



Dossier / Dossier / Dossier

"Analysis beyond Notation in XX<sup>th</sup> and XXI<sup>st</sup> Century Music"

Alessandro Bratus and Marco Lutz (Guest editors)

## Manual Transcription and Instrumental Analysis of Singing through *Praat*

Paolo Bravi

Conservatorio Statale G.P. da "Palestrina", Italy  
pa.bravi@tiscali.it

### Abstract

In the field of ethnomusicology, the main tool-of-the-trade for music analysis has been musical transcription on the score, despite its acknowledged cultural bias and the limits of staff notation as a means of representing music conceived and performed outside of the Western music culture. In the digital era, a number of analytical tools have been designed which may serve to solve some of the problems related to the use of the score as a method for describing and visualizing music. These tools allow analyses to be performed, which were virtually impossible for most ethnomusicologists until a few decades ago. *Praat*, the well known software developed by Paul Boersma and David Weenink and designed for phonetic studies, may also be helpful for the annotation and analysis of the singing voice. This paper deals with some aspects of the use of this program for musicological aims, focussing on the relation between acoustic data and subjective musical interpretations and their relevance in analytic and perceptual investigations.

**Keywords:** Singing analysis, music transcription, singing perception, *Praat*

## La transcripción manual y el análisis instrumental del canto mediante *Praat*

### Resumen

En el campo de la etnomusicología la principal herramienta de mercado para el análisis musical ha sido la transcripción musical sobre la partitura, a pesar de su reconocido sesgo cultural y de los límites que tiene la escritura en pentagrama para representar la música concebida y ejecutada fuera de la cultura musical occidental. En la era digital fue diseñado un número de herramientas analíticas capaces de resolver algunos de los problemas relacionados con el uso de la partitura como método para describir y visualizar la música. Esas herramientas permiten hacer un tipo de análisis que, hasta hace unas pocas décadas, era virtualmente imposible de realizar para la mayoría de los etnomusicólogos. *Praat*, el conocido programa de



Los trabajos publicados en esta revista están bajo la licencia Creative Commons Atribución- NoComercial 2.5 Argentina

computación desarrollado por Paul Boersma y David Weenink y diseñado para los estudios fonéticos, puede también ser útil para la notación y el análisis del canto. Este artículo aborda algunos aspectos del uso musicológico de ese programa con énfasis en la relación entre el dato acústico y las interpretaciones musicales subjetivas, y su relevancia para el análisis y la investigación de la percepción.

**Palabras clave:** análisis del canto, transcripción musical, percepción del canto, *Praat*

---

## A transcrição manual e a análise instrumental do canto através do *Praat*

### Resumo

No domínio da etnomusicologia a principal ferramenta de mercado para a análise musical tem sido a transcrição musical na partitura, apesar do seu reconhecido viés cultural e dos limites da escrita num pentagrama como método para representar a música concebida e realizada fora da cultura musical ocidental. Na era digital, foram concebidