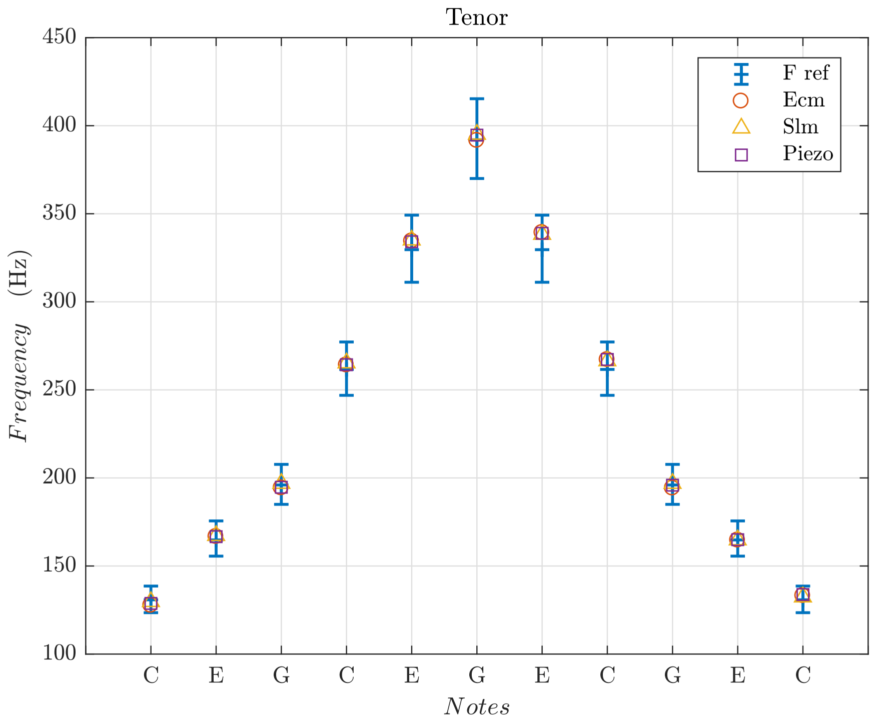


Figure 3.2. *Pitch Accuracy - Legato and slow.*

The results of all the singers are avaiable. There are not all of the results because in one case the recording was lost and in the two other cases the function of auto-correlation gave not complying results. This is probably due to the high variability of the original signal.

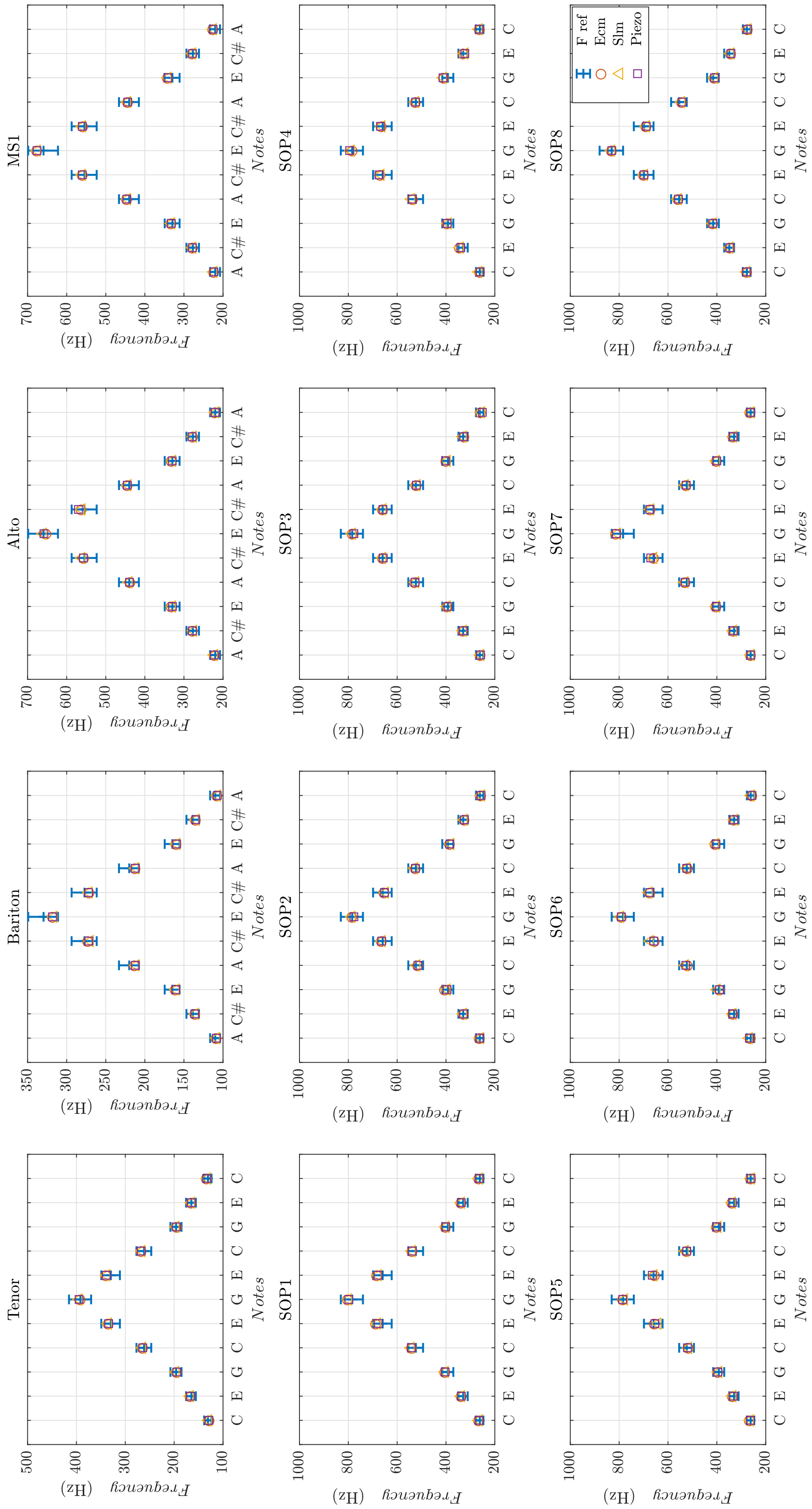


Figure 3.3. *Pitch Accuracy - Legato and slow.*

In this Figure 3.4, is shown the behaviour of a parameter called "Hysteresis of the singer". First of all, it was calculated the difference between the output result of the microphone in air (SLM) and the reference note. Then this result is shown here. The result about the ascending and descending arpeggio are differentiated. Even more the output result of the SLM is closed to the reference note, even more that difference is smaller. When the results between the ascendent and discendent parts are entirely in line, it means that the output result in these two cases are equal. Obviously, the output of the last note is equal because the last note is the only note that is comprised both of ascending and descending semi-phrase of the arpeggio. In the Figure 3.4 is shown the behaviour of the output results of Tenor.

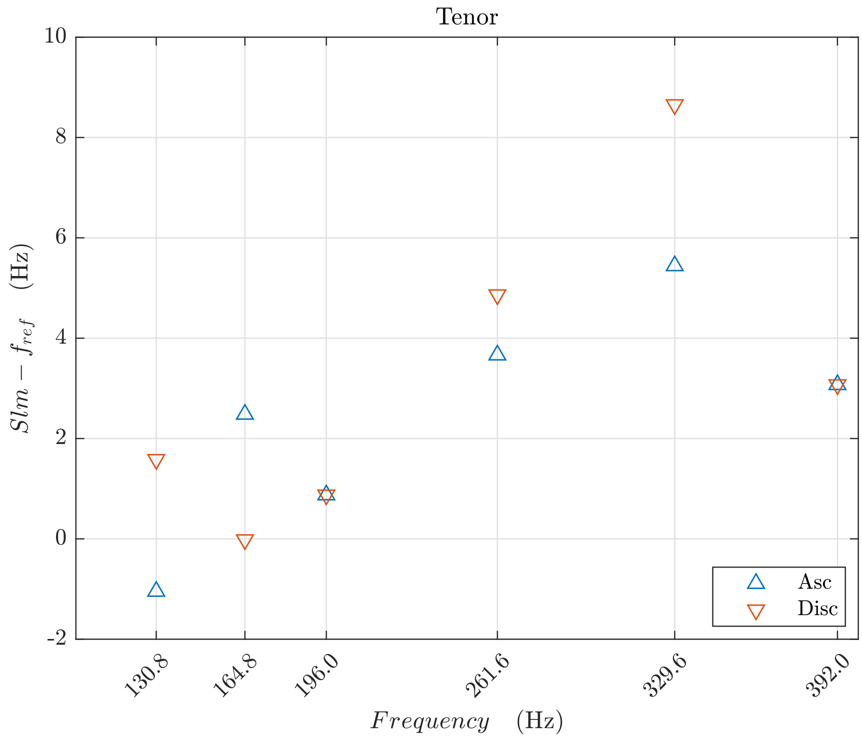


Figure 3.4. *Histeresys of singer - Legato and slow.*

From this figure, it was calculated the difference between the ascending and descending semi-phrase results. Even more is this difference high, even more the signal of SLM is far from the reference note. For Tenor, these results are resume in the Table 3.2 :

C	E	G	C	E	G
-2.6	2.5	0	-1.2	-3.2	0

Table 3.2. *Difference for each pitch for Tenor.*

Now, to evaluate if the singer is affected by ascendent or discendent part of arpeggio, is calculated the following parameter into the Formula 3.1. As this parameter increases, the singer is more affected by the ascending and descending semi-phrase and the singer is more inaccurate to do the same note in these types of arpeggios.

$$H = \frac{\sum_{k=1}^{N} |a_i|}{N} \tag{3.1}$$

Here N is the number of the considered notes (in this case N=5); the last note was never considered because was sung just once and this note is the same for the two semi-phrase (the difference is always 0). $|a_i|$ is absolute value of each difference between the ascending and descending semi-phrase for each note (for example, $|a_i|$ are the results obtained in the Table 3.2). For the example singers, the value of H is 1.9 Hz.
This value H, is calculated for a range of 8 tones and 3 semi-tones.
In Figure 3.5 are shown the results of all of the subjects.

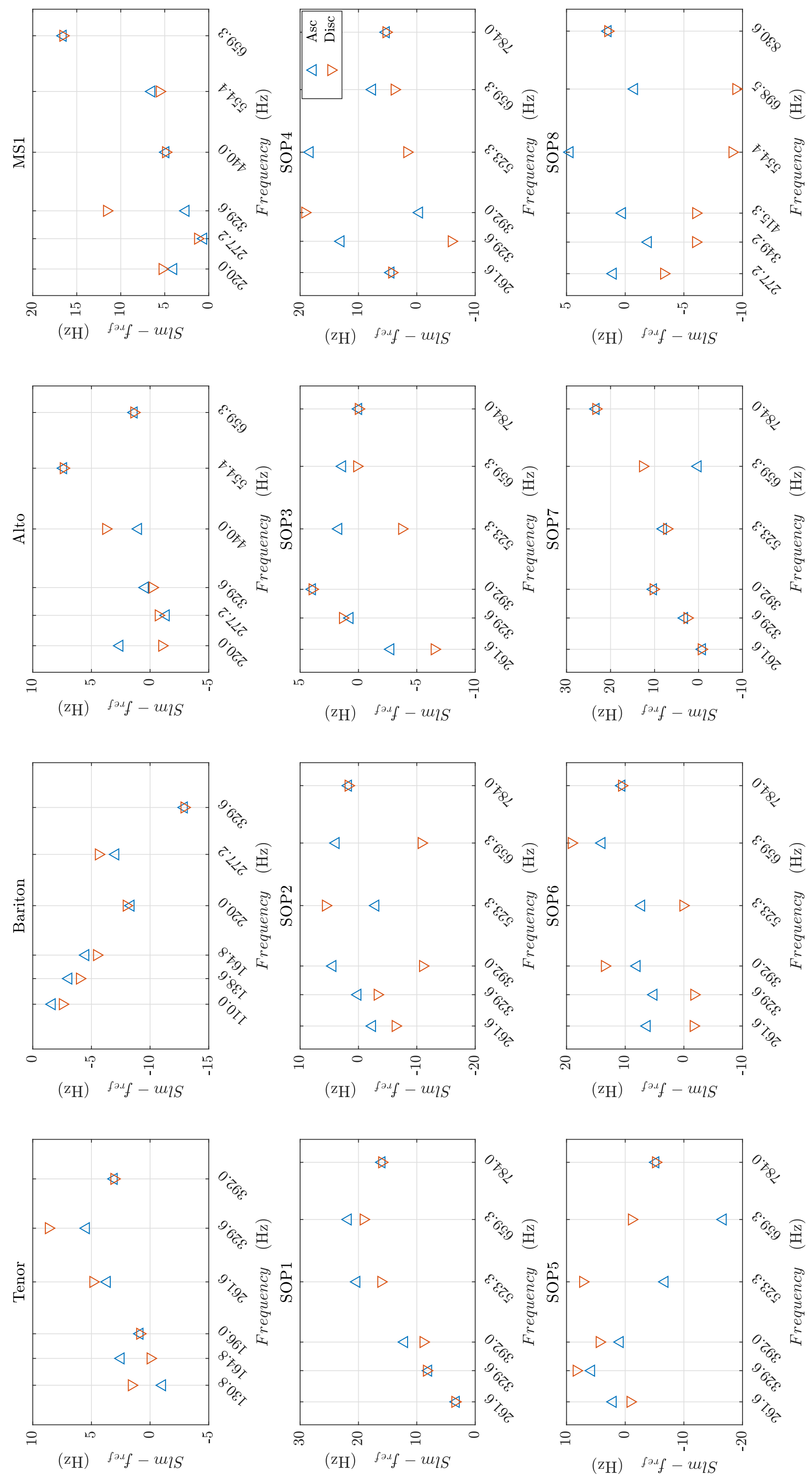


Figure 3.5. *Histeresys - Legato and slow.*

The results of the differences of all of the singers are shown in Table 3.3:

	1st Note	*2nd Note*	*3rd Note*	*4th Note*	*5th Note*	H (Hz)
Tenor	-2.6	2.5	0	-1.2	-3.2	1.9
Bariton	0.9	0.9	0.9	-0.4	-1.5	0.9
Alto	3.6	-0.6	0.6	-2.8	0	1.5
Ms1	-1.3	-0.7	9.0	0	0.9	2.4
Sop1	-0.2	-0.3	3.2	4.2	2.6	2.1
Sop2	3.9	3.4	15.4	-8.6	14.6	9.2
Sop3	3.7	-0.6	0	5.4	1.2	2.2
Sop4	0.2	18.9	-19.6	16.6	3.8	11.8
Sop5	2.9	-2.6	-3.5	-14.0	-15.6	7.7
Sop6	7.9	6.8	-5.5	7.1	-5.2	6.5
Sop7	-0.2	0.6	0	0.8	-12.6	2.8
Sop8	4.3	4.1	6.3	13.8	8.7	7.4

Table 3.3. *Differences and Histeresys of all of the singers.*

Following, is avaiable the last part of this section. Here, two results obtained from the equations 3.2 and 3.3 were appreciated:

$$f_{0_S} - f_{0_E} = S - E \quad (3.2) \qquad\qquad f_{0_S} - f_{0_P} = S - P \quad (3.3)$$

Where S represents the vector of output values of the frequency feels by SLM, E of ECM and P of piezoelectric microphones. Then these values are represent into the Figure 3.6 for the Tenor only. The two blue lines represent the *Absolute Resolution* range due to the sample frequency. The Absolute Resolution Equation 3.4 is:

$$\Delta f_o = f_o^2 \cdot T_s \tag{3.4}$$

Where f_o is the fundamental frequency of the note and T_s is the sampling step.
In this case T_s is the reverse of the sample frequency (there was an initial over-sampling of the signal at a sample frenquency of 352800 Hz) equal to 2.8 μs. The Equation 3.5 allows to obtain T_s:

$$T_s = \frac{1}{f_s} \tag{3.5}$$

If the results fall into these blue lines, the resolution will be due to the sample frequency; the ear does not perceive the difference between the note and the reference note if this difference falls into these blue lines. Figure 3.6 represents the output results of the Tenor.

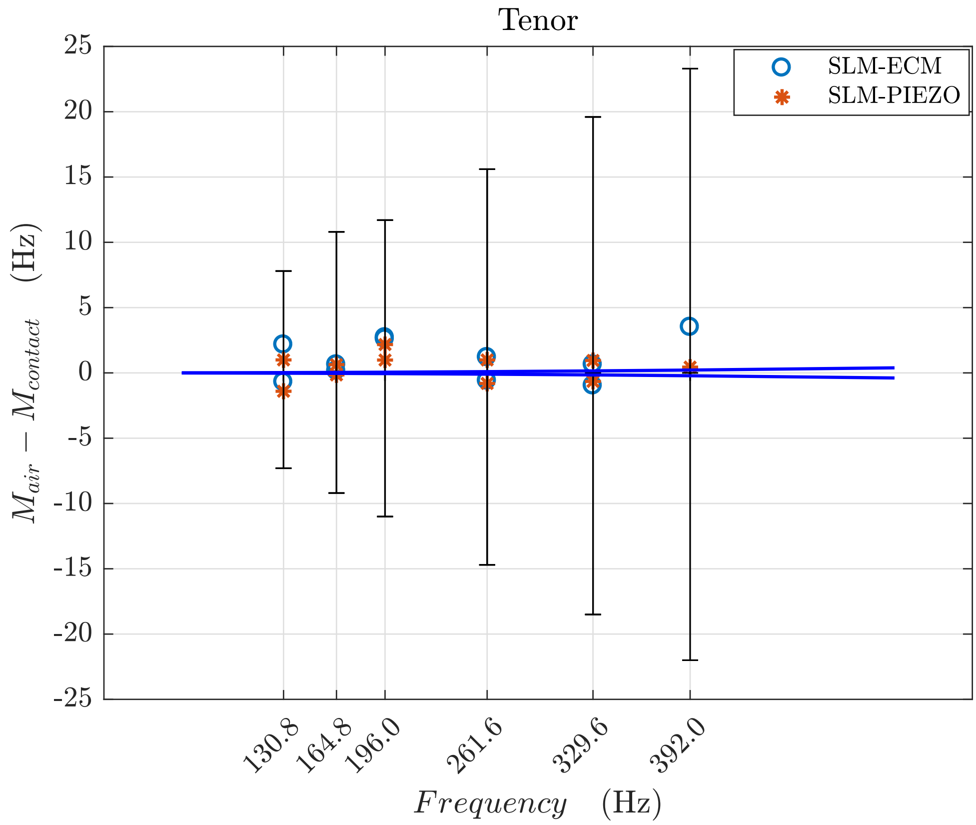


Figure 3.6. *Comparison between microphones - Legato and slow.*

In the Figure 3.6 there are the output results of the contact microphones relative to the microphone in air. The error bars represent the maximum value which is not possible to exceed to consider the singer in tune with that particular microphone. In this case, the output results for both contact microphones are in tune. The maximum value of Δf_o for tenor is 0.4 Hz that coresponds at the frequency of 392.0 Hz. For sopranos, the maximum value of Δf_o 1.7 Hz for a frequency of 784 Hz. Here are available all of the obtained results. The Figure 3.7 shows these results.

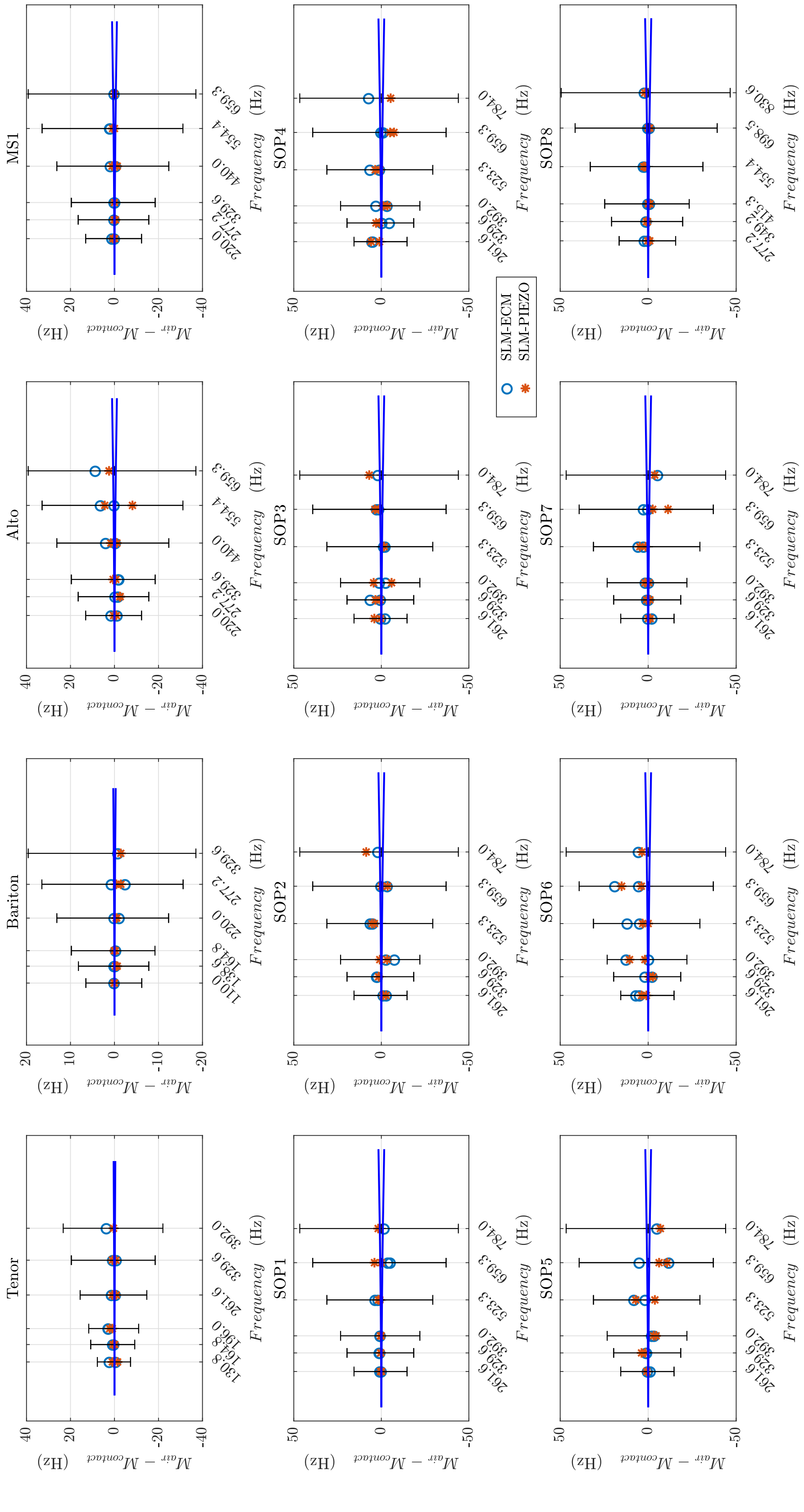


Figure 3.7. *Comparison between microphones - Legato and slow.*

These analysis were carried out also for the legato and fast, staccato and slow and staccato and fast arpeggios. Following, are avaiable the results for each arpeggio.

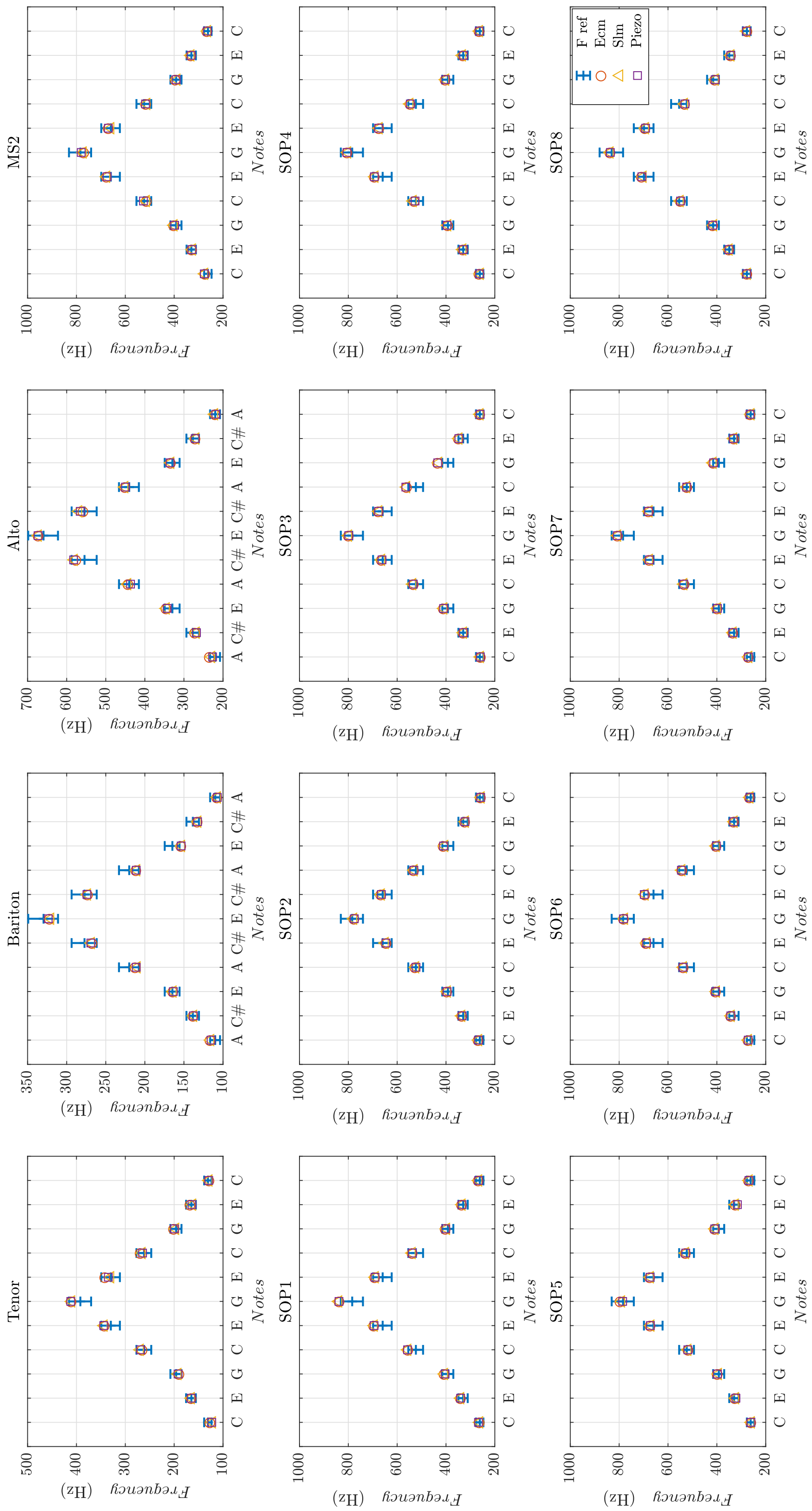


Figure 3.8. *Pitch Accuracy - Legato and fast.*

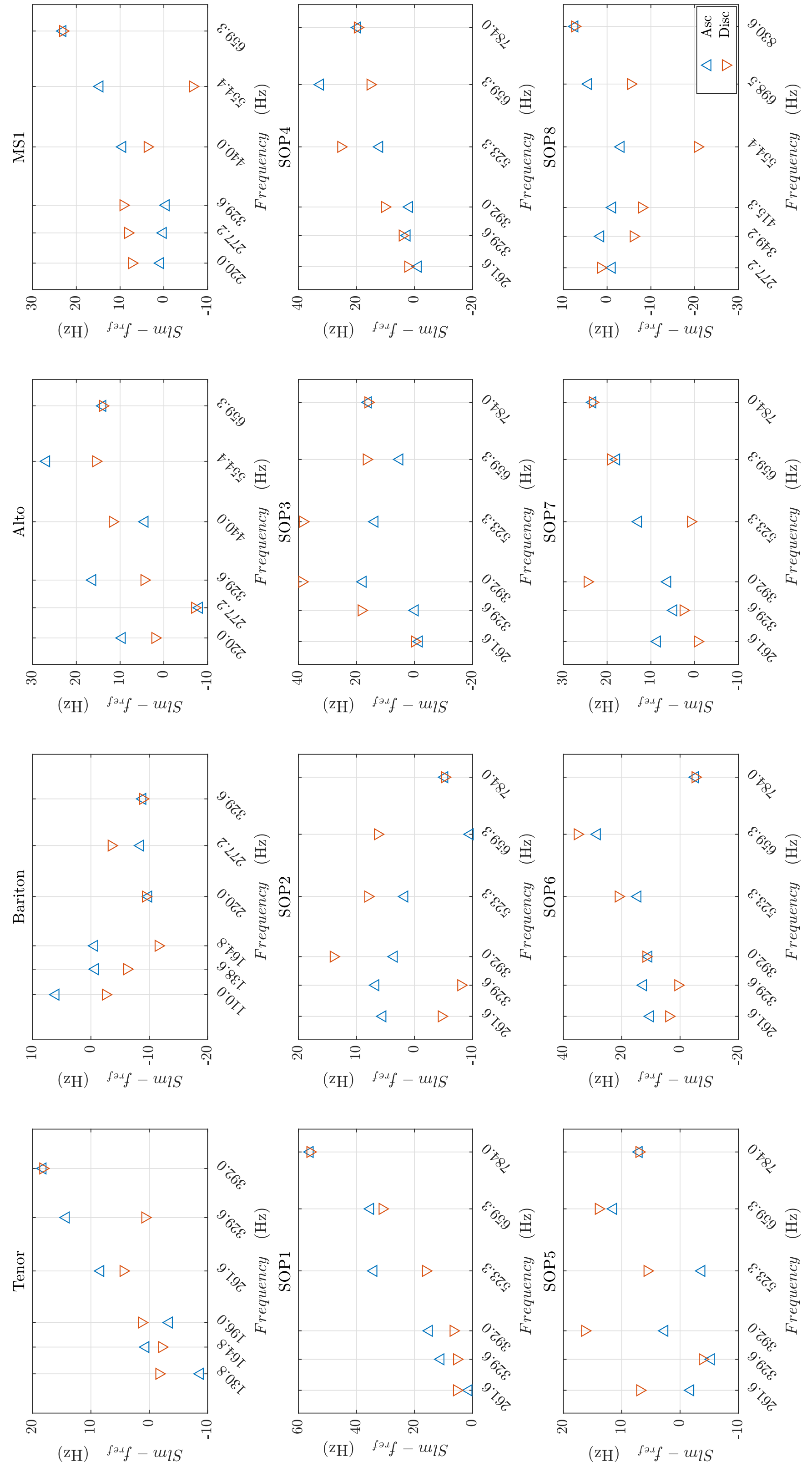


Figure 3.9. *Histeresys - Legato and fast.*

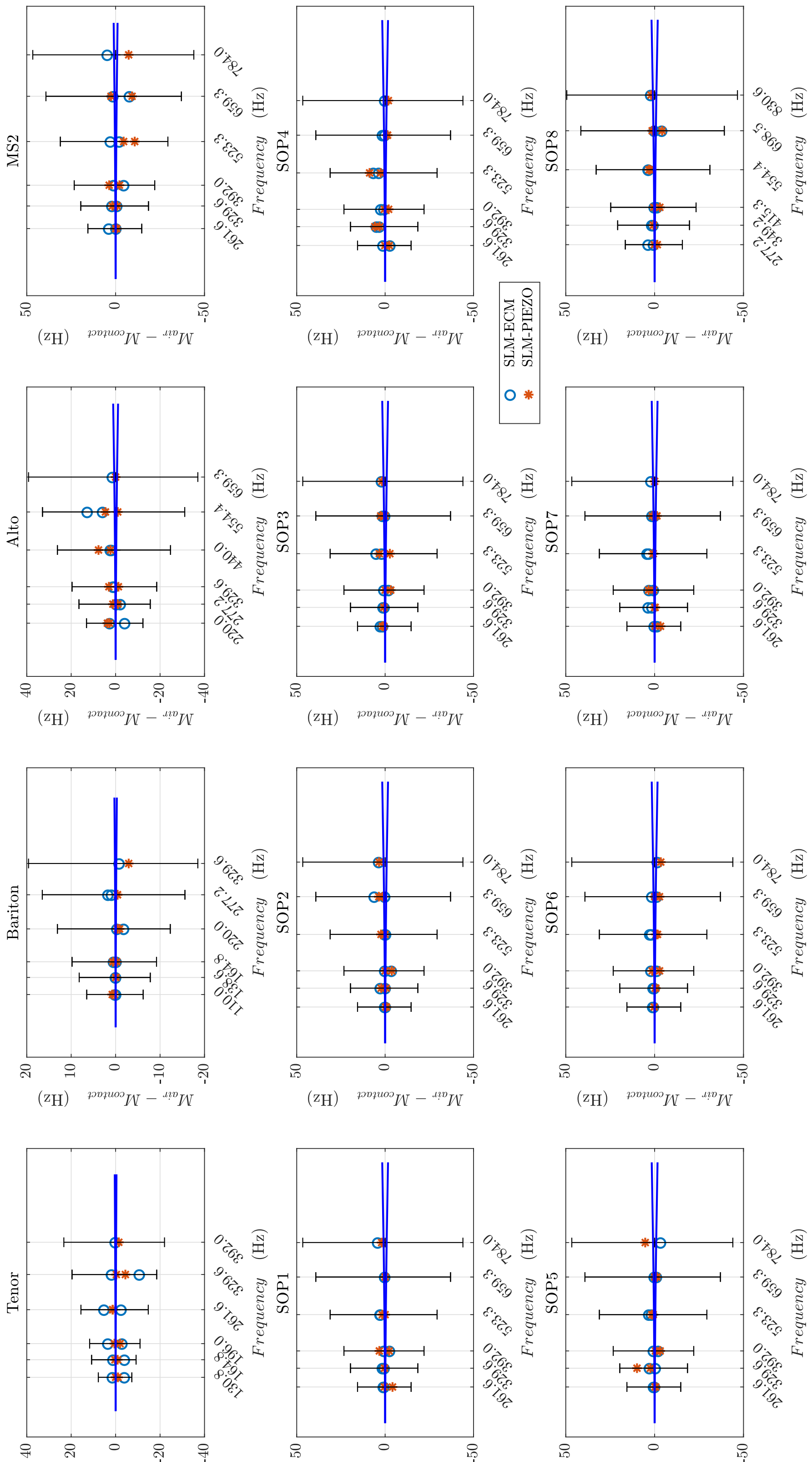


Figure 3.10. *Comparison between microphones - Legato and fast.*

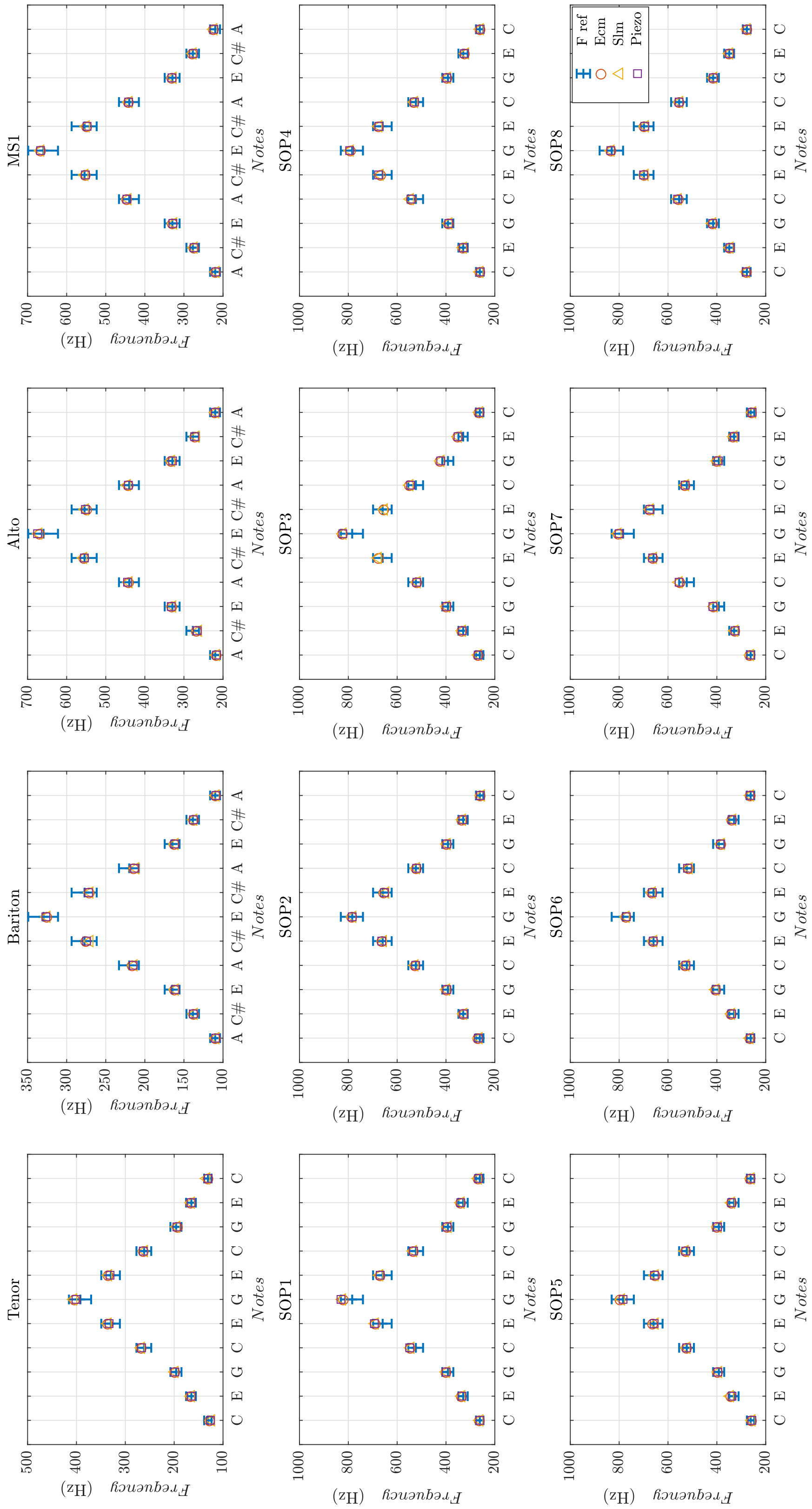


Figure 3.11. *Pitch Accuracy - Staccato and slow.*

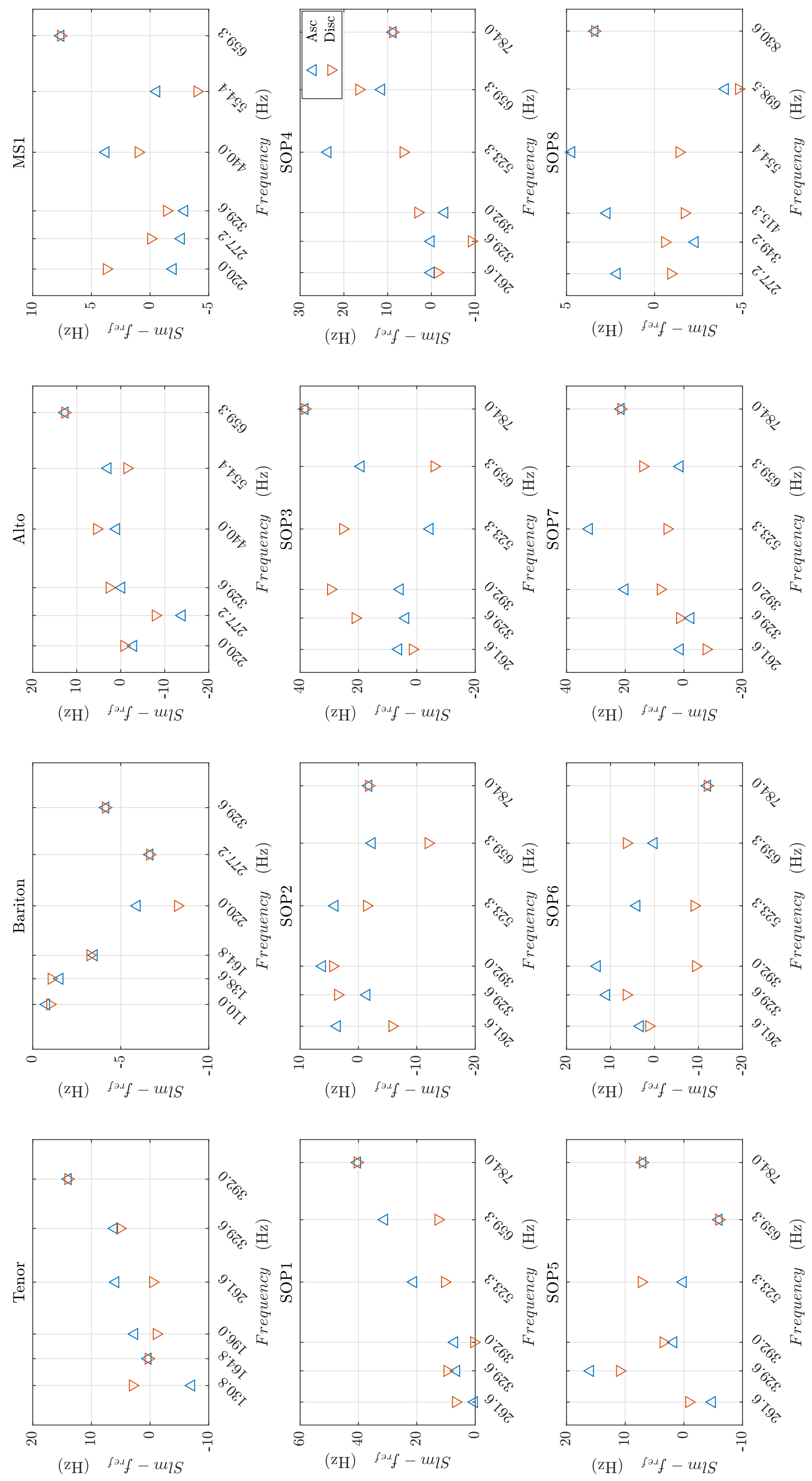


Figure 3.12. *Histeresys - Staccato and slow.*

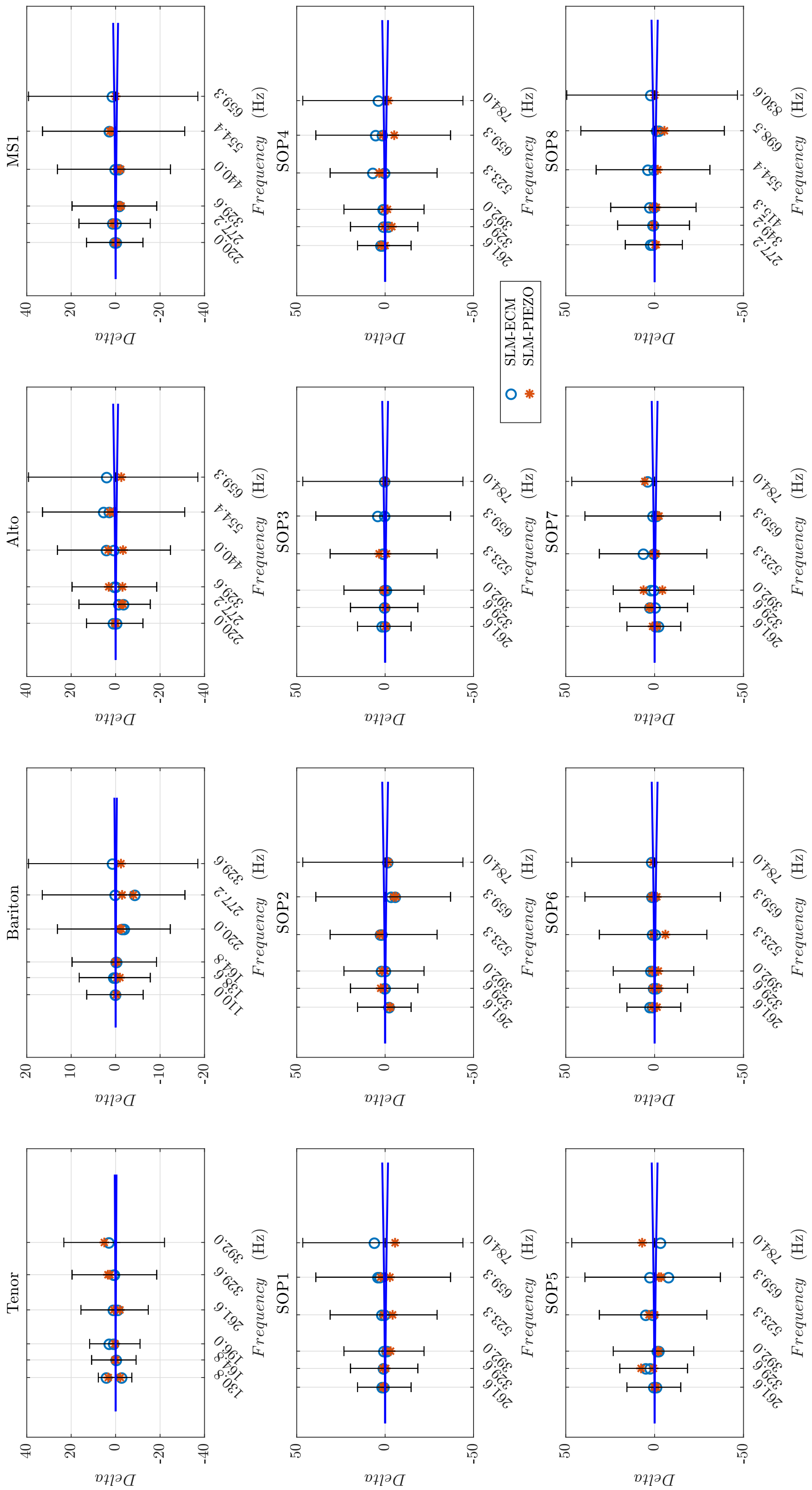


Figure 3.13. *Comparison between microphones - Staccato and slow.*

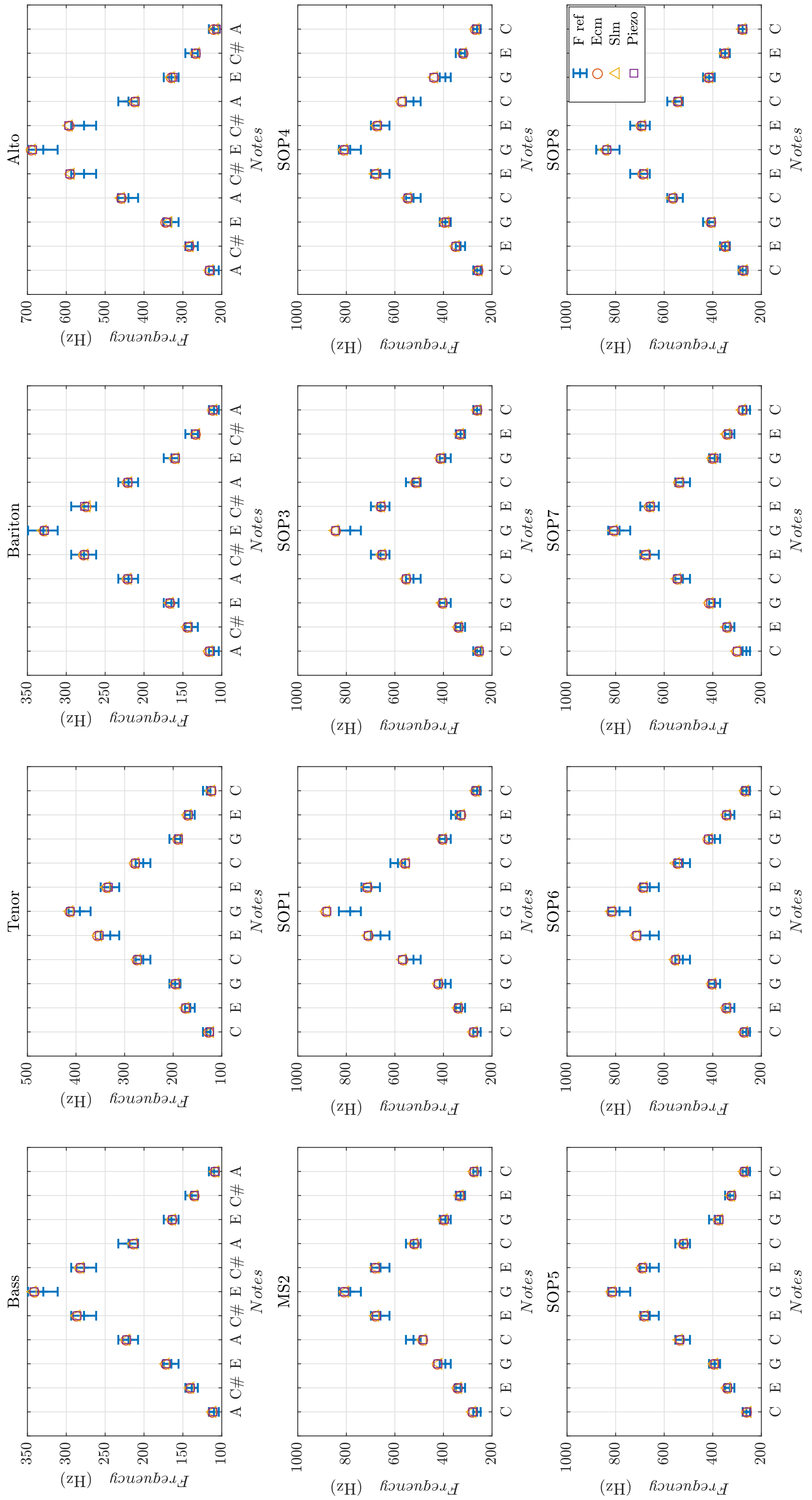


Figure 3.14. *Pitch Accuracy - Staccato and fast.*

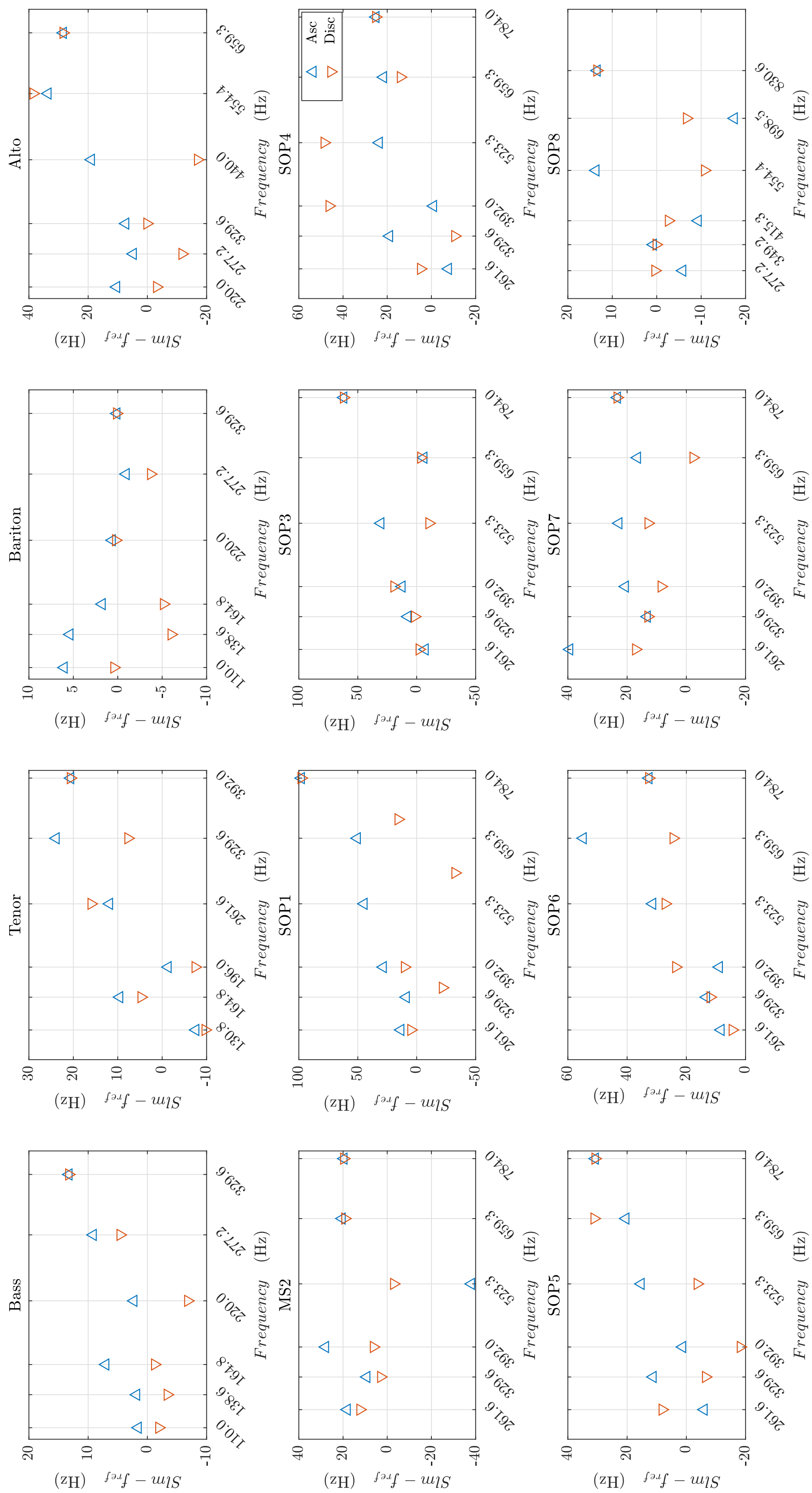


Figure 3.15. *Histeresys - Staccato and fast.*

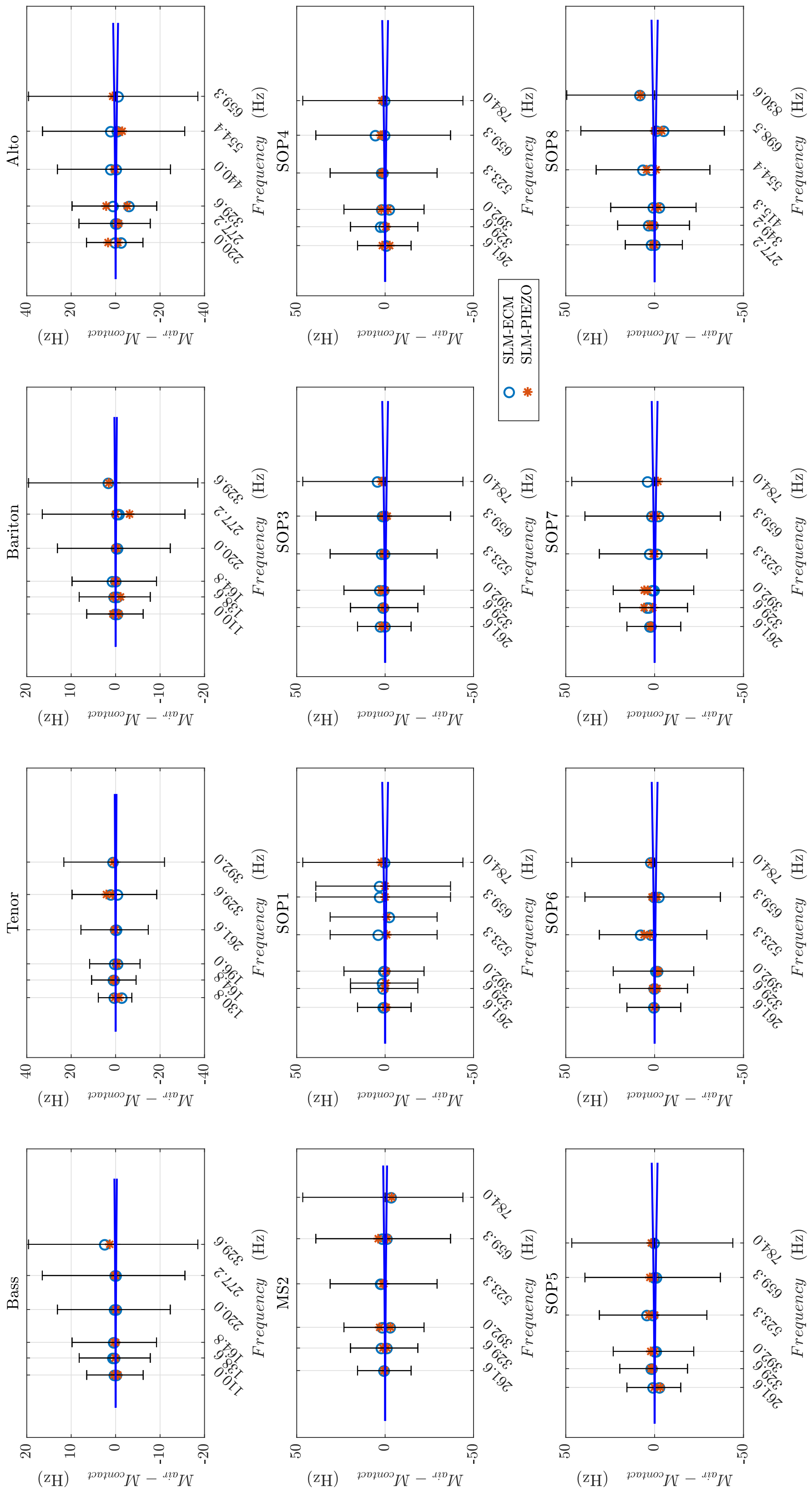


Figure 3.16. *Comparison between microphones - Staccato and fast.*

At the end, the comparison between the four types of arpeggios was obtained for professional singers. In the Figure 3.17 it is shown how is important the method to do the arpeggio: the worse results are obtained for the arpeggio staccato and fast. The 6^{th} subject sung a very different note from the reference one in the staccato and fast compared to the legato and slow one. The results are worse with the increasing of difficulty of the arpeggio.

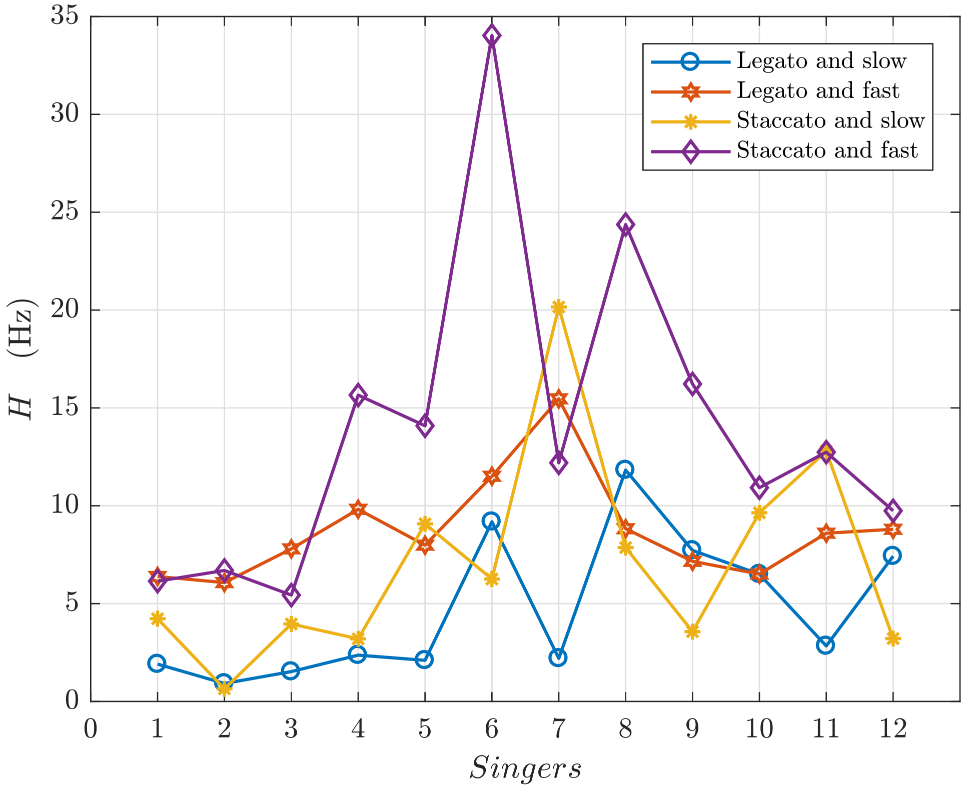


Figure 3.17. *H value for four arpeggios.*

The Table 3.4 summarizes the values of H for each subject for all of arpeggios.

Subjects	*Legato and slow*	*Legato and fast*	*Staccato and slow*	*Staccato and fast*
1^{st}	1.9	6.4	4.2	6.2
2^{nd}	0.9	6.1	0.6	6.7
3^{rd}	1.5	7.8	4.0	5.4
4^{th}	2.4	10.3	3.2	15.7
5^{th}	2.1	8.0	9.1	14.1
6^{th}	9.2	11.5	6.3	34.0
7^{th}	2.2	15.5	20.2	12.2
8^{th}	11.8	2.2	7.9	24.4
9^{th}	7.7	7.2	3.6	16.2
10^{th}	6.5	6.5	9.6	10.9
11^{th}	9.0	8.6	12.8	12.7
12^{th}	7.4	7.8	3.2	9.7

Table 3.4. *H value for trained singers.*

The same study was performed for the legato and slow arpeggio for untrained singers. There are more cases in which the singers are not in tune compared to the study about trained singers. The correlation of the two microphones is high in all of the cases. Then, the same provious studies were conducted also for unrtained singers. Following, it is possible to find the obtained results.

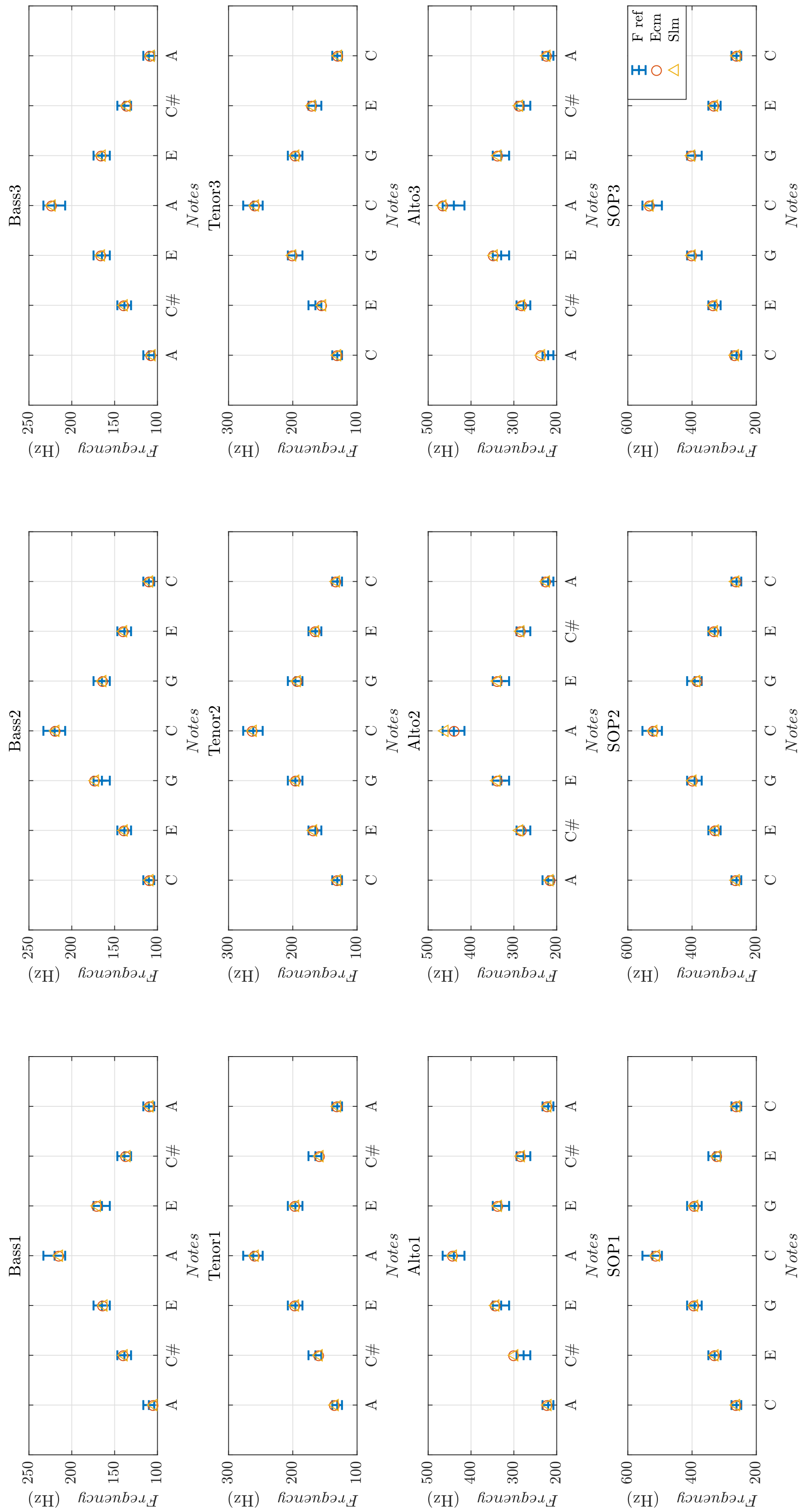


Figure 3.18. *Pitch Accuracy.*

The results of the differences between the output of SLM and the reference note and the parameter H of all of the singers are shown in Table 3.5:

	1st Note	*2nd Note*	*3rd Note*	H
Bass1	-4.58	3.42	-7.69	5.2
Bass2	-0.44	-0.66	8.66	3.3
Bass3	-0.9	3.26	0.85	1.7
Tenor1	3.58	1.35	0.22	1.7
Tenor2	-1.81	-3.14	2.36	2.4
Tenor3	1.24	-16.1	4.89	7.4
Alto1	-0.41	15.3	5.89	7.2
Alto2	-8.38	2.77	1.3	4.2
Alto3	12.52	-4.8	8.61	8.7
Sop1	0.97	7.2	0.44	2.9
Sop2	-1.64	-3.08	11.75	5.5
Sop3	4.72	2.18	-1.37	2.8

Table 3.5. *H value for untrained singers.*

To calculate the parameter H, the Formula 3.1 was used with N=3 because the untrained singers did an arpeggio of one octave and not an octave and half.

Figure 3.19. *Histeresys.*

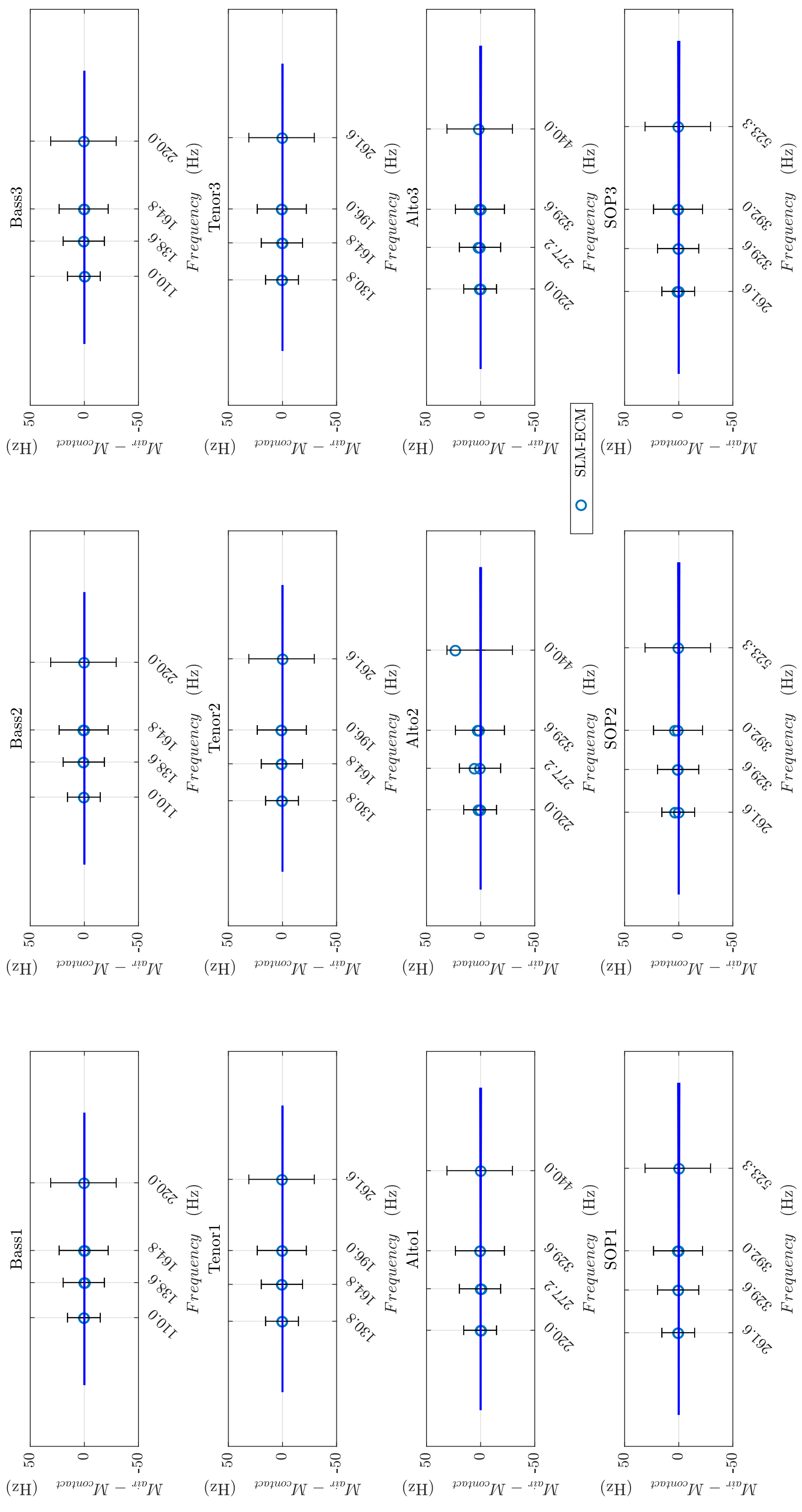


Figure 3.20. *Comparison between ECM and SLM.*

The results of H were compared for trained and not untrained singers for an octave.

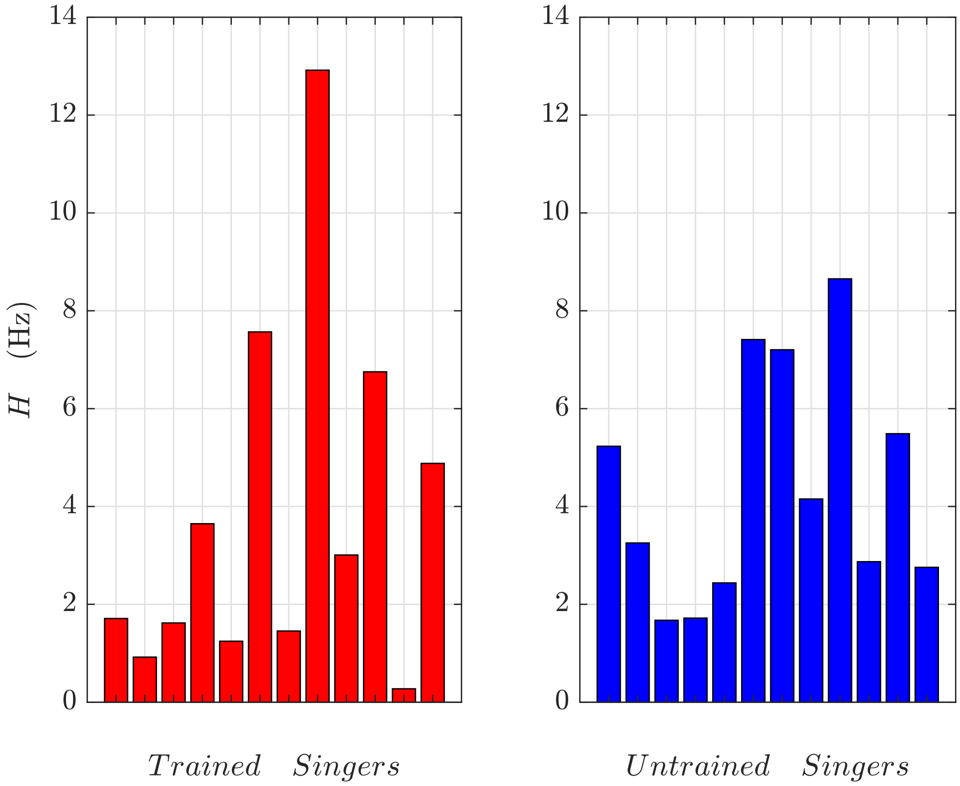


Figure 3.21. *Comparisonof H value between trained and untrained singers.*

It was obtained the summary value of H for the two categories of singers. In this case two values were obtained: the first one is the sum of the each value of H of all of the trained singers; the second one is the same value for untrained singers. In Table 3.6 the results are shown:

	Tr. Singers	*Untr. Singers*
H_{tot}	46 Hz	52.8 Hz

Table 3.6. *Comparison of* H_{tot}.

This parameter indicates that untrained singers are more influenced compared to the trained ones by the type of arpeggio (ascending and descending).

3.2 Cepstral Peak Prominence Smoothed

In this section it was calculated the CPPS distribution for the reading text and the songs. For songs, the CPPS was calculated in two observation range: the first one is between 50 Hz and 300 Hz and the second one is between 50 Hz and 600 Hz. Afterwards, there are the results for trained and untrained singers about reading text and songs.

The most different results are obtained for trained singers: the shape of CPPS is quite different between the reading text and the singing tasks. It is possible to do a qualitative analysis of the shape of this distribution. When the CPPS has a right-distribution, it means that the subject is healty. The distribution of CPPS has a better shape for the analysis of the singing tasks than the reading test.
Into the Figure 3.22 are shown the results about the reading text for three microphones. The features of the used microphones and the interpretation of results are analysed in detail from Castellana [31]. The bi-modal distribution indicates that the subject is sick. It is possibile to find the same distribution in the following figures and see how the distribution changes. For the singing tasks, the distribution becomes an healty distribution. This result is visible for the two categories of singers.

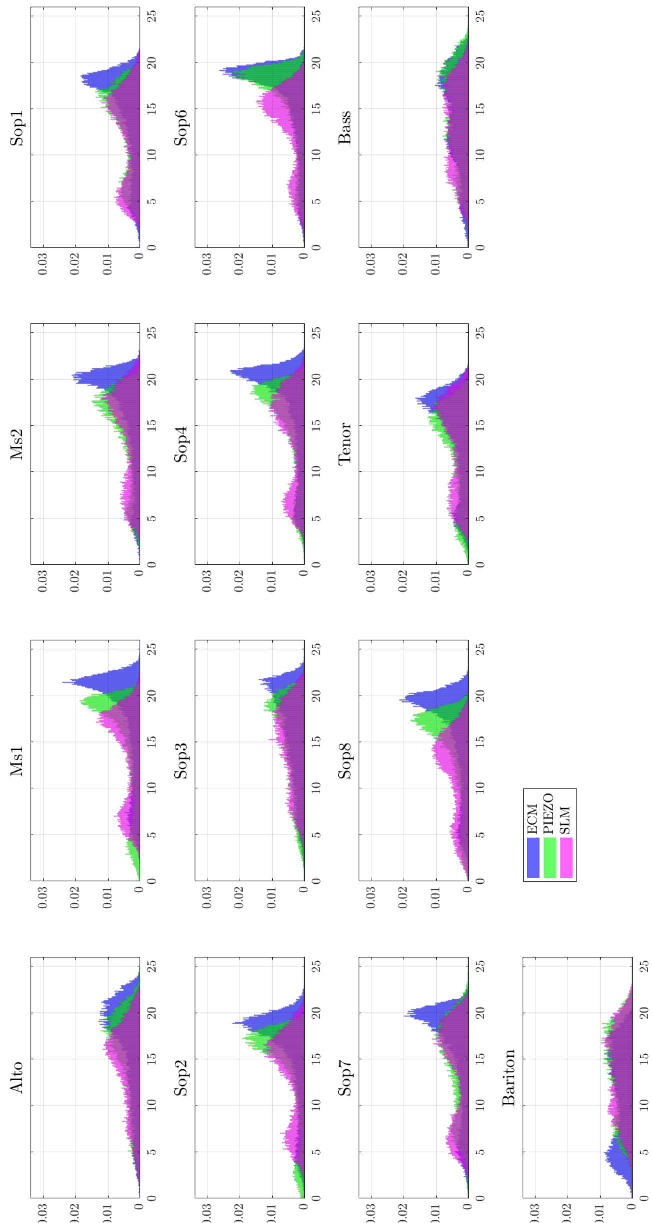


Figure 3.22. *CPPS reading text - Tra. Singers.*

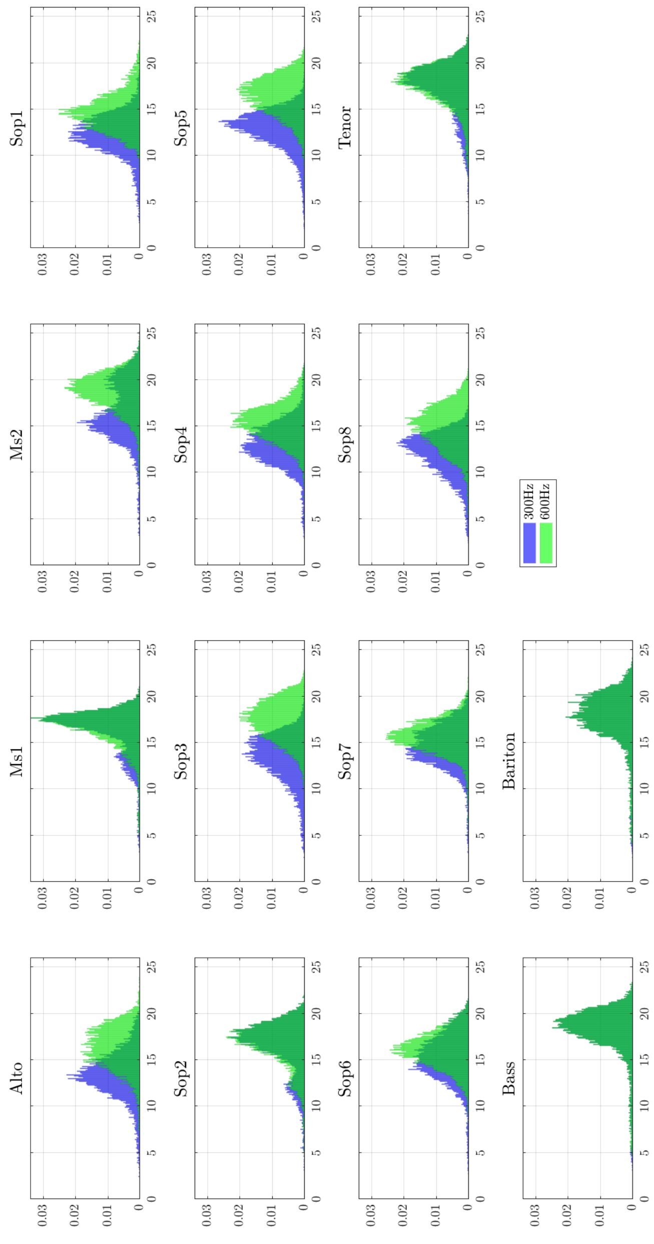


Figure 3.23. *CPPS "Tanti Auguri" Song - Tra. Singers - ECM.*

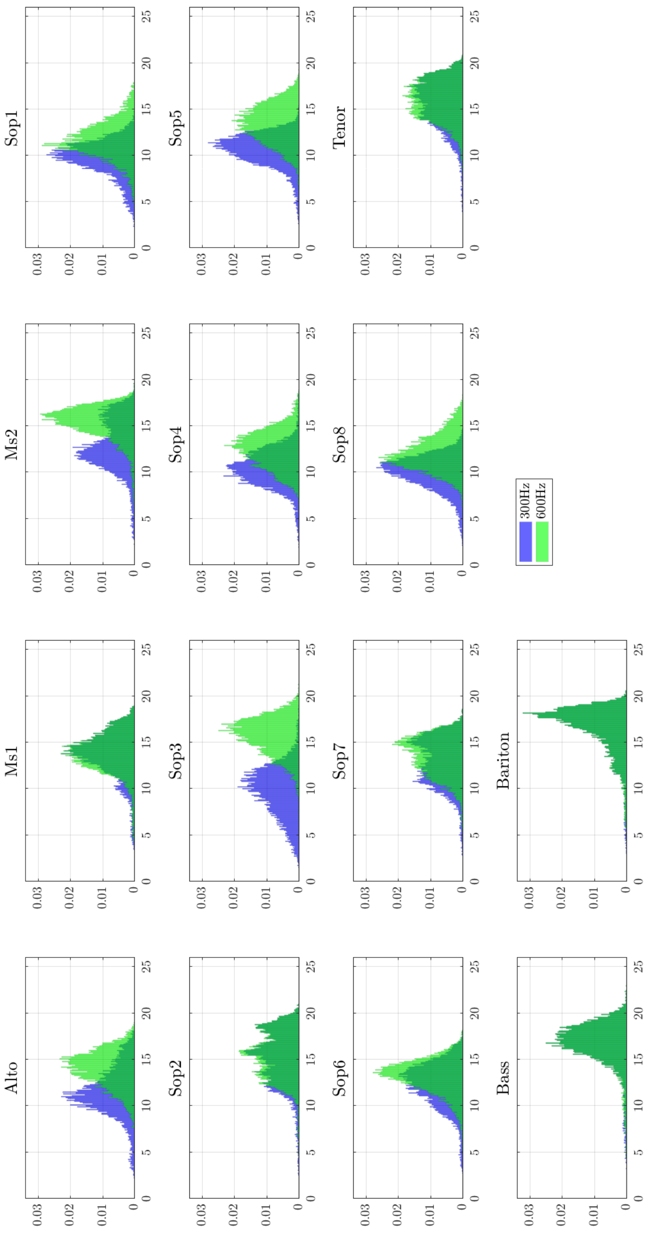


Figure 3.24. *CPPS "Tanti Auguri" Song - Tra. Singers - PIEZO.*

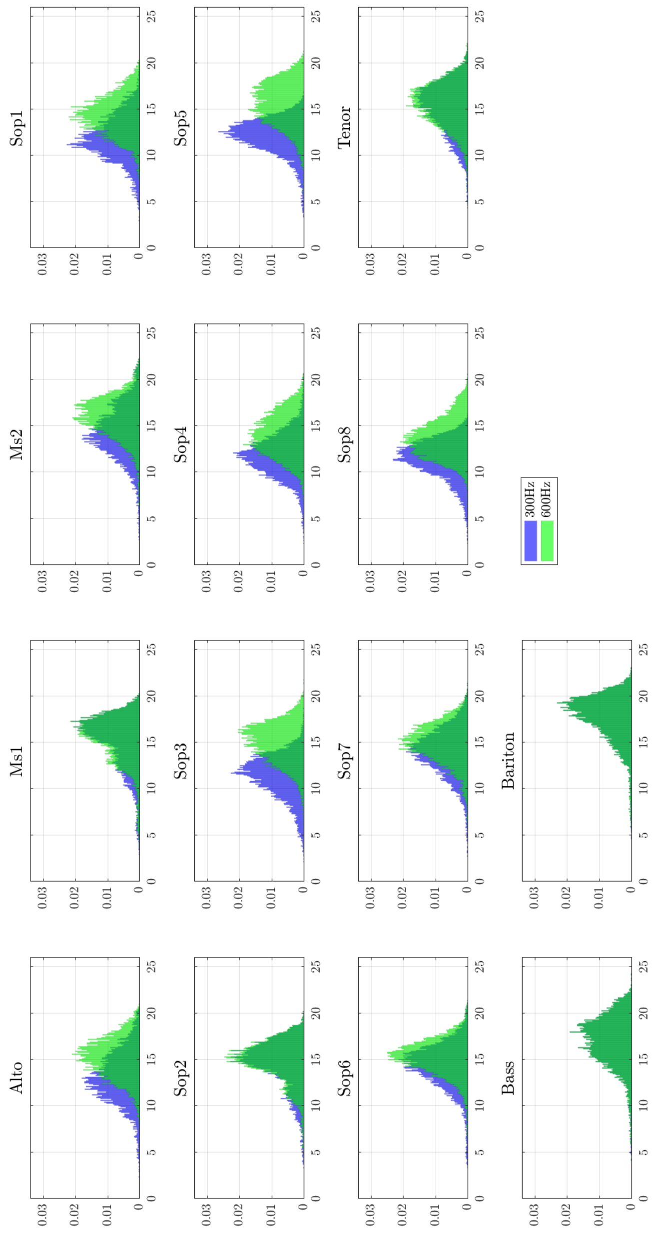


Figure 3.25. *CPPS "Tanti Auguri" Song - Tra. Singers - SLM.*

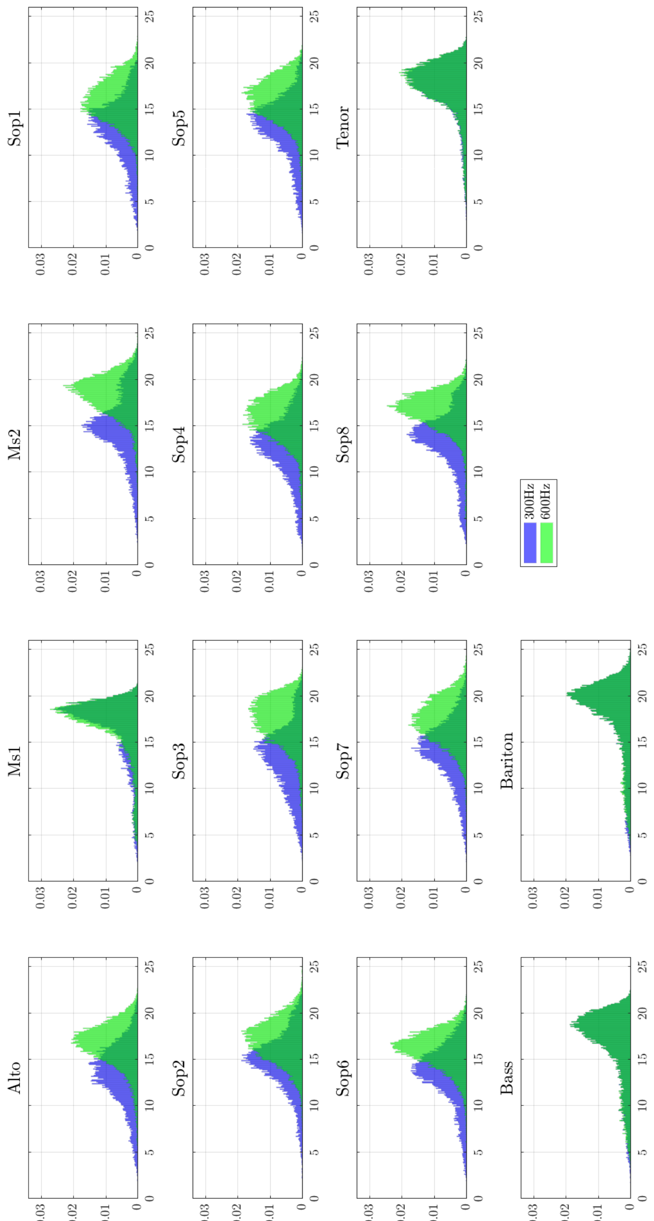


Figure 3.26. *CPPS "Inno di Mameli" Song - Tra. Singers - ECM.*

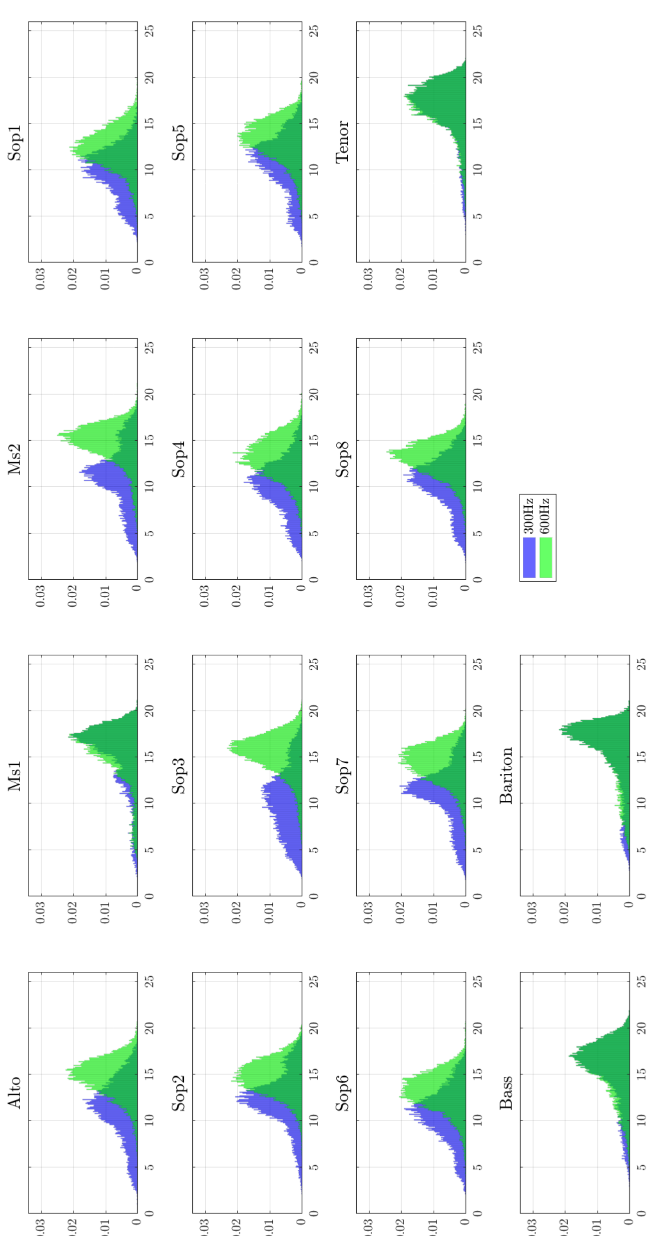


Figure 3.27. *CPPS "Inno di Mameli" Song - Tra. Singers - PIEZO.*

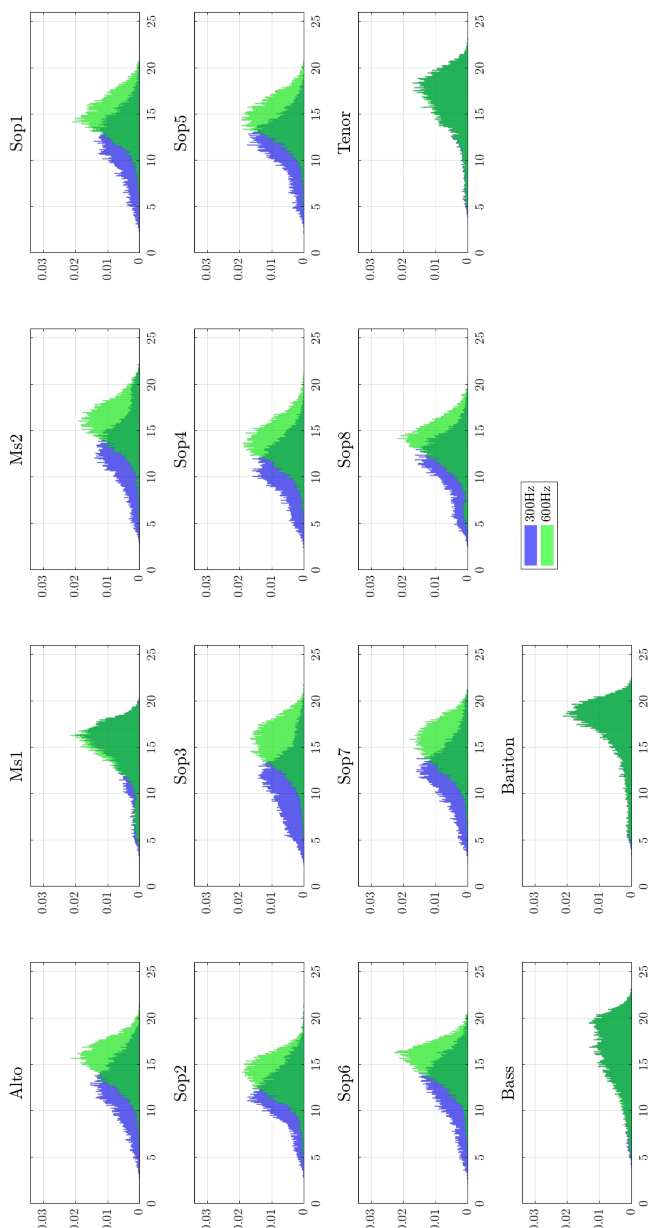


Figure 3.28. *CPPS "Inno di Mameli" Song - Tra. Singers - SLM.*

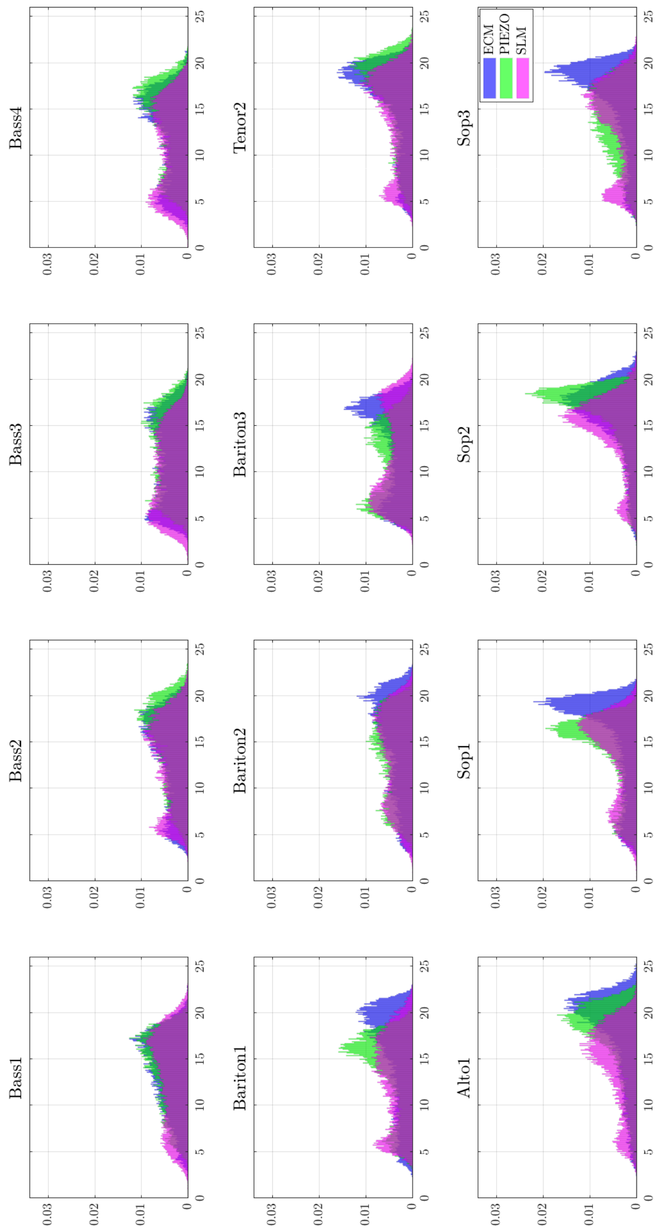


Figure 3.29. *CPPS reading text - Untra. Singers.*

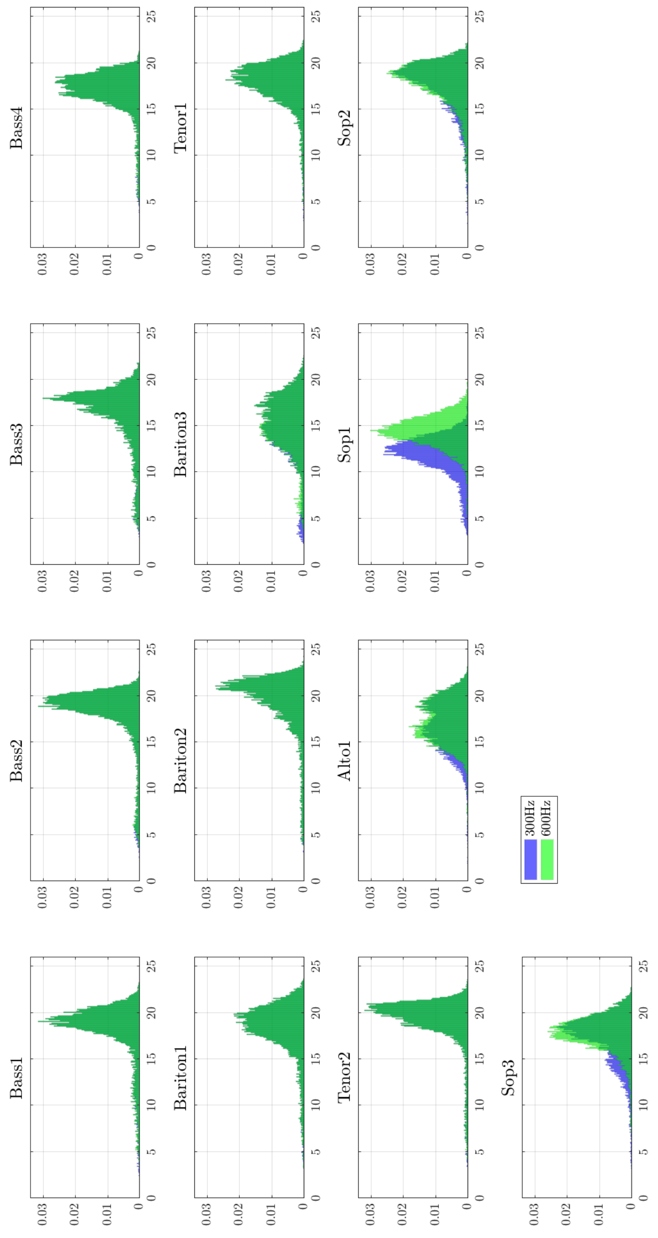


Figure 3.30. *CPPS "Tanti Auguri" Song - Untra. Singers - ECM.*

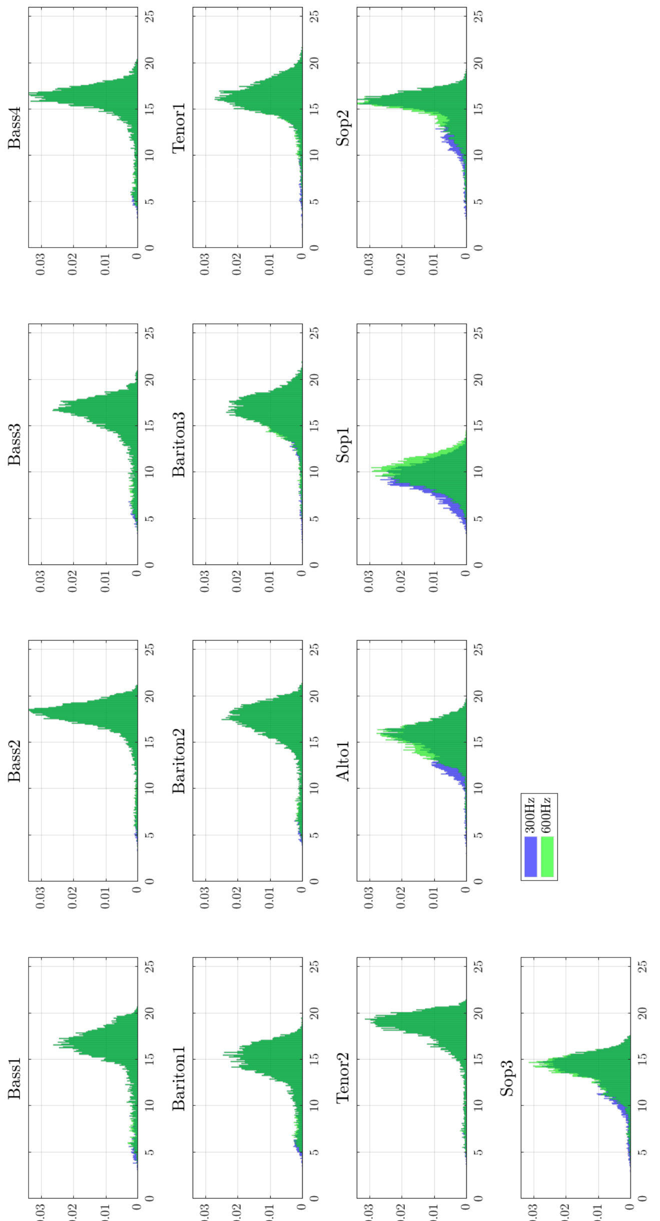


Figure 3.31. *CPPS "Tanti Auguri" Song - Untra. Singers - PIEZO.*

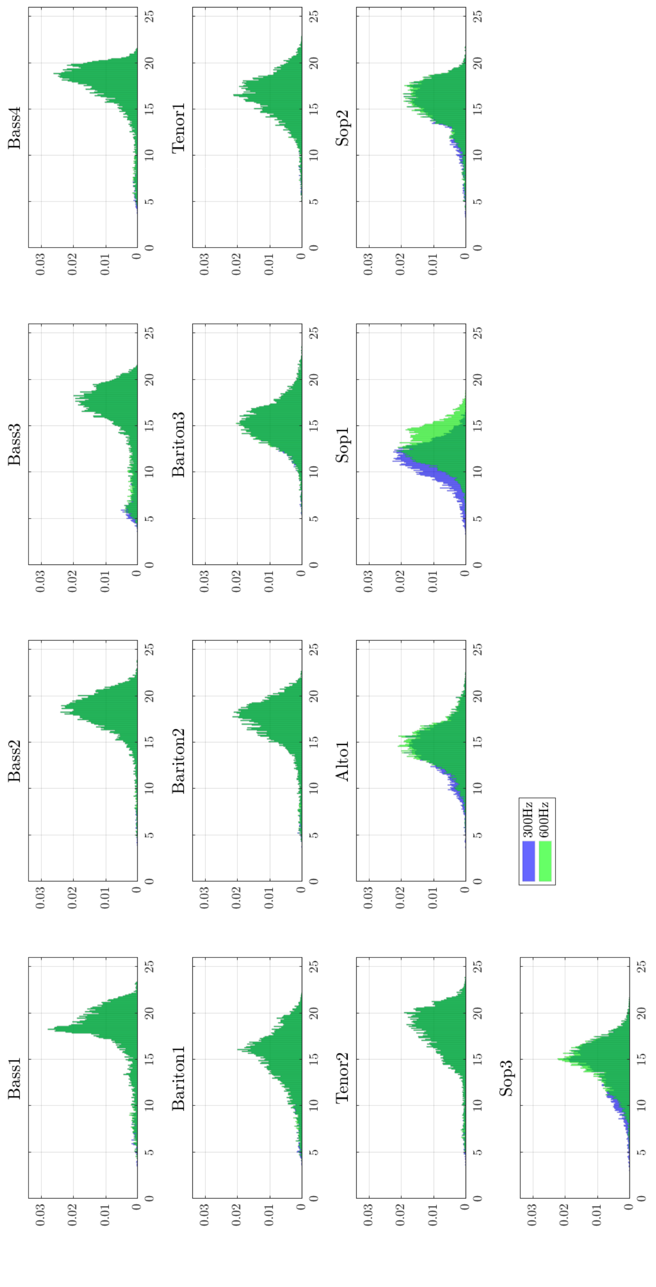


Figure 3.32. *CPPS "Tanti Auguri" Song - Untra. Singers - SLM.*

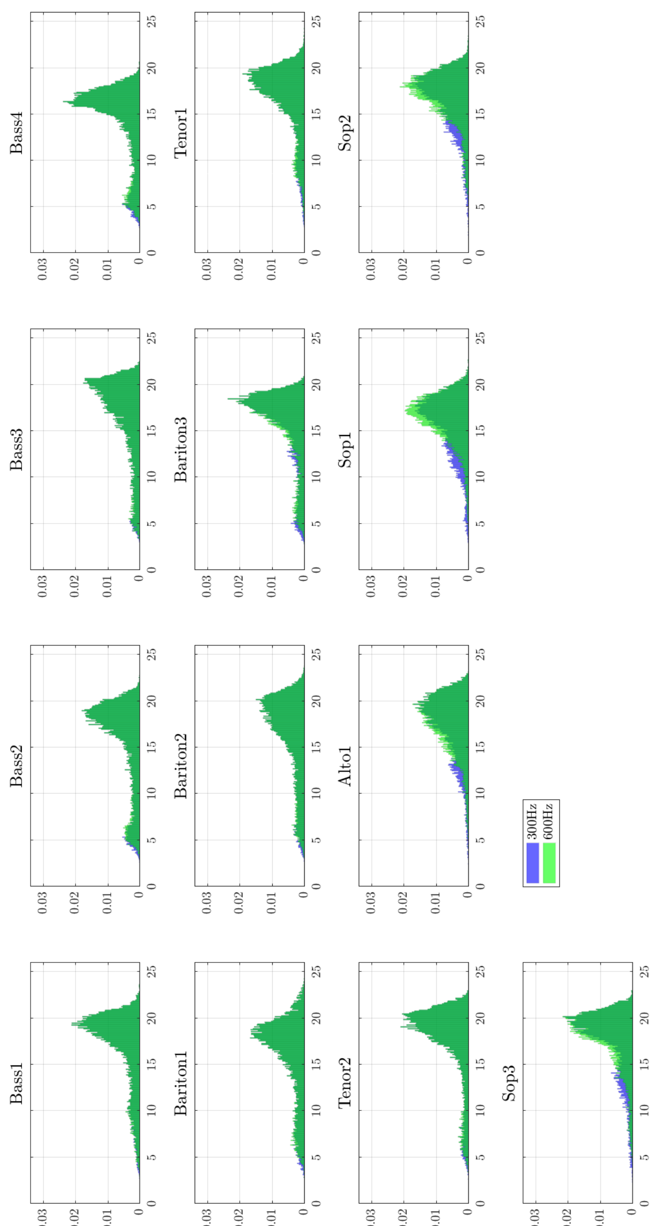


Figure 3.33. *CPPS "Inno di Mameli" Song - Untra. Singers - ECM.*

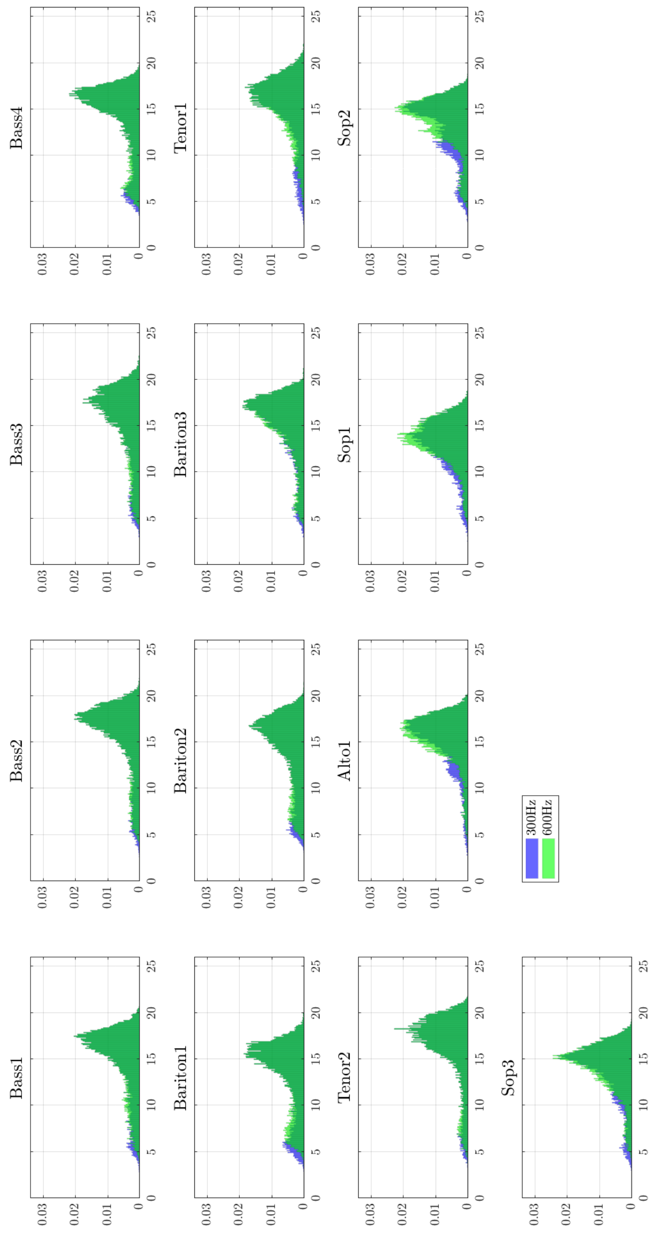


Figure 3.34. *CPPS "Inno di Mameli" Song - Untra. Singers - PIEZO.*

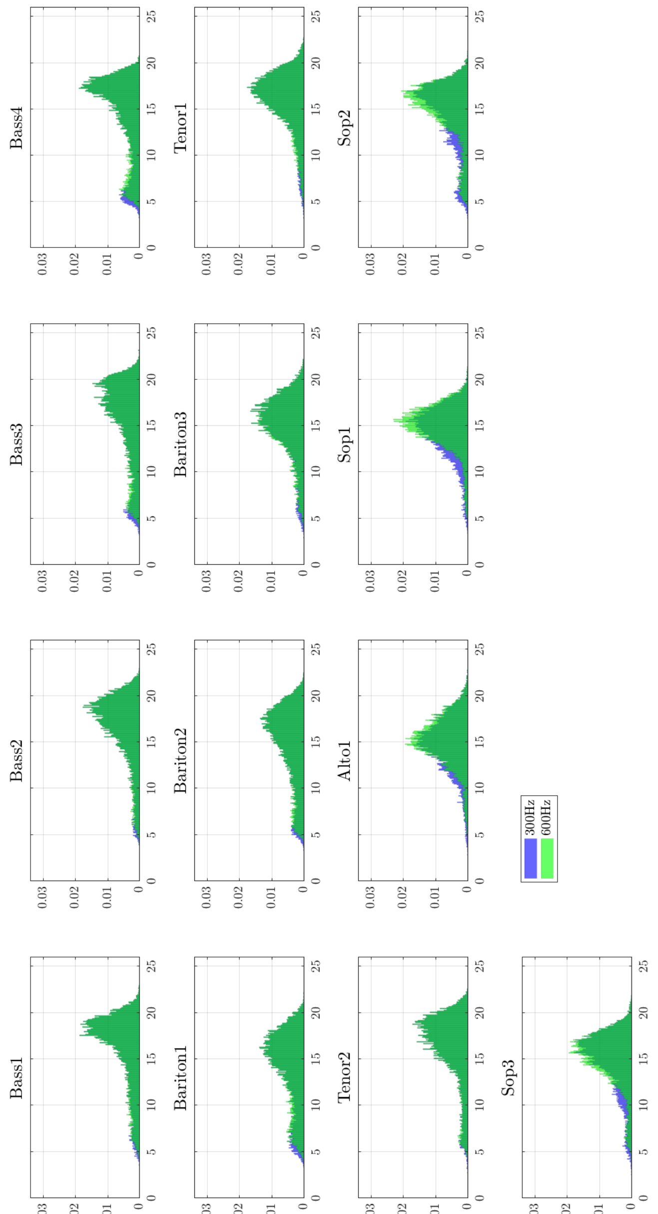


Figure 3.35. *CPPS "Inno di Mameli" Song - Untra. Singers - SLM.*

3.3 Long Time Average Spectrum

In this section are shown the results obtained with the LTAS analysis for the three microphones. The results for the songs sung by tenor are shown below in the Figure 3.36 and Figure 3.37. It is possible to see very well the Singer's Formant in the third graph on both figures (the microphone in air result). The Singer's Formant is a feature of trained singers that arise after years and years of training and it is a feature of the vocal tract (this is the reason why the Singer's Formant is so clear in the spectrum of the microphone in air). In the graphs of the spectrum of contact microphones, there is not a clear evidence of this peak into the range around 3000 Hz. The frequency information of these spectra might show the behaviours of glottis when a singer sing.
For each subject and each song, the differences show into the section 2.5.3. LTAS (ΔB_{low} and ΔB_{high}) were calculated. For the example subject, these results were obtained:

		Tanti Auguri		***Inno di Mameli***	
		ΔB_{low}	ΔB_{high}	ΔB_{low}	ΔB_{high}
Tenor	*ECM*	-10.0	-29.6	-13.8	-33.3
	PIEZO	-12.3	-29.5	-14.9	-36.6
	SLM	2.6	-12.6	0.1	-20.5

Table 3.7. *Obtained results for Tenor.*

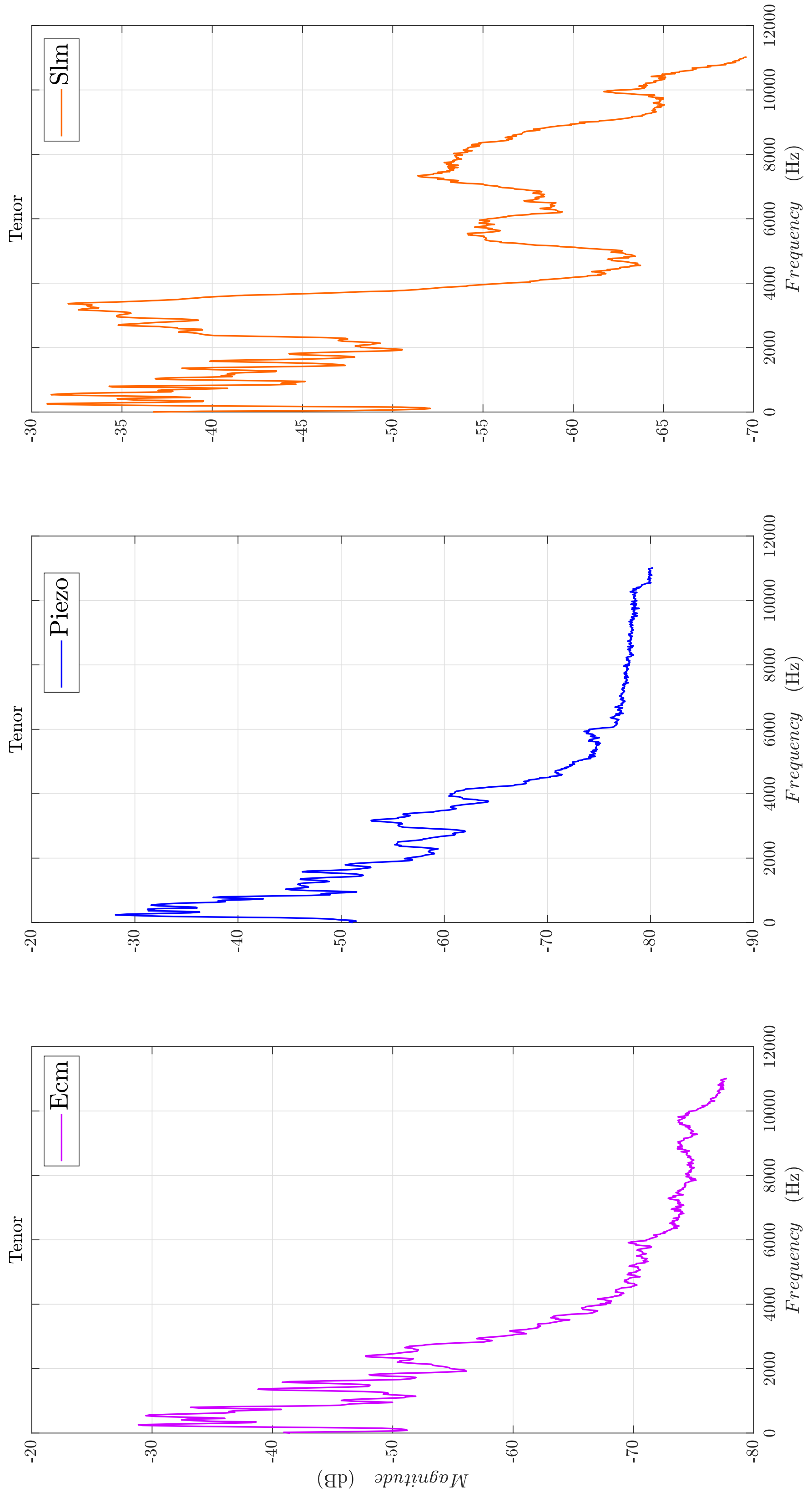


Figure 3.36. *LTAS of "Tanti auguri" song obtained with three devices.*

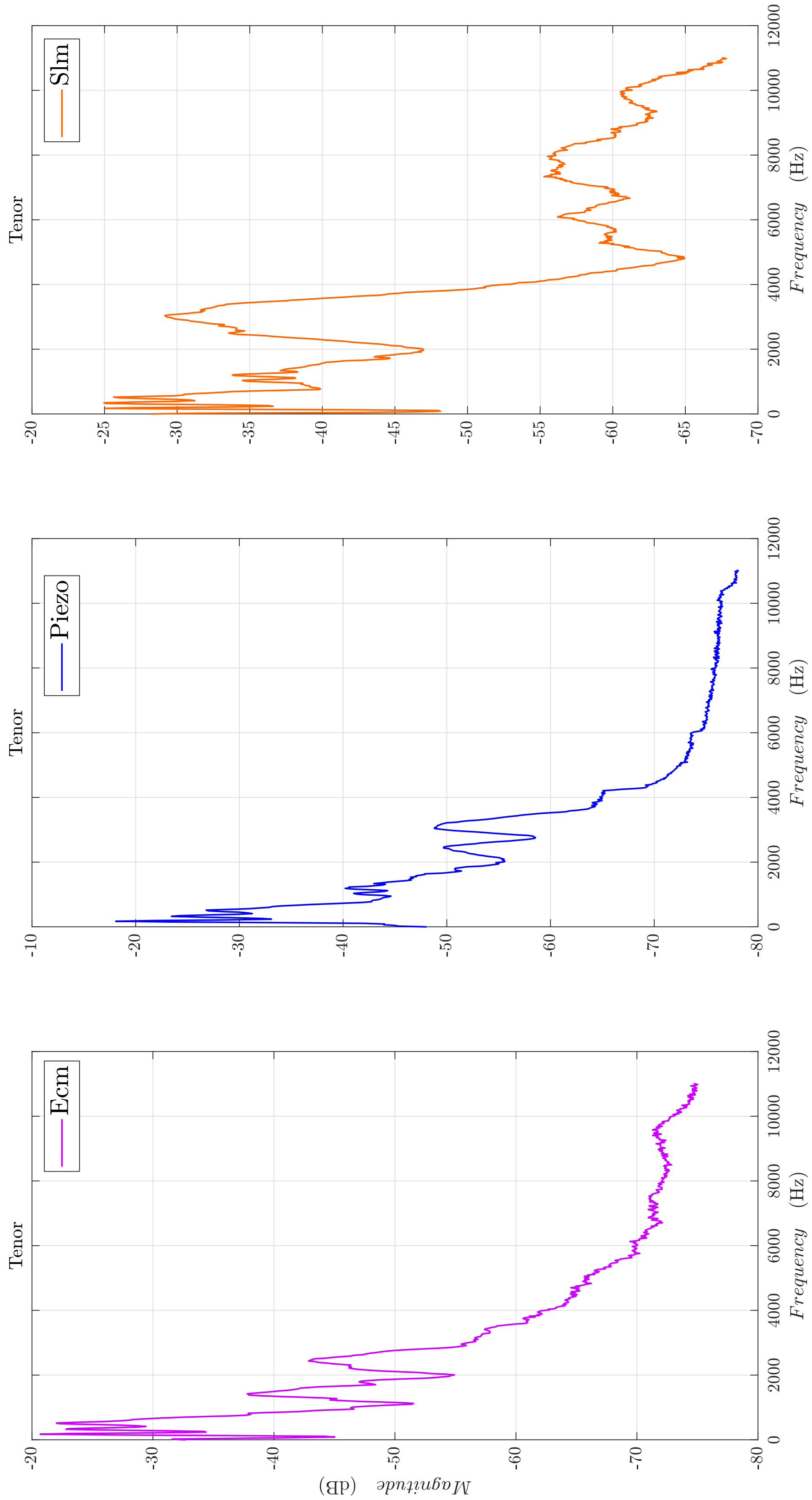


Figure 3.37. *LTAS of "Inno di Mameli" song obtained with three devices.*

These results were obtained for all of the singers. In the following Table 3.8 are sumamrized the results of trained singers obtained from the microphone in air and in the Table 3.9 are summarized the results for untrained singers:

	ΔB_{low}		ΔB_{high}	
	T. auguri	***Inno***	***T. auguri***	***Inno***
Bass	0.9	-0.9	-18.7	-22.5
Bariton	-2.6	-0.4	-22.9	-23.8
Tenor	2.6	0.1	-12.6	-20.5
Ms1	-9.3	-6.9	-15.3	-15.9
Ms2	5.7	1.7	-15.5	-17.0
Alto	4.7	1.2	-6.2	-12.9
Sop1	-3.2	-4.1	-9.2	-11.8
Sop2	-11.1	-5.8	-13.1	-13.1
Sop3	-3.4	-3.5	-12.1	-14.1
Sop4	-6.2	-5.9	-18.1	-14.4
Sop5	-5.8	-6.5	-17.3	-16.5
Sop6	1.6	0.1	-13.1	-15.8
Sop7	-8.9	-3.3	-15.1	-16.1
Sop8	-5.4	-6.3	-19.3	-16.9
Overall Mean	-2.9 ± 5.3	-2.9 ± 3.1	-14.9 ± 3.9	-16.8 ± 3.5
Mean F.	-3.7 ± 5.6	-3.6 ± 3.2	-14.4 ± 3.6	-15.6 ± 2.5
Mean M.	0.3 ± 2.7	-0.4 ± 0.5	-18.1 ± 5.2	-22.3 ± 1.7

Table 3.8. *Results for trained singers - SLM.*

	ΔB_{low}		ΔB_{high}	
	T. auguri	***Inno***	***T. auguri***	***Inno***
Bariton1	-15.1	-13.7	-19.8	-20.1
Bariton2	-10.3	-10.7	-21.1	-21.6
Bariton3	-5.7	-8.4	-20.6	-17.4
Bass1	-11.4	-9.7	-24.0	-24.7
Bass2	-4.2	-4.0	-21.2	-21.5
Bass3	-19.5	-14.2	-23.8	-24.3
Bass4	-14.5	-15.8	-21.2	-22.0
Sop1	-6.3	-5.1	-10.2	-10.9
Sop2	-9.1	-9.1	-14.3	-14.2
Sop3	-7.1	-7.7	-12.9	-16.4
Alto1	-3.7	-3.8	-12.7	-13.9
Tenor1	-8.6	-8.3	-14.5	-17.0
Tenor2	-9.4	-7.3	-18.4	-18.1
Overall Mean	-9.6 ± 4.6	-9.1 ± 3.8	-18.0 ± 4.6	-18.6 ± 4.2
Mean F.	-6.6 ± 2.2	-6.4 ± 2.4	-12.5 ± 1.7	-13.9 ± 2.3
Mean M.	-8.8 ± 8.3	-10.2 ± 3.8	-20.5 ± 2.9	-20.7 ± 2.8

Table 3.9. *Results for untrained singers - SLM.*

Following it is possible to observe the results for the other microphones in the Table 3.10 and 3.11 for trained singers and in the Table 3.12 and 3.13 for untrained ones:

		Tanti Auguri		***Inno di Mameli***	
		ΔB_{low}	ΔB_{high}	ΔB_{low}	ΔB_{high}
	ECM	-18.3	-34.4	-17.4	-35.4
Bass	*PIEZO*	-13.1	-34.5	-15.3	-32.0
	SLM	0.9	-18.7	-0.9	-22.5
	ECM	-11.3	-30.2	-5.5	-35.8
Bariton	*PIEZO*	-14.9	-32.6	-12.4	-34.2
	SLM	-2.6	-22.9	-0.4	-23.8
	ECM	-10.0	-29.6	-13.8	-33.3
Tenor	*PIEZO*	-12.3	-29.5	-14.9	-36.6
	SLM	2.6	-12.6	0.1	-20.5
	ECM	-20.2	-27.9	-22.7	-31.9
Ms1	*PIEZO*	-19.1	-22.4	-19.9	-29.8
	SLM	-9.3	-15.3	-6.9	-15.9
	ECM	-5.7	-31.2	-8.4	-32.3
Ms2	*PIEZO*	-7.1	-22.5	-8.7	-24.2
	SLM	5.7	-15.5	1.7	-17.0
	ECM	-9.1	-20.9	-8.3	-20.9
Alto	*PIEZO*	-6.0	-18.5	-9.0	-20.8
	SLM	4.71	-6.2	1.2	-12.9

Table 3.10. *Results for trained singers - ECM, PIEZO, SLM.*

		Tanti Auguri		*Inno di Mameli*	
		ΔB_{low}	ΔB_{high}	ΔB_{low}	ΔB_{high}
Sop1	*ECM*	-12.7	-26.4	-12.4	-27.5
	PIEZO	-4.4	-11.4	-4.0	-11.5
	SLM	-3.2	-9.2	-4.1	-11.8
Sop2	*ECM*	-20.0	-30.3	-12.9	-30.9
	PIEZO	-17.6	-32.2	-14.6	-29.1
	SLM	-11.1	-13.1	-5.8	-13.1
Sop3	*ECM*	-7.8	-20.6	-11.8	-24.3
	PIEZO	-5.8	-15.9	-9.8	-22.1
	SLM	-3.4	-12.1	-3.5	-14.1
Sop4	*ECM*	-11.0	-24.5	-13.7	-27.9
	PIEZO	-15.1	-26.1	-11.3	-22.7
	SLM	-6.2	-181	-5.9	-14.4
Sop5	*ECM*	-10.4	-23.8	-15.5	-27.7
	PIEZO	-15.3	-23.3	-12.4	-21.1
	SLM	-5.8	-17.3	-6.5	-16.5
Sop6	*ECM*	-14.4	-30.9	-15.1	-32.0
	PIEZO	-10.2	-20.6	-11.5	-22.5
	SLM	1.6	-13.1	0.1	-15.8
Sop7	*ECM*	-15.5	-26.8	-13.4	-29.6
	PIEZO	-12.9	-20.2	-12.9	-24.3
	SLM	-8.9	-15.1	-3.3	-16.1
Sop8	*ECM*	-11.6	-24.6	-17.6	-29.8
	PIEZO	-9.5	-19.6	-12.2	-22.2
	SLM	-5.4	-19.3	-6.3	-16.9
Overall Mean	*ECM*	-12.7 ± 4.4	-27.3 ± 4.1	-13.5 ± 4.3	-30.0 ± 4.1
	PIEZO	-11.7 ± 4.6	-23.5 ± 6.7	-12.1 ± 3.7	-25.2 ± 6.5
Mean F.	*ECM*	-12.6 ± 4.6	-26.2 ± 3.7	-13.8 ± 4.1	-28.6 ± 3.5
	PIEZO	-11.2 ± 5.1	-21.2 ± 5.4	-11.5 ± 4.0	-22.7 ± 4.8
Mean M.	*ECM*	-13.2 ± 4.5	-31.4 ± 2.6	-12.2 ± 6.1	-34.8 ± 1.3
	PIEZO	-13.4 ± 1.3	-32.2 ± 2.5	-14.2 ± 1.6	-34.3 ± 2.3

Table 3.11. *Results for trained singers - ECM, PIEZO, SLM.*

		Tanti Auguri		***Inno di Mameli***	
		ΔB_{low}	ΔB_{high}	ΔB_{low}	ΔB_{high}
	ECM	-27.4	-39.2	-24.5	-36.3
Bariton1	*PIEZO*	-15.8	-18.2	-17.6	-23.9
	SLM	-15.1	-19.8	-13.7	-20.1
	ECM	-25.5	-38.8	-22.1	-36.4
Bariton2	*PIEZO*	-13.4	-22.3	-15.8	-24.9
	SLM	-10.3	-21.1	-10.7	-21.6
	ECM	-18.6	-30.7	-24.6	-32.2
Bariton3	*PIEZO*	-14.7	-29.9	-11.9	-26.0
	SLM	-5.7	-20.6	-8.4	-17.4
	ECM	-20.3	-36.1	-21.9	-37.9
Bass1	*PIEZO*	-18.6	-26.6	-17.2	-27.8
	SLM	-11.4	-24.0	-9.7	-24.7
	ECM	-22.6	-32.9	-23.3	-32.5
Bass2	*PIEZO*	-11.7	-29.7	-11.4	-29.8
	SLM	-4.2	-21.2	-4.0	-21.5
	ECM	-34.8	-39.0	-28.8	-36.6
Bass3	*PIEZO*	-21.6	-25.9	-18.1	-28.9
	SLM	-19.5	-23.8	-14.2	-24.3
	ECM	-31.1	-37.6	-32.3	-35.9
Bass4	*PIEZO*	-19.6	-24.2	-21.1	-26.9
	SLM	-14.5	-21.2	-15.8	-22.0

Table 3.12. *Results for untrained singers. - ECM, PIEZO, SLM.*

		Tanti Auguri		*Inno di Mameli*	
		ΔB_{low}	ΔB_{high}	ΔB_{low}	ΔB_{high}
Sop1	*ECM*	-17.2	-24.2	-19.9	-29.2
	PIEZO	-7.2	-9.5	-9.5	-14.3
	SLM	-6.3	-10.2	-5.1	-10.9
Sop2	*ECM*	-18.5	-34.0	-18.8	-34.3
	PIEZO	-13.4	-21.9	-13.5	-19.7
	SLM	-9.1	-14.3	-9.1	-14.2
Sop3	*ECM*	-18.3	-28.4	-19.5	-32.4
	PIEZO	-10.3	-14.7	-10.9	-17.5
	SLM	-7.1	-12.9	-7.7	-16.4
Alto1	*ECM*	-16.31	-29.0	-19.1	-31.4
	PIEZO	-12.2	-24.4	-12.6	-27.4
	SLM	-3.7	-12.7	-3.8	-13.9
Tenor1	*ECM*	-17.9	-36.3	-16.1	-37.4
	PIEZO	-13.9	-33.7	-14.2	-35.2
	SLM	-8.6	-14.5	-8.3	-17.0
Tenor2	*ECM*	-25.5	-34.6	-25.1	-35.2
	PIEZO	-14.3	-25.2	-12.0	-27.0
	SLM	-9.4	-18.4	-7.3	-18.1
Overall Mean	*ECM*	-22.6 ± 5.8	-33.9 ± 4.6	-22.8 ± 4.4	-34.4 ± 2.7
	PIEZO	-14.4 ± 3.9	-23.6 ± 6.5	-14.3 ± 3.4	-25.3 ± 5.5
Mean F.	*ECM*	-17.6 ± 1.0	-28.9 ± 4.0	-19.3 ± 0.5	-31.8 ± 2.1
	PIEZO	-10.8 ± 2.7	-17.6 ± 6.8	-11.6 ± 1.8	-19.7 ± 5.6
Mean M.	*ECM*	-24.9 ± 5.7	-36.1 ± 2.9	-24.3 ± 4.5	-35.6 ± 2.0
	PIEZO	-15.9 ± 3.3	-26.2 ± 4.6	-15.5 ± 3.3	-27.8 ± 3.3

Table 3.13. *Results for untrained singers. - ECM, PIEZO, SLM.*

Here, are avaiable the results for all of the professional singers.

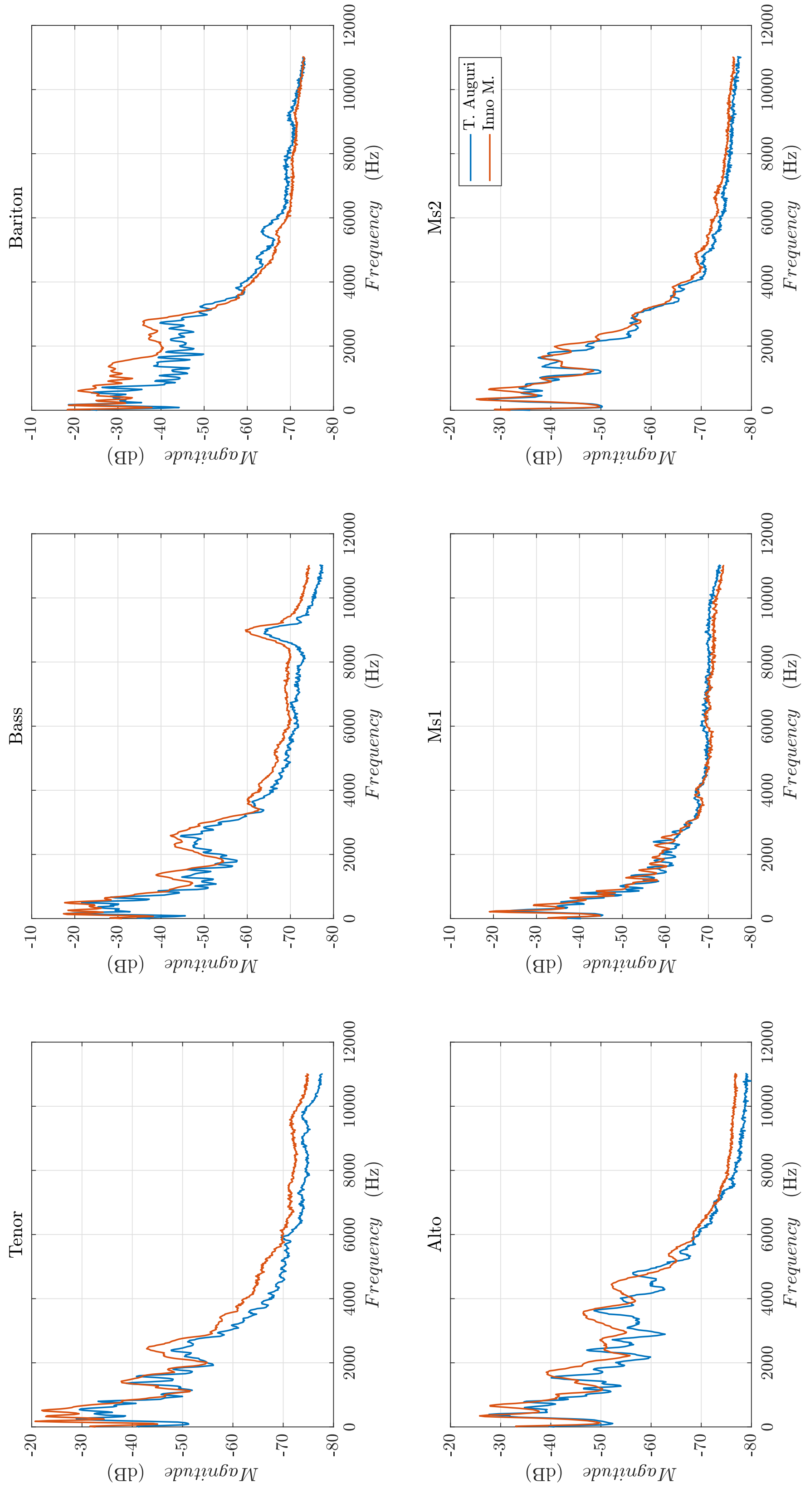


Figure 3.38. *LTAS of "Inno M." and "Tanti auguri" song - ECM.*

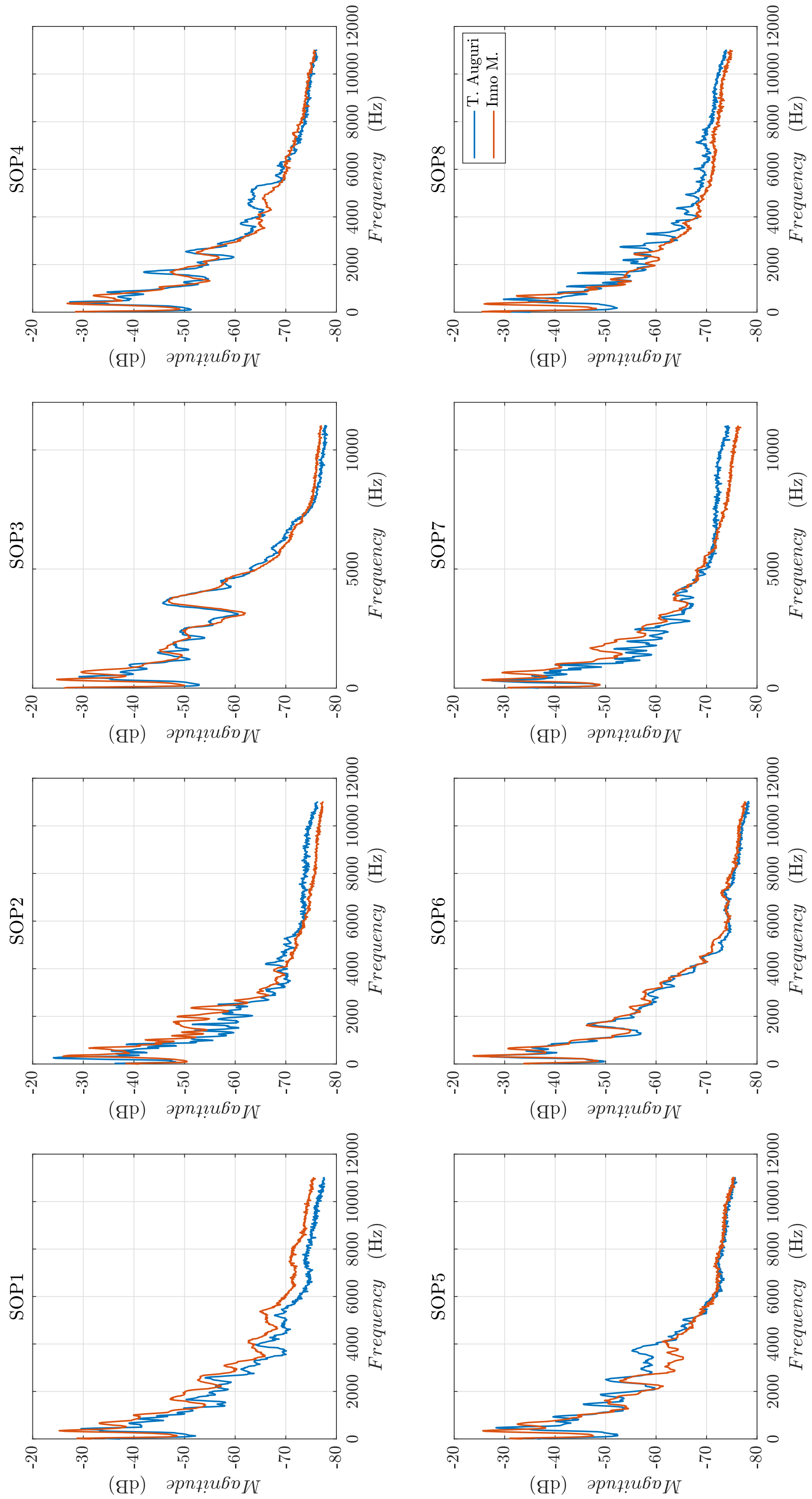


Figure 3.39. *LTAS of "Inno M." and "Tanti auguri" song - ECM.*

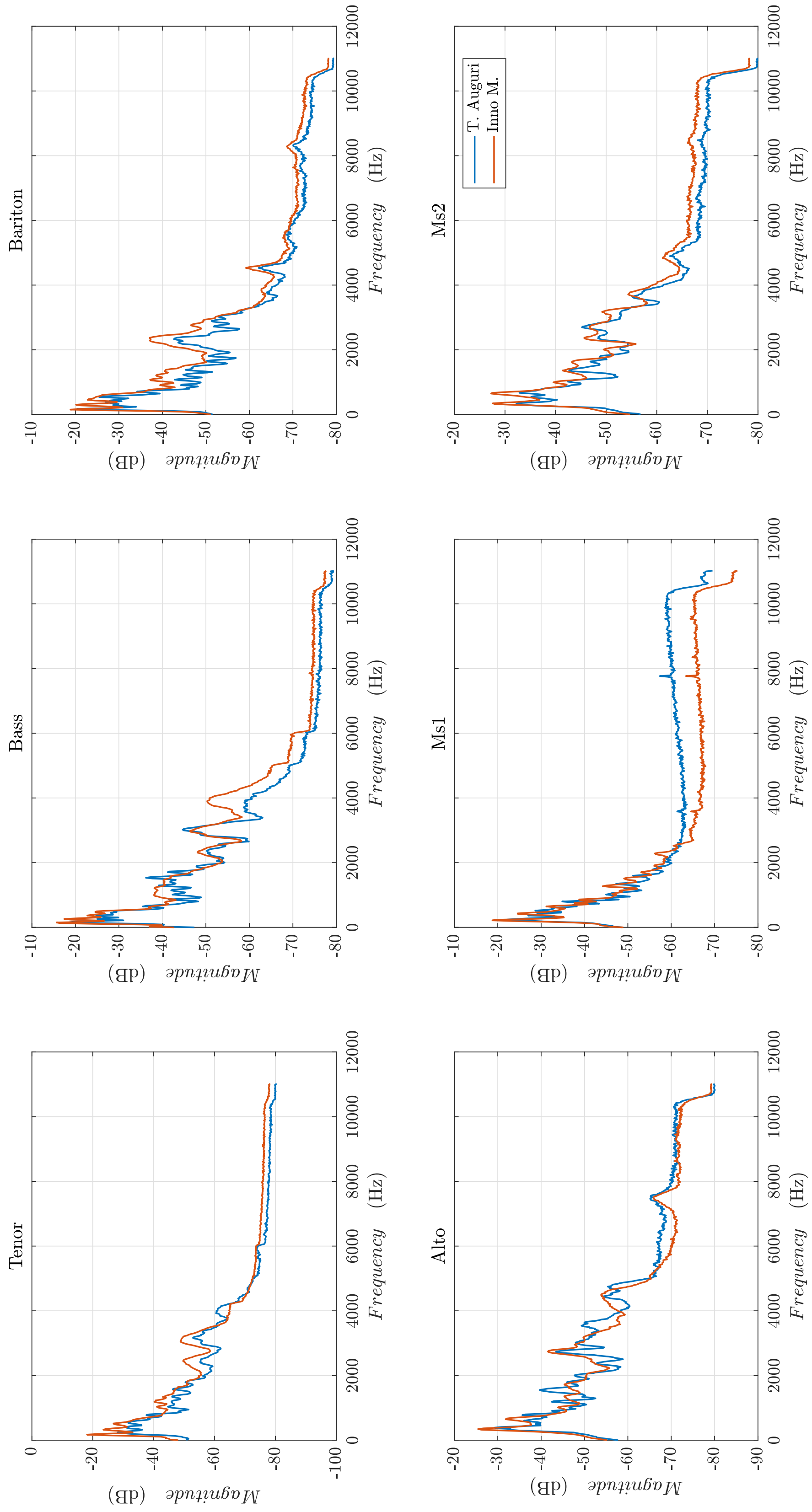


Figure 3.40. *LTAS of "Inno M." and "Tanti auguri" song - PIEZO.*

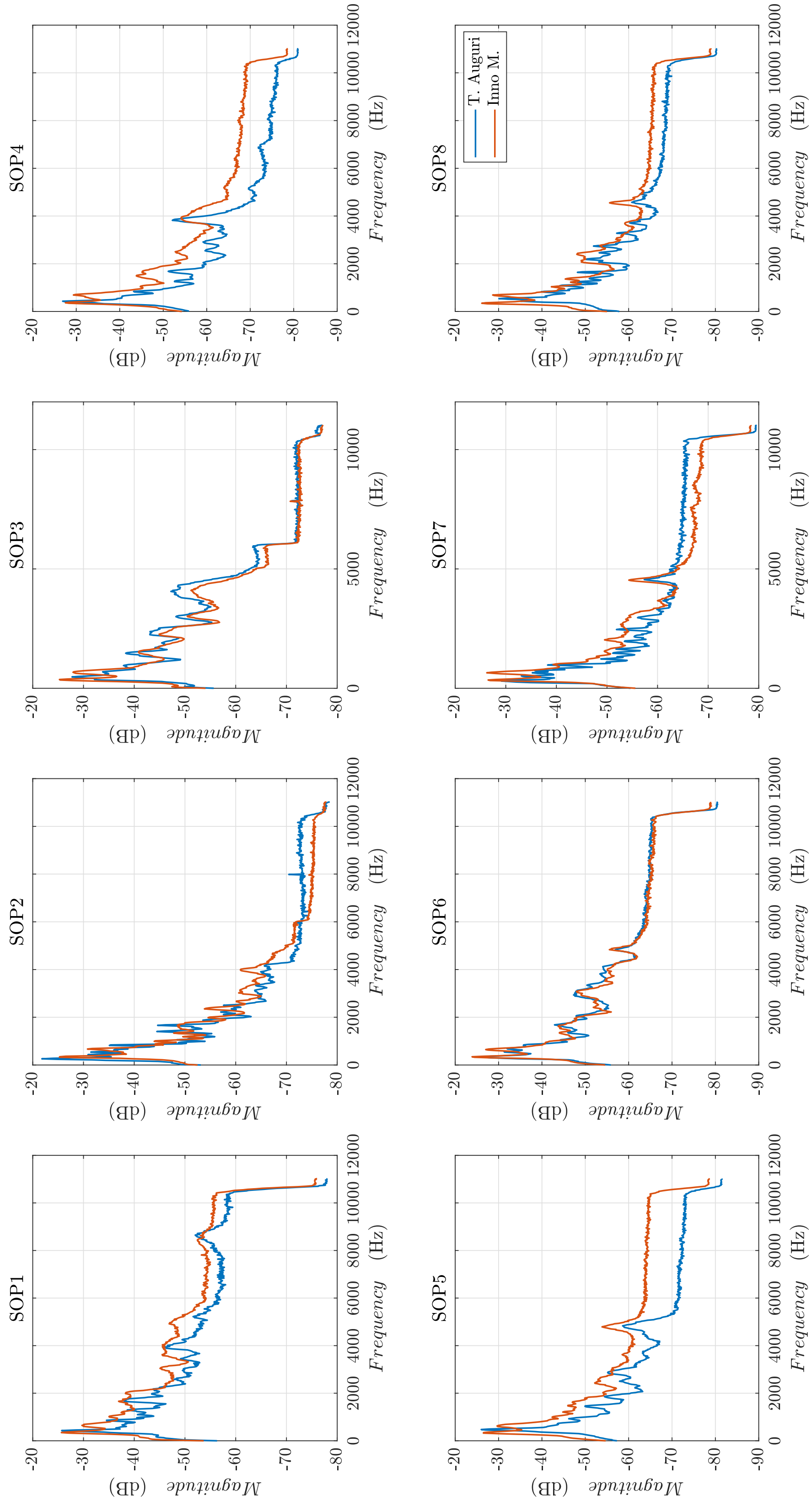


Figure 3.41. *LTAS of "Inno M." and "Tanti auguri" song - PIEZO.*

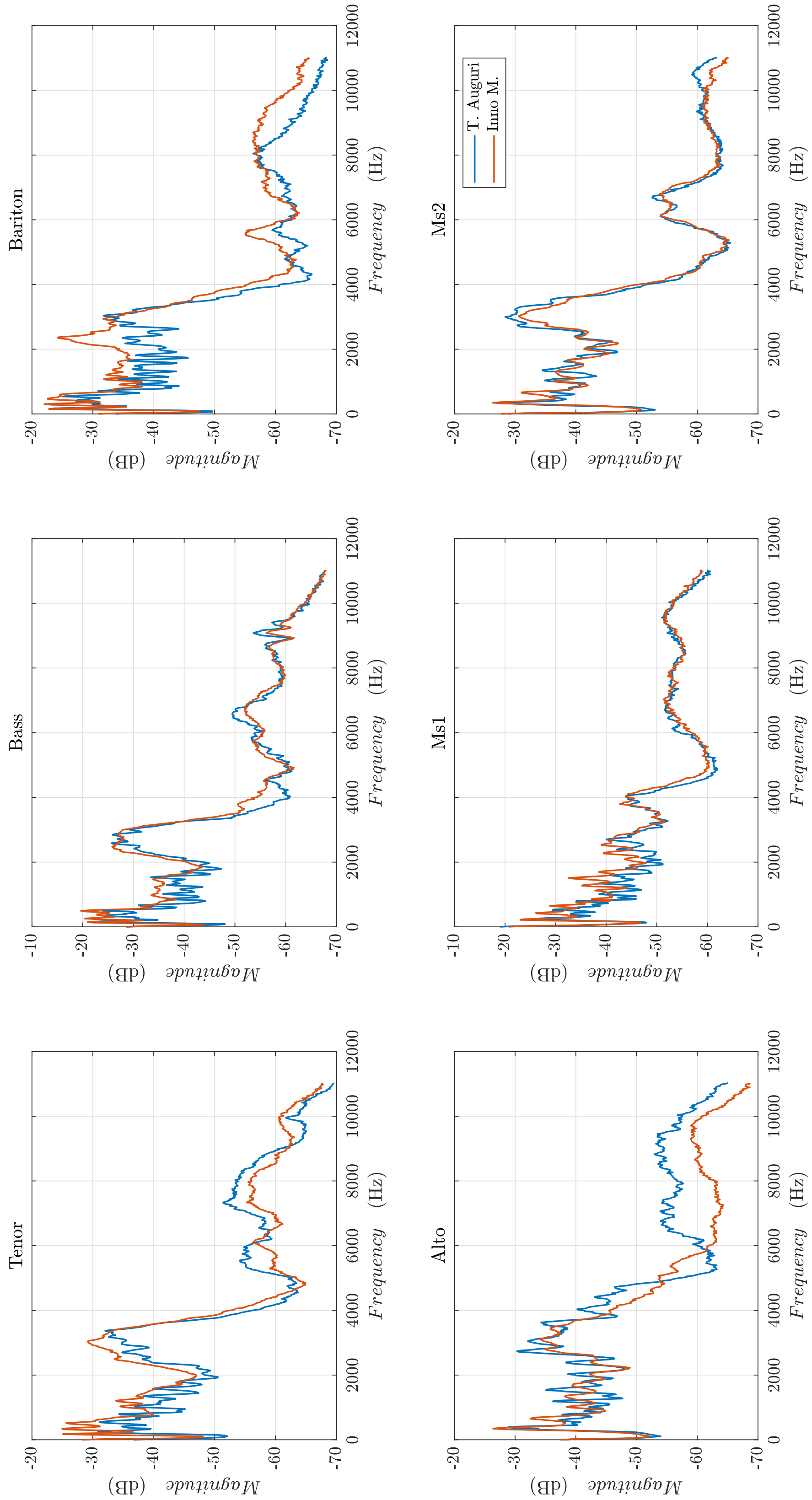


Figure 3.42. *LTAS of "Inno M." and "Tanti auguri" song - SLM.*

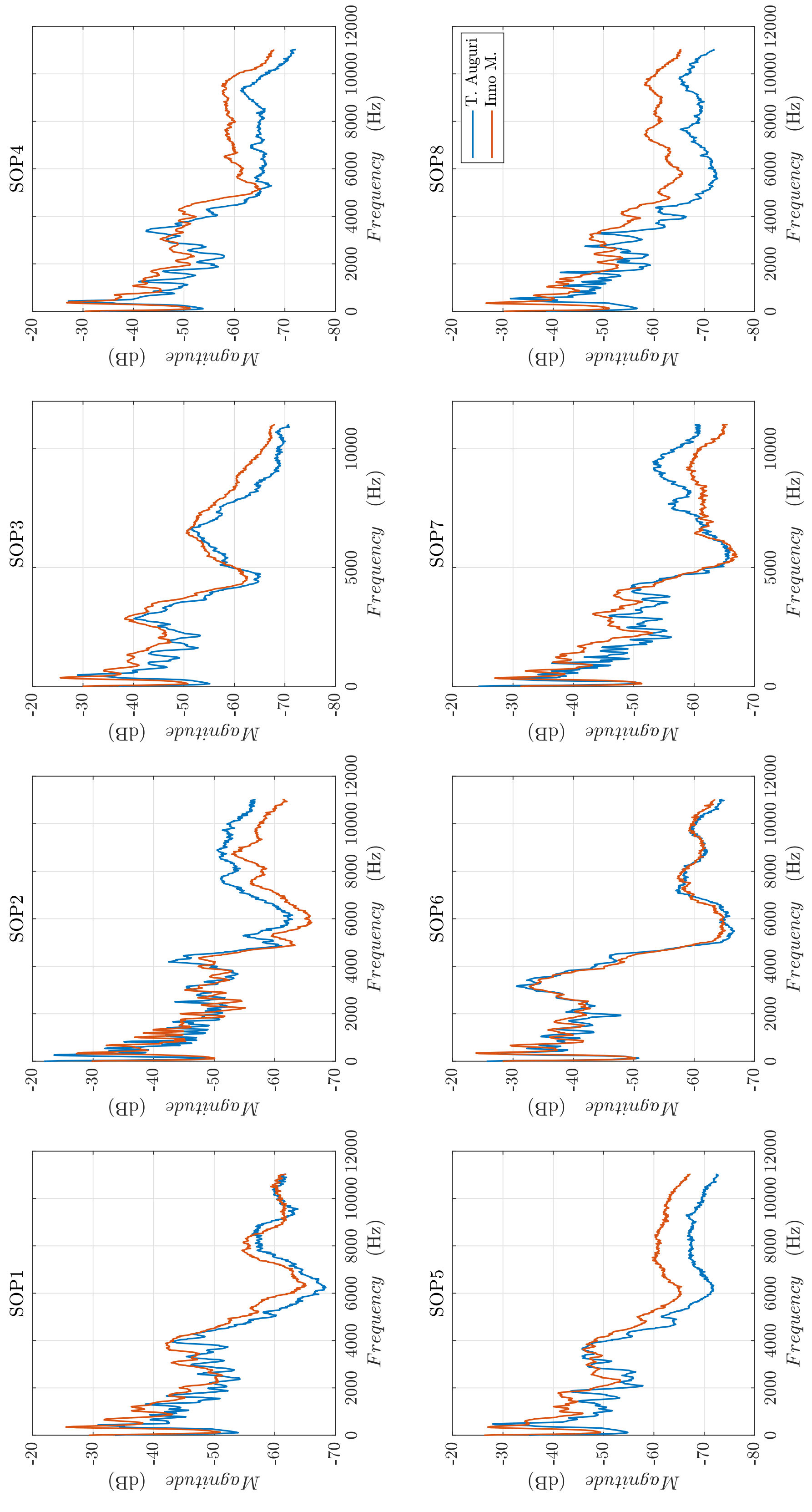


Figure 3.43. *LTAS of "Inno M." and "Tanti auguri" song - SLM.*

Following are avaiable some example graphs to compare the LTAS results for trained and untrained singers. The comparison was carried out between subjects of the same register.

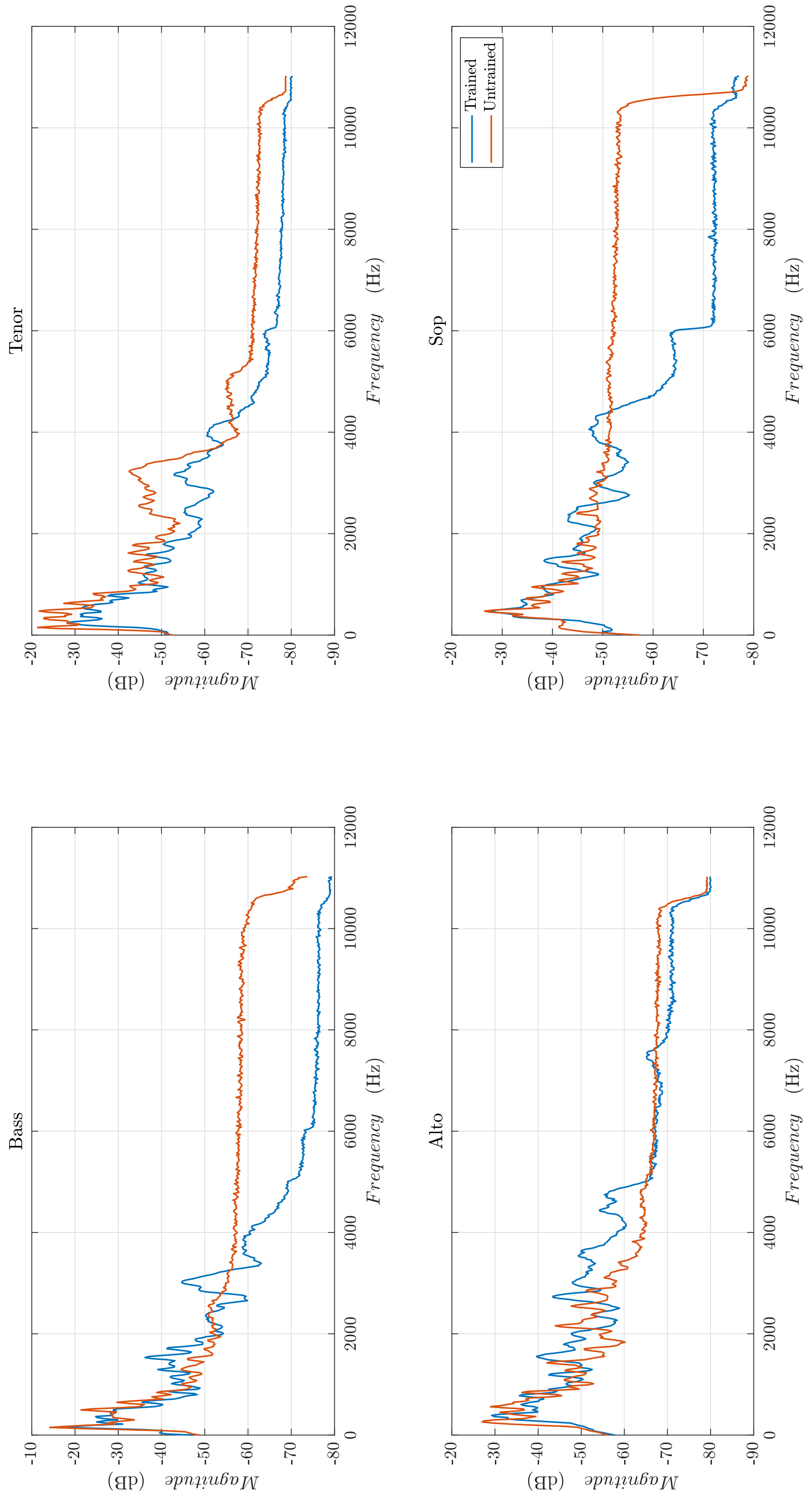


Figure 3.44. *Comparison between trained and untrained singers - "Tanti auguri" song - ECM.*

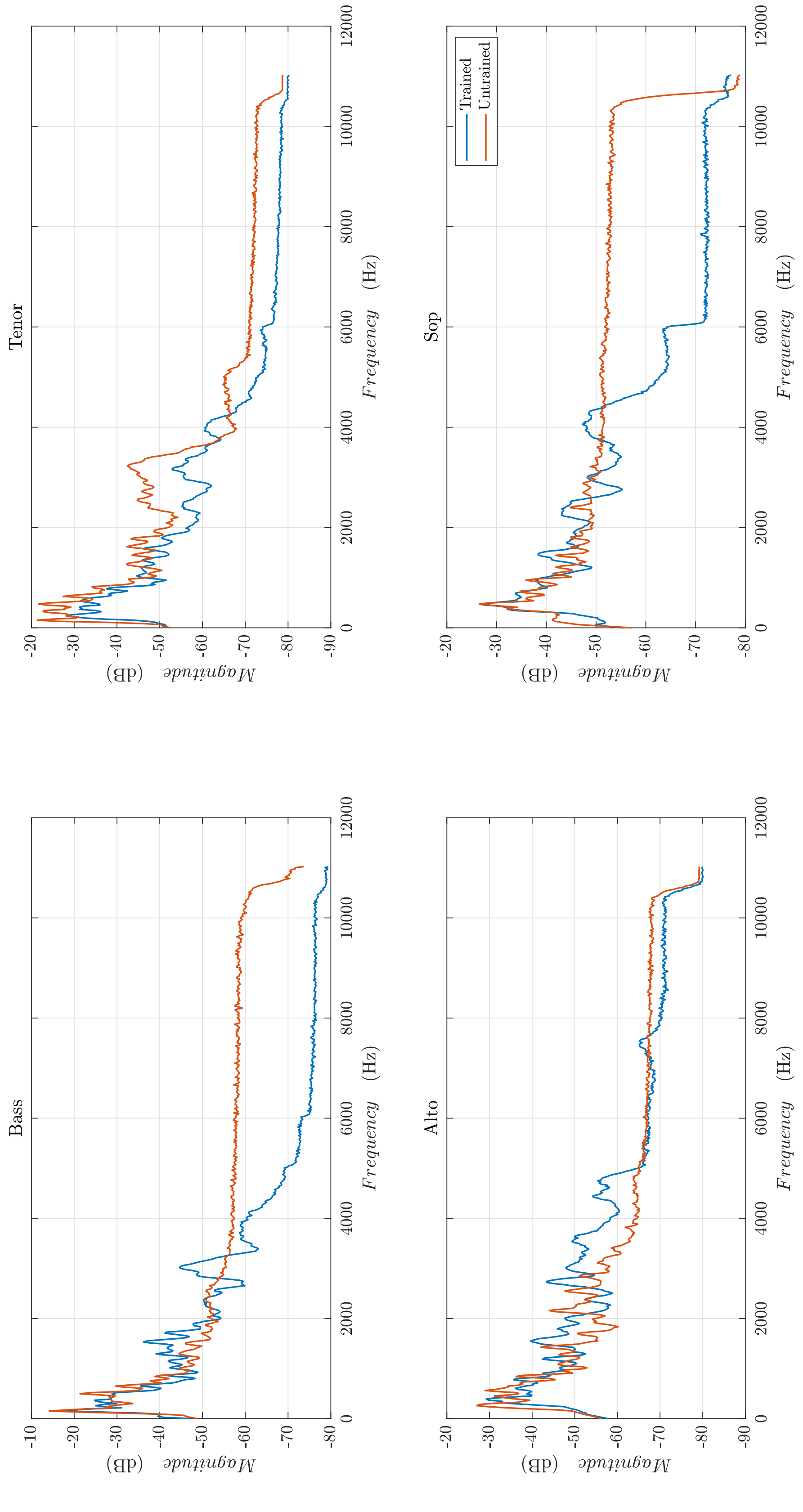


Figure 3.45. *Comparison between trained and untrained singers - "Tanti auguri" song - PIEZO.*

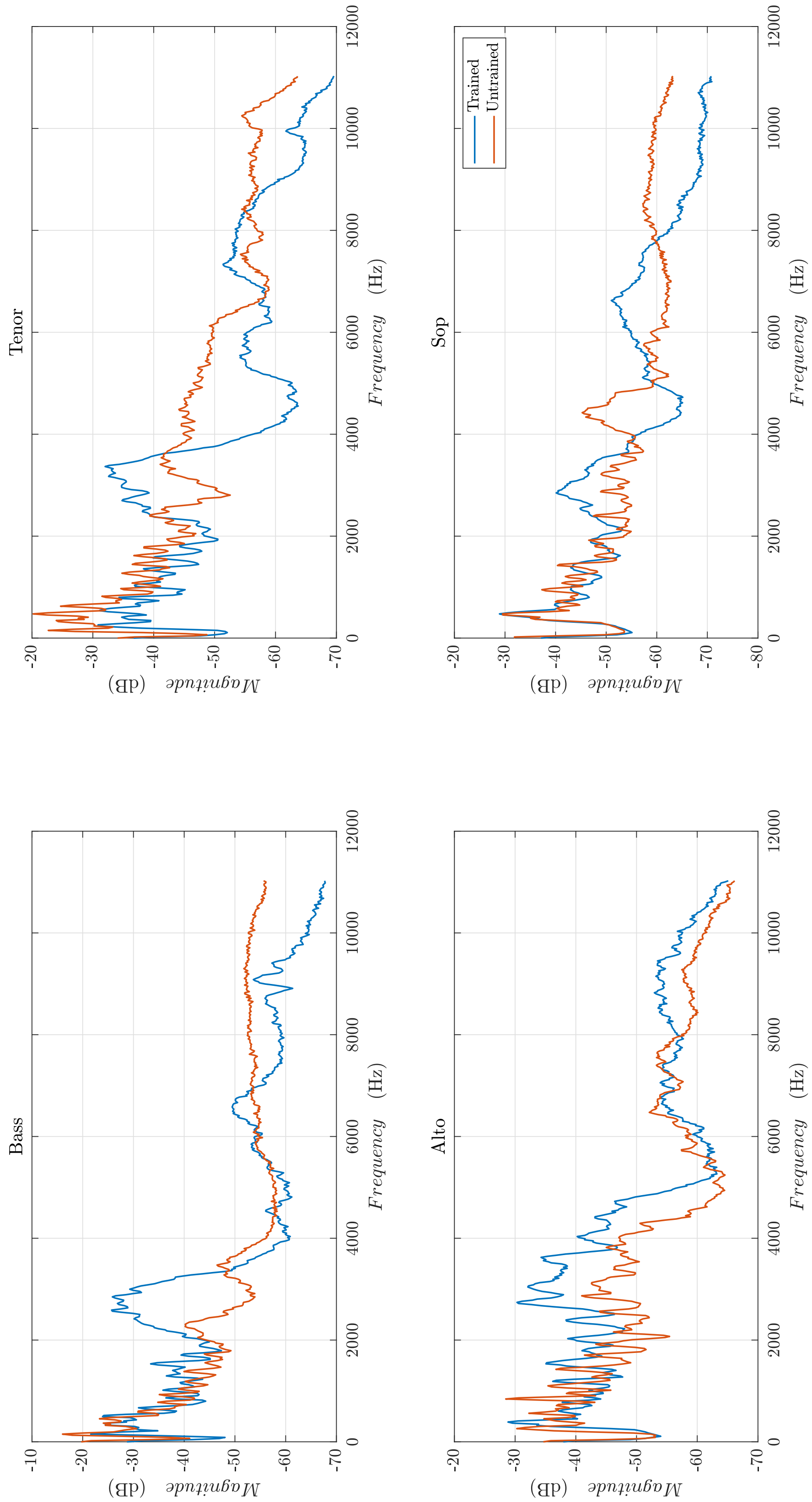


Figure 3.46. *Comparison between trained and untrained singers - "Tanti auguri" song - SLM.*

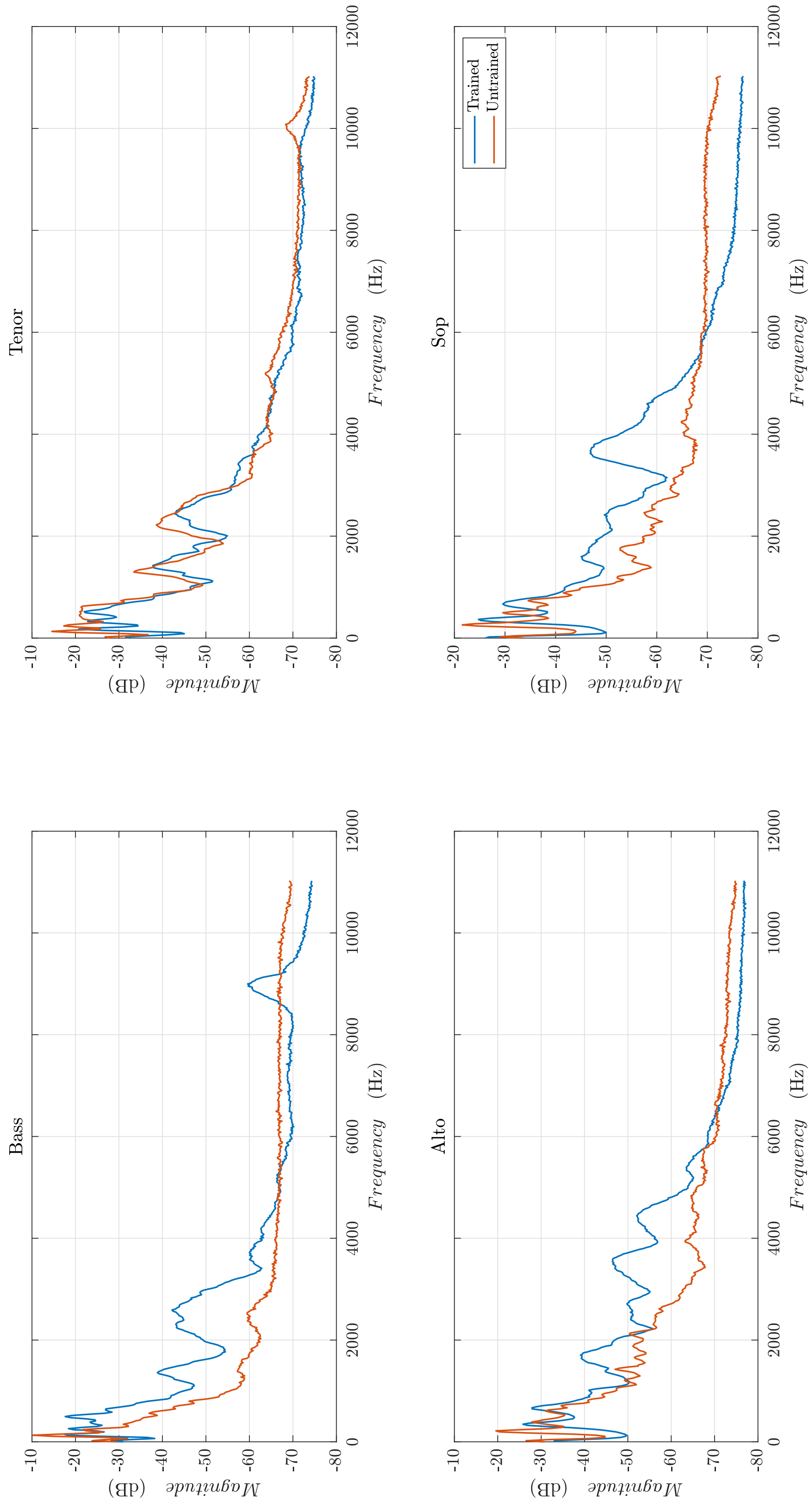


Figure 3.47. *Comparison between trained and untrained singers - "Inno M." song - ECM.*

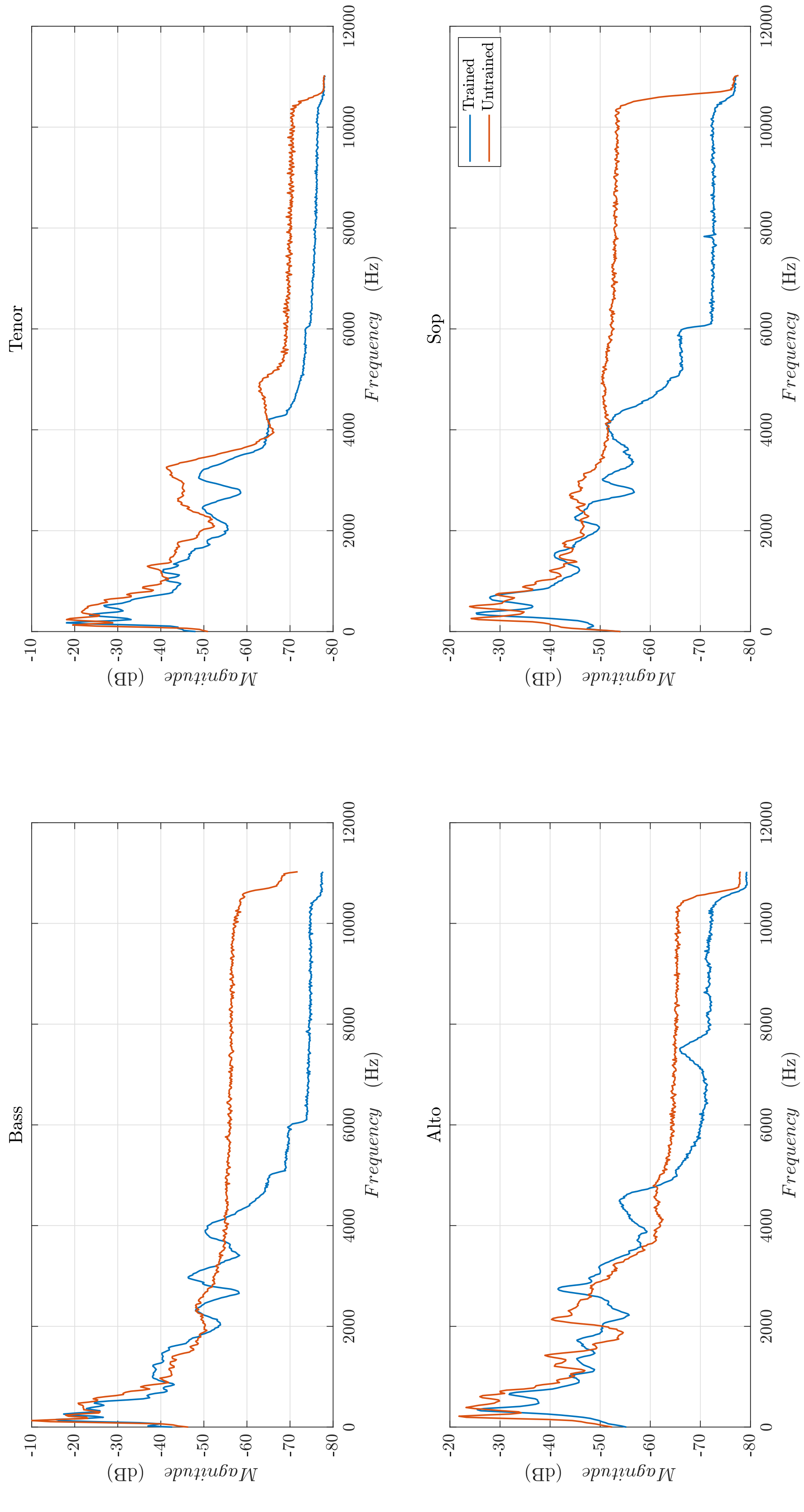


Figure 3.48. *Comparison between trained and untrained singers - "Inno M." song - PIEZO.*

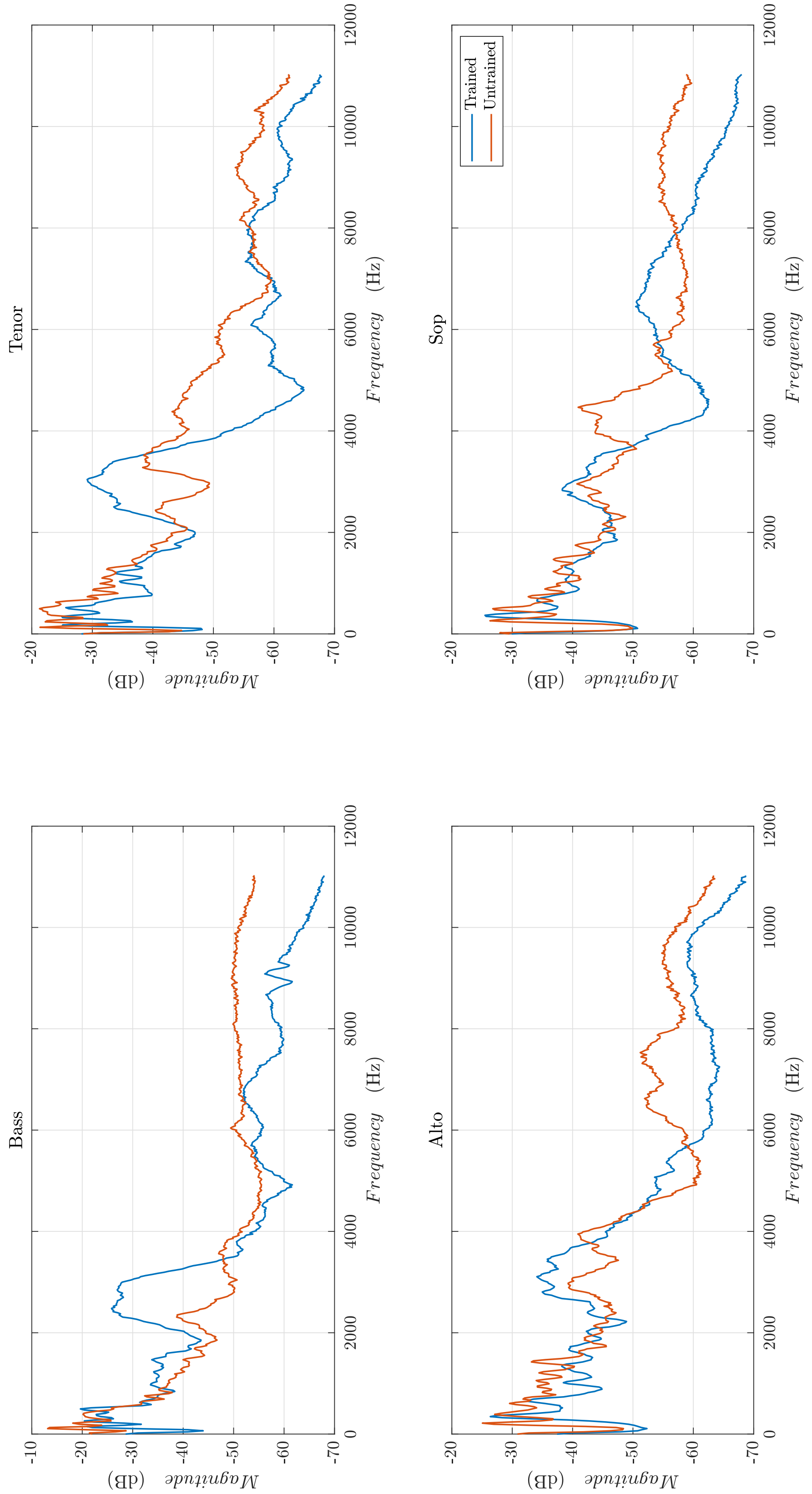


Figure 3.49. *Comparison between trained and untrained singers - "Inno M." song - inno SLM.*

Here there are the results of the comparison between the ranges from trained and untrained singers:

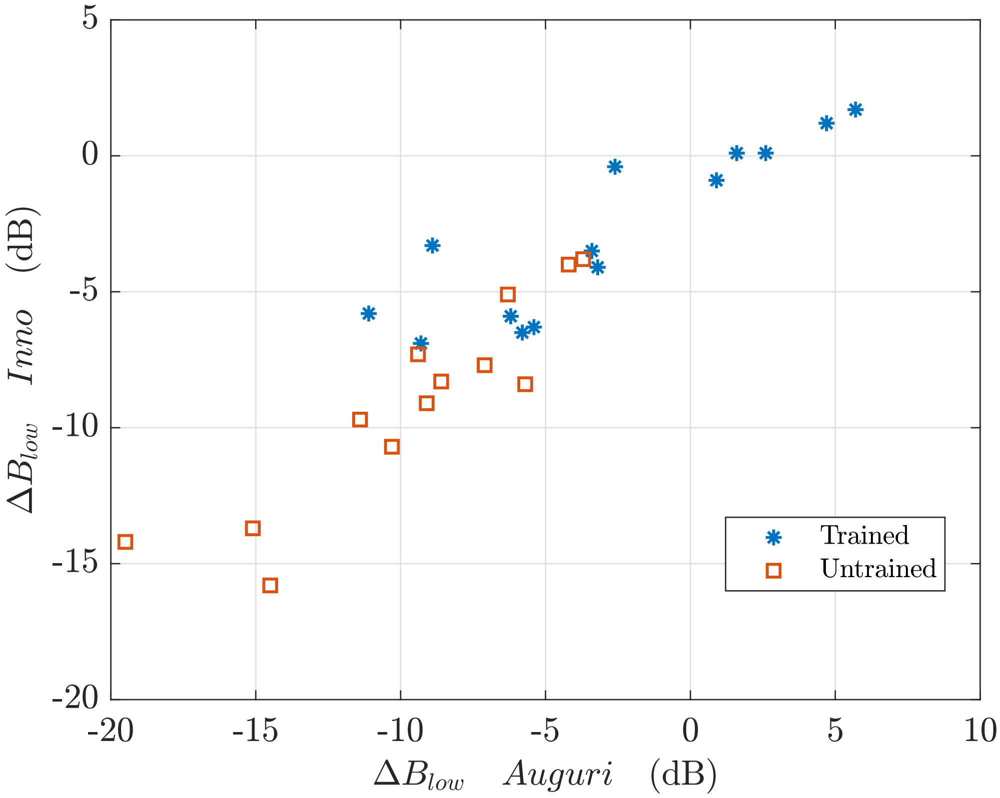


Figure 3.50. *Comparison between* Δ B_{low}.

The Pearson Coefficient was calculated for this value of ΔB_{low} of these two types of songs: 0.92.
The Pearson Coefficient was also calculated for this value of ΔB_{high} of these two types of songs: 0.84. This means that there is an higher correlation for the lower frenquency range of information of two different songs instead of the the higher frequency range.

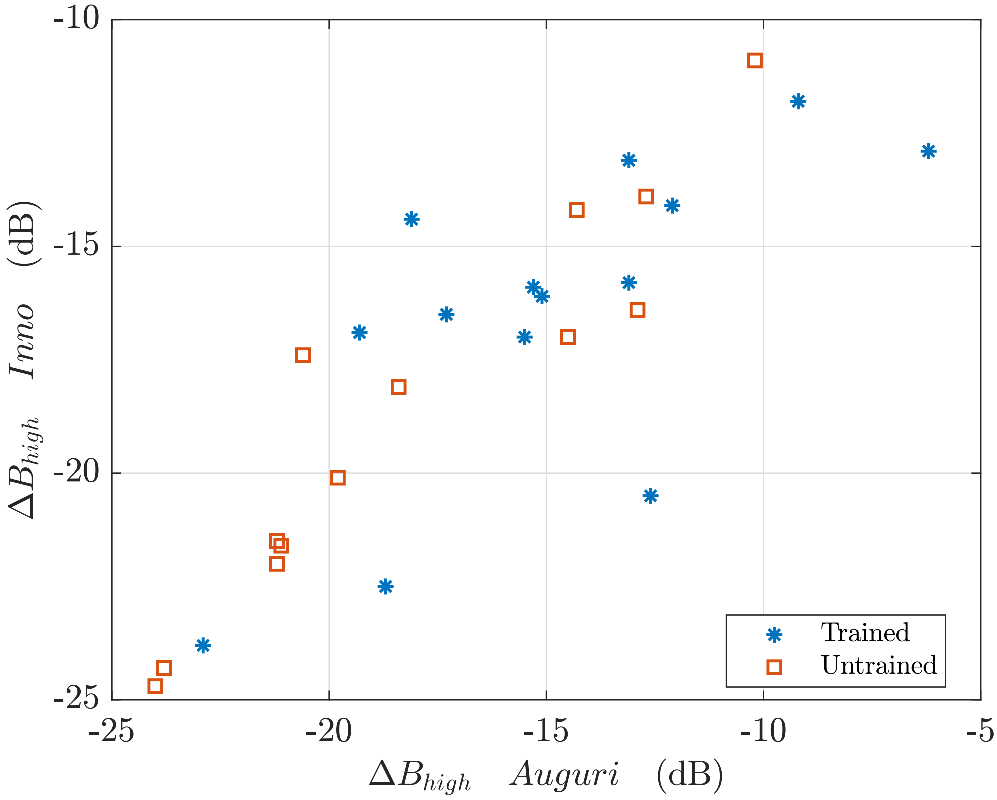


Figure 3.51. *Comparison between* Δ B_{high}.

There is also the comparison between the microphones. The Pearson Coefficient is 0.81 for the comparison between the ΔB_{low} between ECM and SLM; the same coefficient is 0.68 for the comparison between the ΔB_{high} between the same microphones.

The same analysis was conducted for the PIEZO microphone. The Pearson Coefficient is 0.67 for the comparison between the ΔB_{low} between PIEZO and SLM; the same coefficient is 0.56 for the comparison between the ΔB_{high} between the same microphones.

In general, the correlation is higher for the ECM microphone instead of the PIEZO one.

At the end of this section, the Figure 3.54 and Figure 3.55 explain how the devices discriminate the trained from untrained singers. In the Figure 3.54 and in the Figure 3.55 there are the evaluation of the discrimination between trained and untrained singers for all of microphones: the best results were obtained for the ΔB_{low} for ECM microphone. There is a better gap into the graph of ECM instead of the other microphones. For the evaluation of ΔB_{high}, none of the microphones are good discriminator.

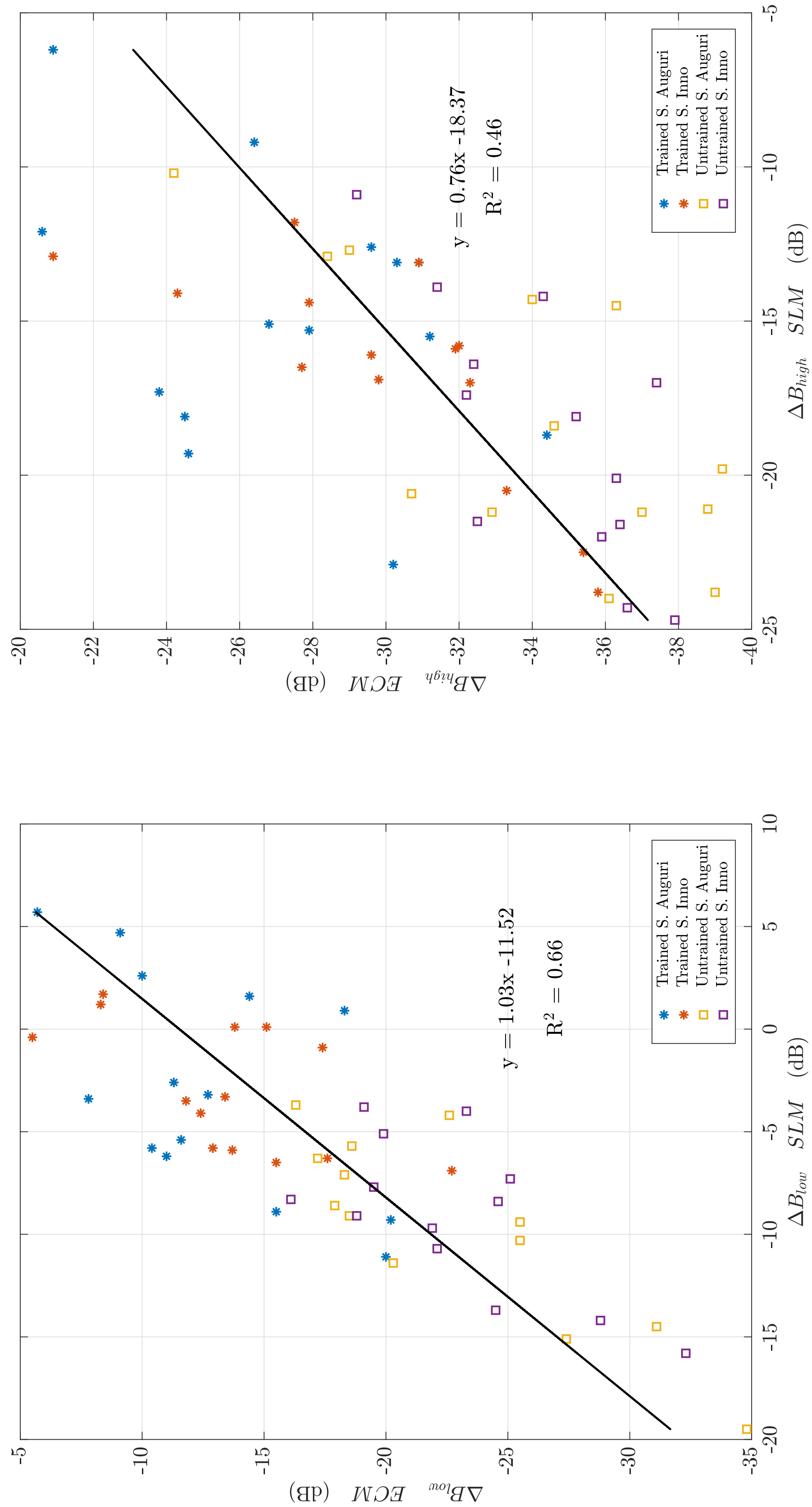
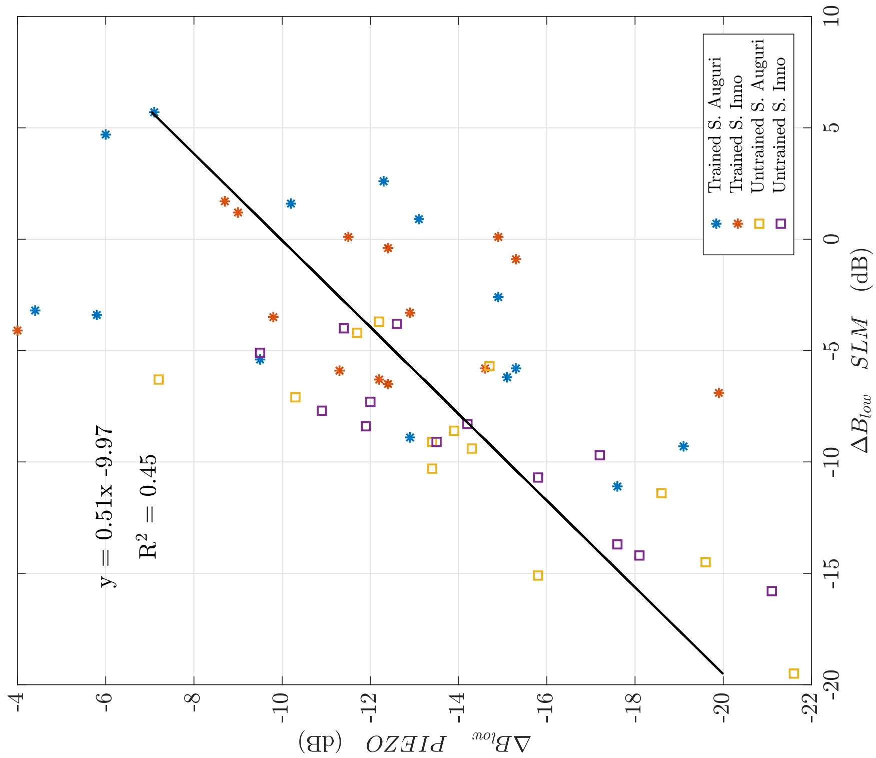


Figure 3.52. *Comparison between the ΔB_{low} and ΔB_{high} for ECM and SLM results.*



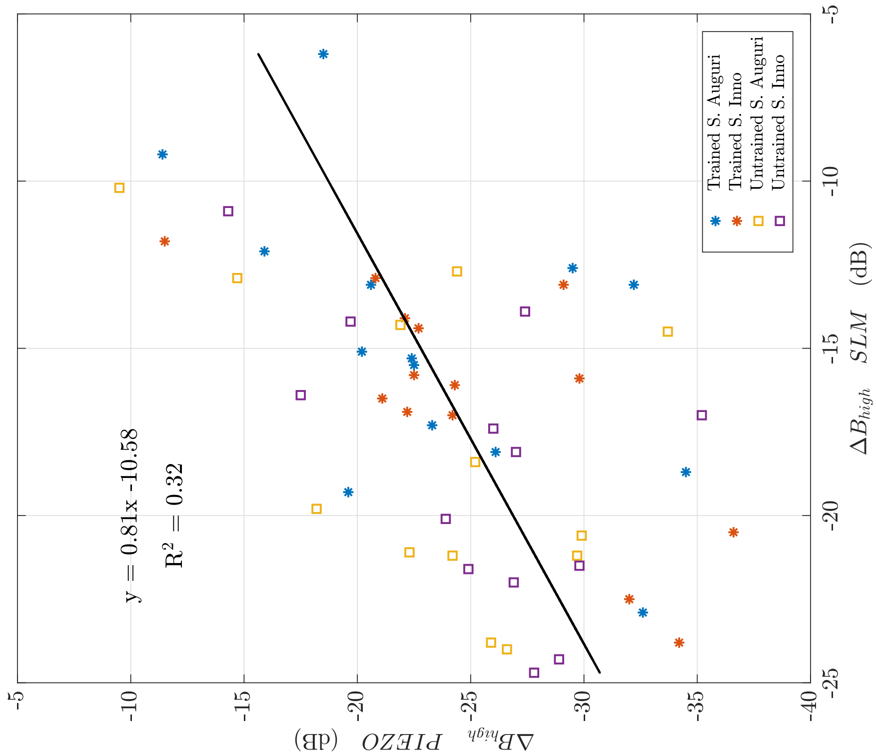


Figure 3.53. *Comparison between the ΔB_{low} and ΔB_{high} for PIEZO and SLM results.*

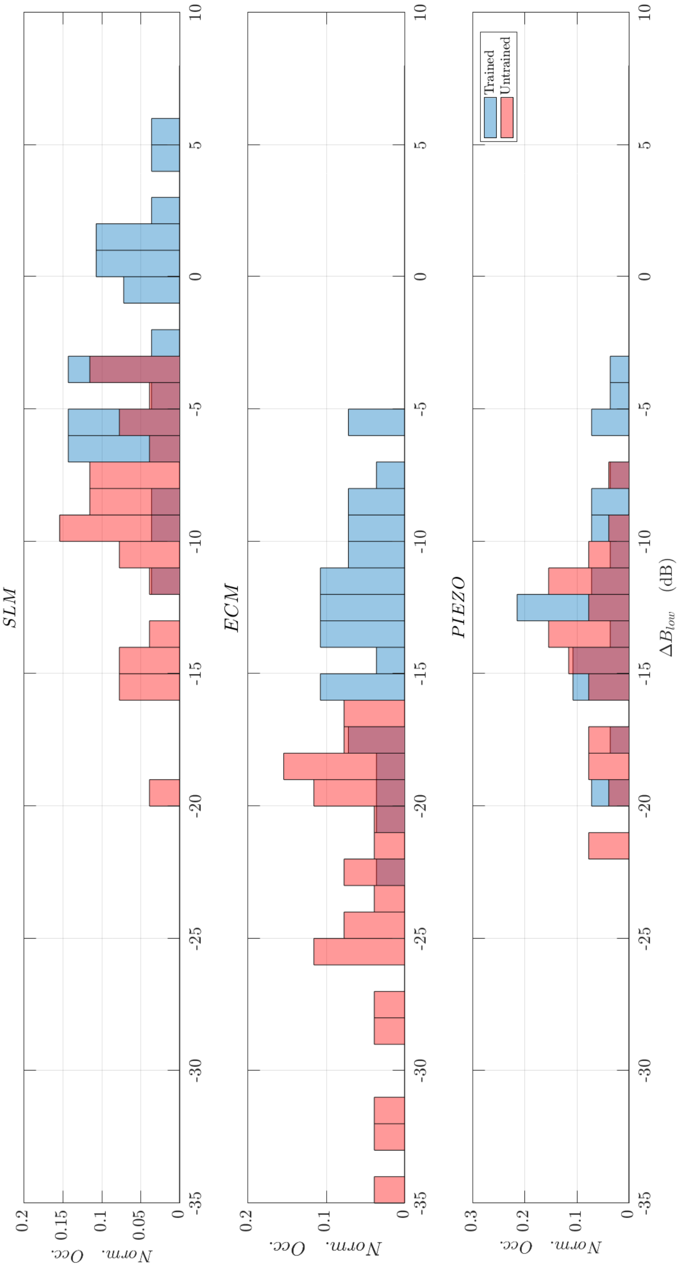


Figure 3.54. *Summary of results about Δ B_{low} for the used devices.*

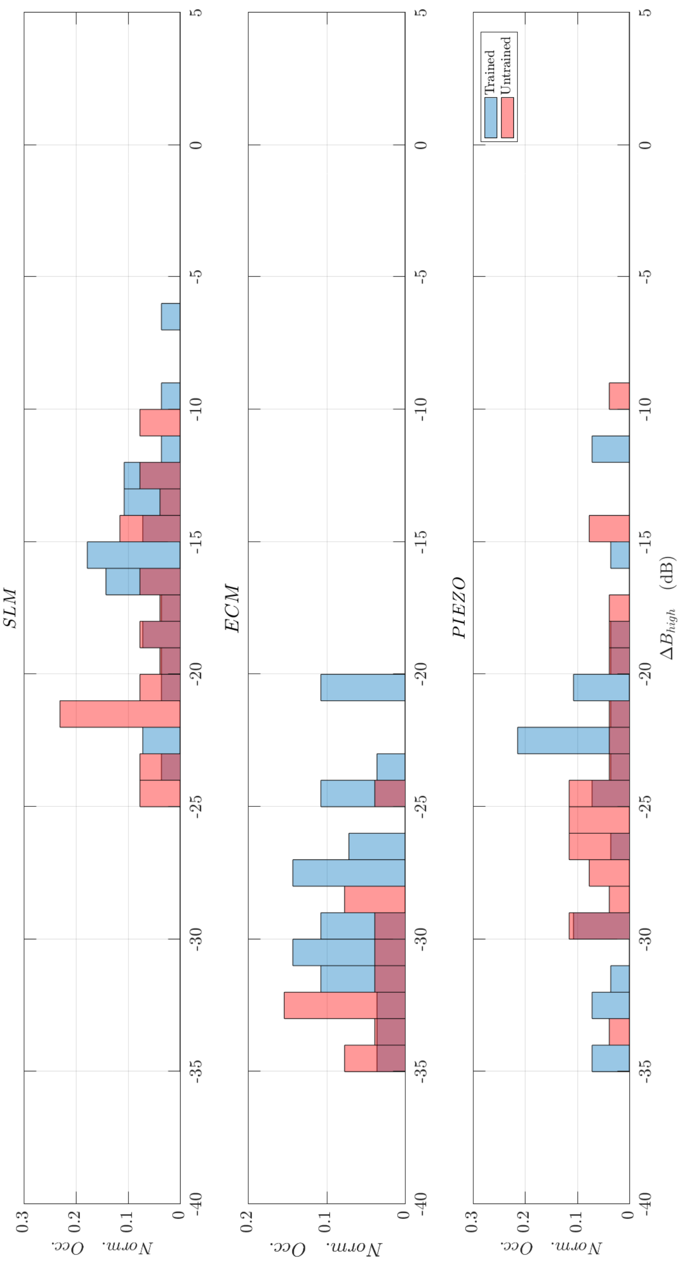


Figure 3.55. *Summary of results about Δ B_{high} for the used devices.*

3.4 Singing Power Ratio

In this section, the results about SPR of the arpeggio of an octave are summarized in the Table 3.14 for trained singers and in the Table 3.15 for untrained singers. These tables contain the results of both microphones.

	ECM	***SLM***
Bass	-35.6 ± 4.2	-17.5 ± 6.5
Bariton	-30.3 ± 4.5	-17.0 ± 5.4
Tenor	-38.6 ± 7.4	-21.5 ± 5.2
Alto	-31.2 ± 4.8	-12.7 ± 3.3
Ms1	-46.2 ± 8.7	-20.1 ± 6.3
Ms2	-37.2 ± 7.1	-14.1 ± 9.5
Sop1	-37.1 ± 6.0	-22.9 ± 5.1
Sop2	-39.9 ± 10.3	-27.9 ± 8.0
Sop3	-36.8 ± 6.7	-20.3 ± 4.7
Sop4	-38.3 ± 7.5	-22.3 ± 4.6
Sop5	-35.9 ± 13.5	-25.2 ± 9.4
Sop6	-43.9 ± 6.4	-15.9 ± 4.6
Sop7	-51.7 ± 8.9	-29.3 ± 7.5
Sop8	-52.2 ± 5.6	-30.1 ± 2.9

Table 3.14. *Mean and standard deviation for trained singers.*

	ECM	***SLM***
Bass1	-49.7 ± 5.9	-28.9 ± 6.1
Bass2	-50.7 ±3.3	-23.2 ± 1.8
Bass3	-53.6 ± 5.7	-29.2 ± 4.6
Tenor1	-49.6 ± 10.5	-23.7 ± 5.9
Tenor2	-38.2 ± 6.3	-19.6 ± 4.3
Tenor3	-35.5 ± 5.9	-18.0 ± 4.6
Alto1	-47.9 ± 5.8	-31.3 ± 3.6
Alto2	-47.7 ± 2.9	-25.1 ± 2.4
Alto3	-49.8 ± 4.7	-24.4 ± -3.0
Sop1	-44.4 ± 1.8	-26.3 ± 7.2
Sop2	-45.7 ± 3.7	-23.9 ± 4.5
Sop3	-44.2 ± 2.8	-26.3 ± 2.0

Table 3.15. *Mean and standard deviation for untrained singers.*

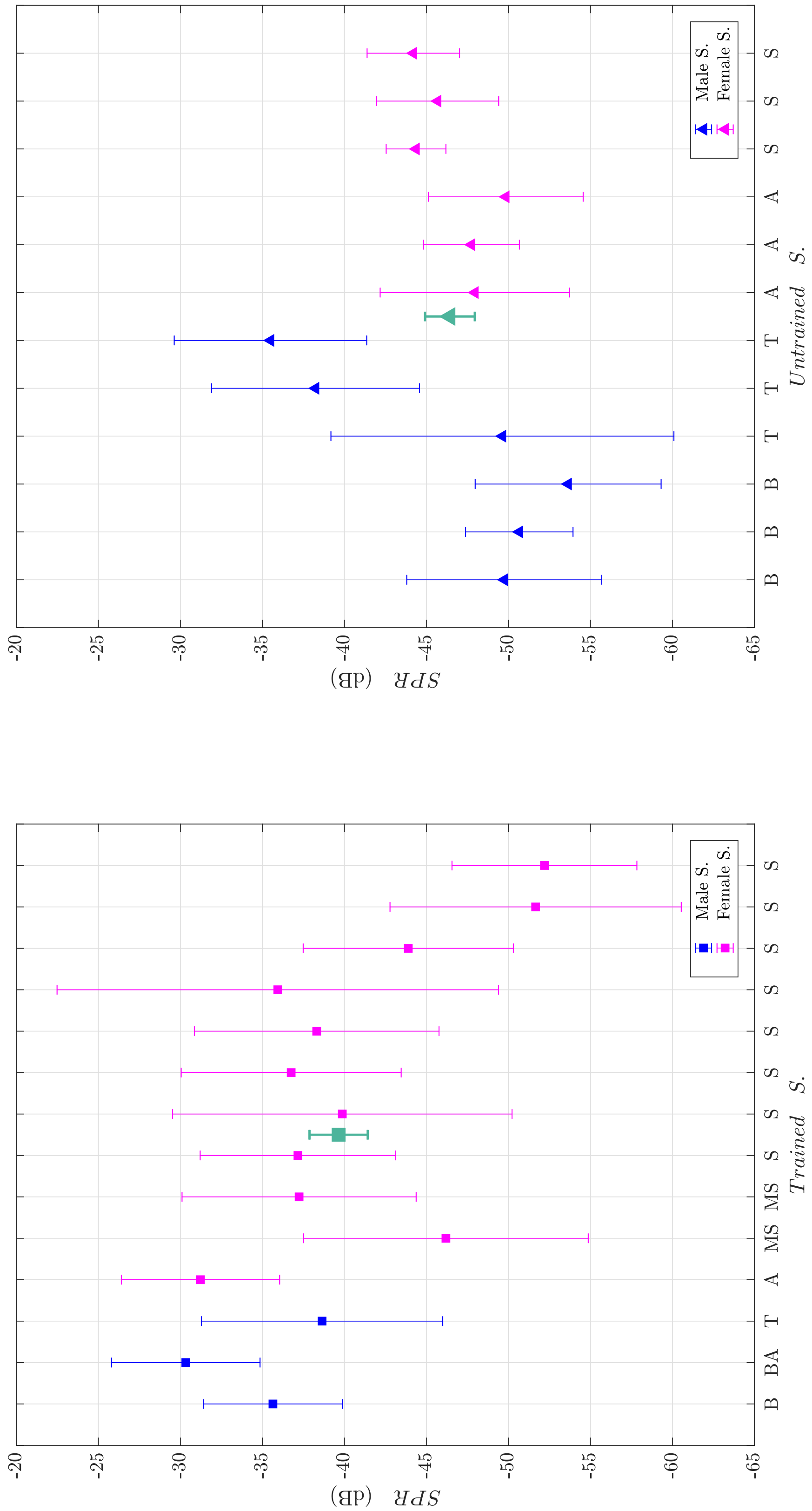


Figure 3.56. *SPR comparison between trained and untrained singers - ECM*

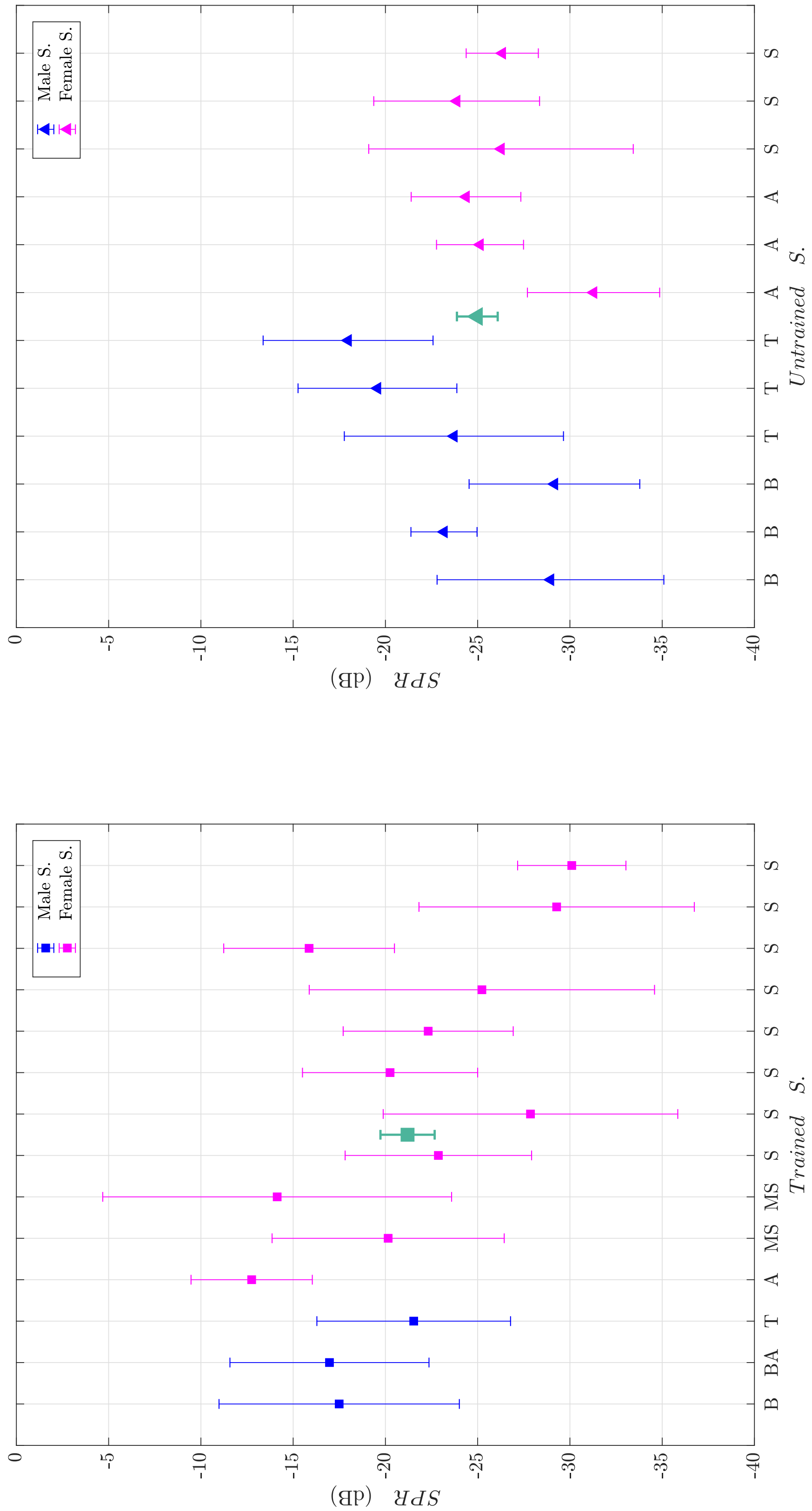


Figure 3.57. *SPR comparison between trained and untrained singers - SLM*

Following, there is the Table 3.16 with the codifing of the letters used into the Figure.

Letters	Definition
B	Bass
BA	Bariton
T	Tenor
A	Alto
MS	Mezzo-Soprano
S	Soprano

Table 3.16. *Codification.*

Here is avaiable the overall means of SPR of singers.

		ECM	*SLM*
Trained	*Overall Mean*	-39.6 ± 6.6	-21.2 ± 5.5
	Mean F.	-40.9 ± 6.7	-21.9 ± 6.0
	Mean M.	-34.8 ± 4.2	-18.7 ± 2.5
Untrained	*Overall Mean*	-46.4 ± 5.3	-25.0 ± 3.8
	Mean F.	-46.6 ± 2.2	-26.2 ± 2.7
	Mean M.	-46.2 ± 7.5	-23.8 ± 4.6

Table 3.17. *Mean and standard error for two groups of singers.*

The results are higher for trained singers: this means that the quality of their voice is better than the untrained singers' one. The gap between the results is higher for ECM microphone instead of the SLM one comparing the two categories of singers. This is probably due to the "preparation" of singer. A professional singer prepares the phonatory apparatus before sing: the information about this internal preparation are provided from ECM microphone due to the position of this microphone. The signal of SLM microhpone provides more general information: there are the formants due to the vocal tract and other information about the "internal" voice before get off it. The ECM microphone gives information of the previous part of the vocal tract.

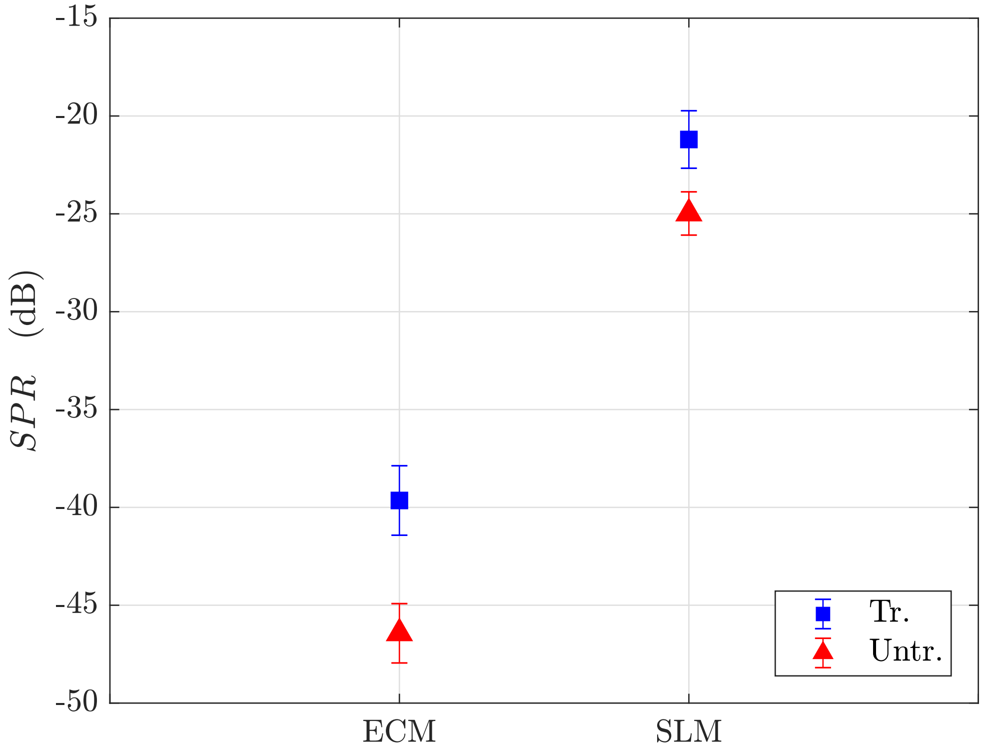


Figure 3.58. *Overall mean - comparison.*

Chapter 4

Conclusions

The starting point of this study, is to evaluate the quality of the singers' voice. Have a good voice is an important element for the healty aspects and, for singers, to have success during an exhibition. To pursue this aim, a study in the domain of frequency of the vocal signal is carried out. This study was conducted for the same signal obtained with three different microphones. The goal of this study is to obtain a new objective parameters to classify the voice of singers analyzing the vocal signal in a not invasive way. Firstly it was evaluated the mean of the values obtained from the contact microphones during the arpeggio compared to the microphone in air: in both cases the results are comparable. So, this suggests that these microphones are good devices to detected the fundamental frequency of the note sung by singer. Then, the evaluation of the "Hysteresis of the singer" show how the singers are able to sing the same note in a different semi-phrase direction. If the parameter is closed to zero, the singer is less influenced by the type of arpeggio. This parameter is evaluated only for the signal obtained from the microphone in air because this microphone rapresents the "audience" during an exhibition; for a singer is essential do a very good performance in front of the audience. Therefore, this parameter indicates a numeric value of the great ability of the singer during an exhibition. It is possible use this parameter to evaluate the same task performed in different moments: this parameter can help to monitor the improvements of the singer in a particular exercises or tasks. After that, the goodness of the contact microphones is evaluated. In every situation, the results of the contact microphone was closed to the microphone in air's one. This means that contact microphone is a good alternative to the microphone in air. The contact microphones are not influenced by the external noise and this allows to monitor different subjects in a both quiet and noise space compared to the microphone in air that allows to monitor subjects in a very quiet spaces (so, the anechoic chamber). In a second time the distribution of the CPPS was evaluated. A qualitative analysis of CPPS was conducted. The results showed that the shape of the distribution of CPPS is better for the analysis of the song tasks instead of the reading text for trained singers. This means that the voice of singers is cleaner during singing instead of reading. The analysis of the CPPS can be used not only with reading text but also for song to study the healty of the voice of singers. This aspect could be studied in the future to analyze the CPPS in a different way for professional singers and not only for normal people and untrained singers. The third part of this study, analyzed the spectrum of two different songs with different microphones. The results showed that the information in the low and high range of frequency are not depend by the song. It is possible examine in depth this aspect analyzing a more exended songs into the own range of singer. The ECM microphone is more correlated to the SLM instead of PIEZO, specially into the lower frequency range. The SPR analysis confirms that the untrained singers have higher value in modulus compared to the trained one; remember that this parameter is better if its values is closed to zero. The most interesting result is the ECM one. This difference, bigger than 5 dB between the two categories of singers, suggests that the "preparation" of the phonatory apparatus to sing is greater in the trained singer instead of the other one. More detailed studies can be conducted to confirm this theory: it is important to enlarge the database to make considerable this parameter also for contact microphone and not only for the microphone in air. This means that it will possibile to monitor different subjects at the same time because the contact microphones are not influenced by the external noise.

Appendix A

Insight of acoustic parameters

Cepstral Peak Prominence Smoothed

The word *Cepstrum* indicates the log-power spectrum of the log-power spectrum of a signal. This parameter is shown in the *Quefrency* domain. This new parameter is the anagram of *Spectrum* and its domain is the anagram of *Frequency*. These new terms are used to underline the difference between the classical spectrum in the frequency domain from this new parameter. The "Quefrency" seems coincide with the time domain but is not like this because there are two logs into the analysis. Into the Figure A.1 is shown the original signal and its spectrum. This spectrum shows a strong component corresponding to the regularity of harmonic peaks; the time at the cepstral peak coresponds to the fundamental period of the signal. The peak of this obtained signal rapresents the fundamental period and, its inverse, the fundamental frequency. This obtained spectrum is normalized for his maximum value. Here there is an example of the *Cepstral Peak Prominance* (CPP) of a signal in the frequency and quefrency domains. Often the *smoothed* CPP is analyzed, so this parameter is called *CPPS*: the modification consist in a smoothing of the cepstrum before calculate the distance between the peak and the regression line. This method was introducted by Hillebrand [1] in 1996 who discovered a good improvement in rediction accuracy introduced by filtering the original cepstrum. Two are the processes of the smoothing:

- Time-domain: frames cepstrum are averaged between the three frames before and the three frames after of the reference one (a totale of seven frames);
- Quefrency-domain: here is the same method but using cepstra insted of frames.

Into the Figure A.2 is show the behaviour of CPP and CPPS.

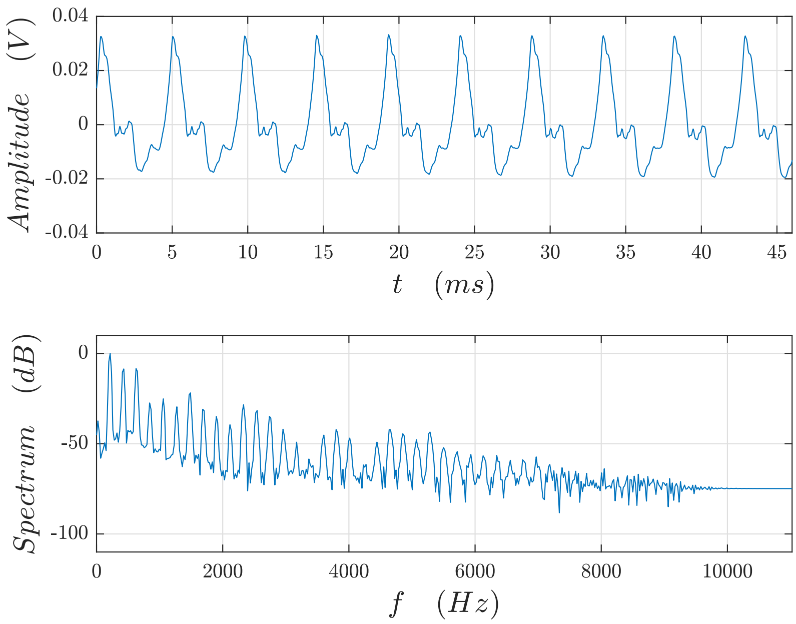


Figure A.1. *Single frame of original signal; Spectrum of signal.*

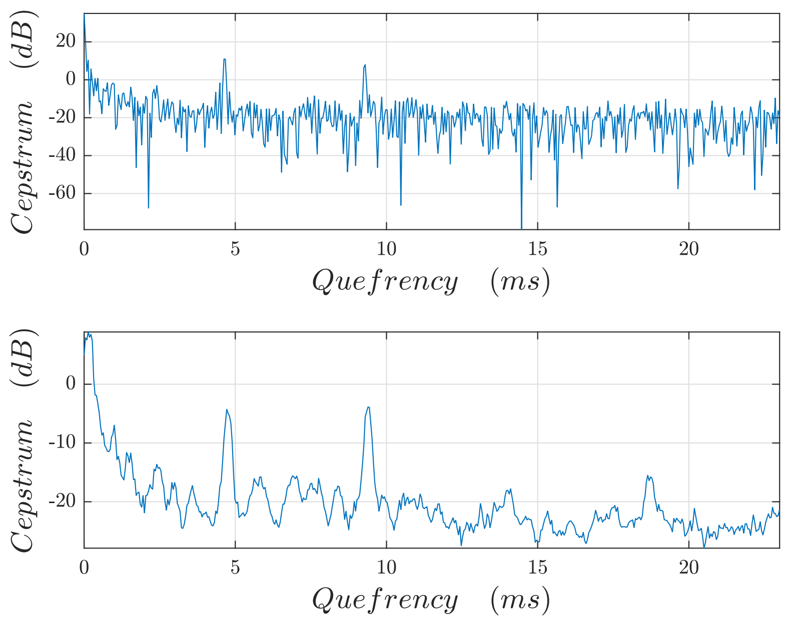


Figure A.2. *Single frame unsmoothed cepstrum; Single frame smoothed cepstrum.*

The amplitude is not the best parameter to observe because is affects from the degree of periodicity and it is influenced by the overall energy of the signal and the window size of the cepstrum analysis [1]. The important parameter here is not the amplitude, but the *Prominance*: the distance between the absolute amplitude of the peak and the value of the regression line at the same quefrency of the peak. The value of prominance is the analyzed parameter to discriminate the healty from unhealty people: for greater value of prominance, the signal is more regular and and the person are healty.
Using this signal, it is possible to calculate the Prominance but, before, it is necessary to create the regression line across the cepstrum. Here the cepstrum peak is found with a subtraction from the regression line with the smoothed cepstrum; this peak is searched into the quefrency range between 3.3 ms and 16.7 ms: in the domain of frequency this range coresponds at 60/300 Hz. In Figure A.3 is shown the behaviour of the regression line.

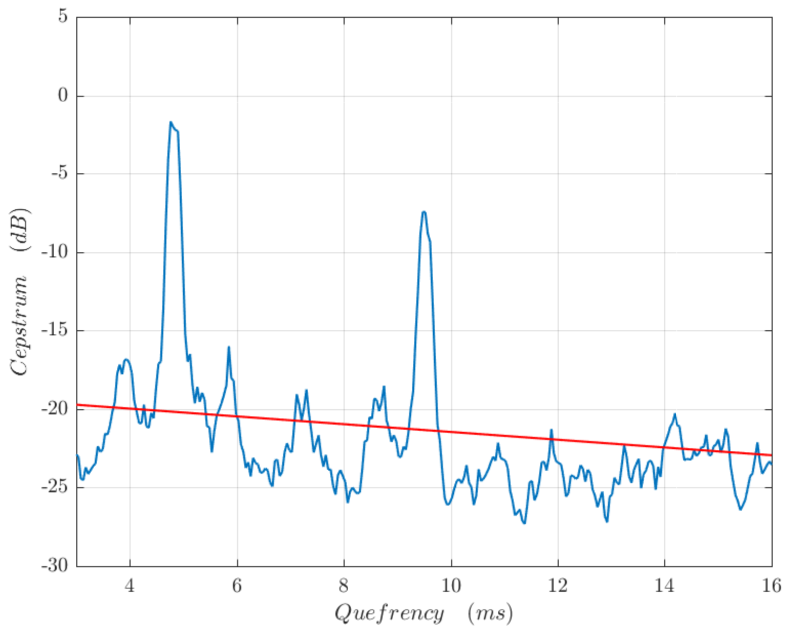


Figure A.3. *Single frame Cepstrum and Regression Line.*

The CPPS is calculated for each frame of the signal, the output of this algorithm is a matrix that contains this CPPS for each frame of every subject.
For example, there is an example of the total distribution of all the cepstra in Figure

A.4:

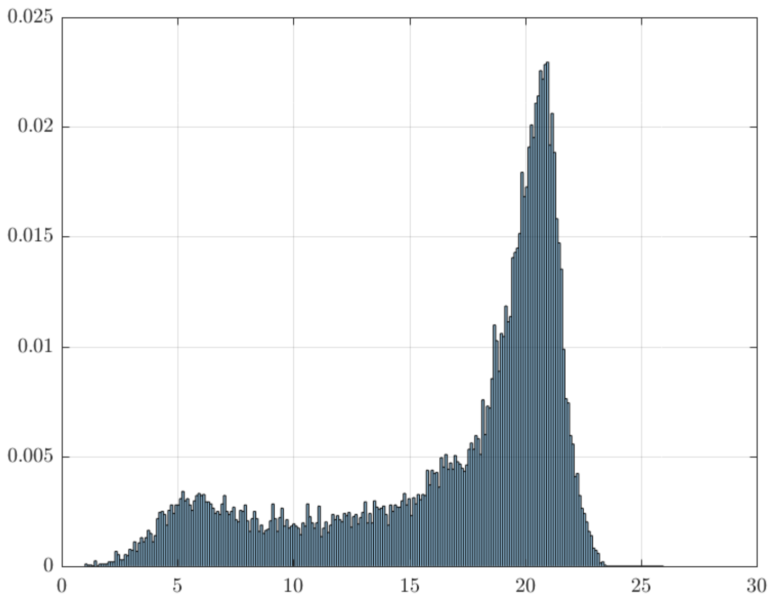


Figure A.4. *CPPS distribution.*

From this distribution, some parameters were calculated that are resumed in the Table A.1:

Parameters
Mean
Median
Mode
Standard Deviation
Range
5th Percentile
95th Percentile
Skewness
Kurtosis

Table A.1. *Parameters.*

These parameters are now object of study to found the best one that allow to discriminate healty and unhealty subjects [20].

Long Time Average Spectrum

The Long-Time Average Spectrum (LTAS) gives information about energy distribution in a spectrum of a signal over a period of time. The LTAS is obtained from the mean of the component of the spectrum in a long signal sample [26]. It is one of the most important averaging measures that works with long speech signals. The best results are obtained with signals no longer than 30-40 seconds [27]. The speech signal is the produced result of the sound source and the vocal tract transfer function. The vocal tract transfer function differs for different sound segments; the averaging process avoid the problem of the short-term variations due to phonetic structure and used language. LTAS is a parameter of frequency that used only voiced segments extracted from the signal because silent segments may require different analytical procedures [27]. The resulting spectrum gives an information about the sound source; if the analysis refers to voiced sound, the results will show that the sound source is the vibrating glottis [28]. The spectrum provides information about the good or poor voice quality in the healthy speakers. LTAS is used also for discriminate healthy from unhealthy people. For example, LTAS could be employ to continuous speech because it may shows a more representative sample of dysphonic speaking patterns than a sustained vowel [26]. This method has a great advantage: it is not invasive. This method does not required the cooperation of the patience and the analysis of the signal can be conducted after the recording data [28].

Anglosaxon Musical Notation

In the section 3.1, **Pitch Accuracy**, it could be find the *Anglosaxon Notation*: it rapresents the notation about notes in a different way. Into the graphs, it can be find letters like *C*, *E*, *G* and others. Into the Table A.2, there are shown the notes and their related anglosaxon notation.

Notes	Angl. Notation
DO	C
RE	D
MI	E
FA	F
SOL	G
LA	A
SI	B

Table A.2. *Anglosaxson Musical Notation.*

This notation is very used in the anglophone Europe and less in the south of Europe (like Italy, Spain and others). There is an only exception: in German, the note *SI* is called ***H*** instead of *B*. The "German" notation is used also in scandinavian countries.

Notes

Every notes have an own frequency; for example, we use as a reference frequency for the note LA (440 Hz) and its multiples. Following, it is avaiable the table used in this work with the frequency of each note.

Note	ottave									
	0	1	2	3	4	5	6	7	8	9
Do	16,35	32,70	65,41	130,8	261,6	523,3	1047	2093	4186	8372
Do#-Reb	17,32	34,65	69,30	138,6	277,2	554,4	1109	2217	4435	8870
Re	18,35	36,71	73,42	146,8	293,7	587,3	1175	2349	4699	9397
Re#-Mib	19,45	38,89	77,78	155,6	311,1	622,3	1245	2489	4978	9956
Mi	20,60	41,20	82,41	164,8	329,6	659,3	1319	2637	5274	10548
Fa	21,83	43,65	87,31	174,6	349,2	698,5	1397	2794	5588	11175
Fa#-Solb	23,12	46,25	92,50	185,0	370,0	740,0	1480	2960	5920	11840
Sol	24,50	49,00	98,00	196,0	392,0	784,0	1568	3136	6272	12544
Sol#-Lab	25,96	51,91	103,8	207,7	415,3	830,6	1661	3322	6645	13290
La	27,50	55,00	110,0	220,0	440,0	880,0	1760	3520	7040	14080
La#-Sib	29,14	58,27	116,5	233,1	466,2	932,3	1865	3729	7459	14917
Si	30,87	61,74	123,5	246,9	493,9	987,8	1976	3951	7902	15804

Figure A.5. *Table of frequencies.*

To calculate the semi-tone it is possible to multiply or divide a certain value (in the Formula A.1):

$$c = \sqrt[12]{2} \tag{A.1}$$

To find the follow semitone it is necessary to multiply the interested frequency for this value; to find the previous semitone the frequency must be divided for this value.

For example, to find the frequency of LA# the calculus to do is the following:

$$f = f_{prev} \cdot c \tag{A.2}$$

where f_{prev} is the frequency of the previous note (440 Hz) and c is the previous value. The result f is 466.2 Hz (as shown also in Figure A.5).

Ringraziamenti

Desidero ringraziare tutti coloro che mi hanno accompagnato lungo questo percorso di tesi. Ringrazio anzitutto il mio relatore, il Prof. Alessio Carullo per l'aiuto, la pazienza e la disponibilità fornitemi in questi mesi sia durante lo svolgimento del lavoro di tesi vero e proprio che durante il periodo di stesura. Ringrazio i miei correlatori ovvero la Prof.ssa Arianna Astolfi, il Prof. Alberto Vallan e soprattutto la Dott.ssa Antonella Castellana che mi ha seguito passo passo offrendomi le sue conoscenze e i suoi consigli per permettermi di lavorare al meglio e acquisire sempre nozioni nuove.
Un ringraziamento particolare va ai miei genitori senza i quali non avrei potuto iniziare il mio percorso di studi in ingegneria; mai è mancato il sostegno economico ed emotivo che mi ha permesso di raggiungere questo obiettivo. Grazie inoltre per l'enorme pazienza mostratami in questi anni e l'affetto con cui avete sempre appoggiato le mie scelte.
Inoltre, è stato possibile iniziare e condurre questo lavoro grazie alla gentile disponibilità dei cantanti che hanno partecipato con entusiasmo alle registrazioni: un sentito grazie va ad ognuno di voi che avete permesso la realizzazione di questo progetto.
Un ringraziamento speciale va alle mie amiche storiche quali Monica, Elena e Sara che da molti anni mi supportano e sopportano "fuori dal Poli" e a Michela che ha avuto la grande fortuna di potermi sopportare anche durante le lezioni e la preparazione degli esami di profitto.
Vorrei infine ringraziare tutte le altre persone a me più care che mi sono state accanto in questo percorso di studi.

Bibliography

[1] James Hillenbrand and Robert A Houde. Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech, Language, and Hearing Research*, 39(2):311–321, 1996.

[2] Apparato fonatorio - definizione. Online, accessed on September 2017.

[3] Wladimiro Del Colle. *Voce e Computer - Analisi Acustica digitale del Segnale Verbale*. Omega Edizioni, 2001.

[4] Wilmer T Bartholomew. A physical definition of good voice-quality in the male voice. *the Journal of the Acoustical Society of America*, 6(1):25–33, 1934.

[5] Antonio Romano, Ugo Cesari, Michele Mignano, Oskar Schindler, and Irene Vernero. La qualità della voce.

[6] Diana Deutsch. *The psychology of Music*. Academic Press, INC., 1982.

[7] Johan Sundberg. Articulatory interpretation of the singing formant. *The Journal of the Acoustical Society of America*, 55(4):838–844, 1974.

[8] Koichi Omori, Ashutosh Kacker, Linda M Carroll, William D Riley, and Stanley M Blaugrund. Singing power ratio: quantitative evaluation of singing voice quality. *Journal of Voice*, 10(3):228–235, 1996.

[9] M Usha, YV Geetha, and YS Darshan. Objective identification of prepubertal female singers and non-singers by singing power ratio using matlab. *Journal of Voice*, 31(2):157–160, 2017.

[10] Christopher Watts, Kathryn Barnes-Burroughs, Julie Estis, and Debra Blanton. The singing power ratio as an objective measure of singing voice quality in untrained talented and nontalented singers. *Journal of voice*, 20(1):82–88, 2006.

[11] Monica A McHenry and Joseph Evans. Aerobic exercise as a warm-up for singing: Acoustic impacts. *Journal of Voice*, 31(4):438–441, 2017.

[12] S Prytz. Long-time-average-spectra (ltas) analyses of normal and pathological voices. *Proc Int Assoc Logop Phon Congr, Copenhagen*, pages 459–475, 1977.

[13] Samantha Warhurst, Catherine Madill, Patricia McCabe, Sten Ternström, Edwin Yiu, and Robert Heard. Perceptual and acoustic analyses of good voice quality in male radio performers. *Journal of Voice*, 31(2):259–e1, 2017.

[14] Sheri L Cook-Cunningham and Melissa L Grady. The effects of three physical and vocal warm-up procedures on acoustic and perceptual measures of choral sound. *Journal of Voice*, 2017.

[15] Johan Sundberg. Level and center frequency of the singer's formant. *Journal of voice*, 15(2):176–186, 2001.

[16] Viviane M Oliveira Barrichelo, Reinhardt J Heuer, Carole M Dean, and Robert T Sataloff. Comparison of singer's formant, speaker's ring, and lta spectrum among classical singers and untrained normal speakers. *Journal of voice*, 15(3):344–350, 2001.

[17] Aaron M Johnson and Gail B Kempster. Classification of the classical male singing voice using long-term average spectrum. *Journal of Voice*, 25(5):538–543, 2011.

[18] BP Bogert, MJR Healy, and JW Tukey. The frequency analysis of time series for echoes: cepstrum, pseudo-auto covariance, cross-cepstrum, and shaft cracking. In *Proceedings of the Symposium on Time Series Analysis (M. Rosenblatt, Ed)*, pages 209–243.

[19] James Hillenbrand, Ronald A Cleveland, and Robert L Erickson. Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4):769–778, 1994.

[20] Antonella Castellana, Alessio Carullo, Simone Corbellini, Arianna Astolfi, M Spadola Bisetti, and J Colombini. Cepstral peak prominence smoothed distribution as discriminator of vocal health in sustained vowel. In *Instrumentation and Measurement Technology Conference (I2MTC), 2017 IEEE International*, pages 1–6. IEEE, 2017.

[21] Christine Johansson, Johan Sundberg, and Hermann Wilbrand. X-ray study of articulation and formant frequencies in two female singers. In *Proc of SMAC*, volume 83, pages 203–18, 1985.

[22] Johan Sundberg. Articulatory configuration and pitch in a classically trained soprano singer. *Journal of Voice*, 23(5):546–551, 2009.

[23] Mark Nolan, Brian Madden, and Edward Burke. Accelerometer based measurement for the mapping of neck surface vibrations during vocalized speech. In *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pages 4453–4456. IEEE, 2009.

[24] Antonella Castellana, Alessio Carullo, Arianna Astolfi, Giuseppina Emma Puglisi, and Umberto Fugiglando. Intra-speaker and inter-speaker variability in speech sound pressure level across repeated readings. *The Journal of the Acoustical Society of America*, 141(4):2353–2363, 2017.

[25] Pasquale Bottalico, Simone Graetzer, and Eric J Hunter. Effect of training and level of external auditory feedback on the singing voice: Pitch inaccuracy. *Journal of Voice*, 31(1):122–e9, 2017.

[26] Antonella Castellana. *Towards vocal-behaviour and vocal-health assessment using distributions of acoustic paramters*. Ph.d. thesis, Politecnico di Torino, 2018.

[27] Soren Y Lowell, Raymond H Colton, Richard T Kelley, and Youngmee C Hahn. Spectral-and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25(5):e223–e232, 2011.

[28] Elvira Mendoza, Juana Muñoz, and Nieves Valencia Naranjo. The long-term average spectrum as a measure of voice stability. *Folia phoniatrica et logopaedica*, 48(2):57–64, 1996.

[29] Anders Löfqvist and Bengt Mandersson. Long-time average spectrum of speech and voice analysis. *Folia Phoniatrica et Logopaedica*, 39(5):221–229, 1987.

[30] Vocal apparatus, physiology. Online, accessed on September 2017.

[31] Rudolf Weiss, WS Brown, and Jack Moris. Singer's formant in sopranos: fact or fiction? *Journal of Voice*, 15(4):457–468, 2001.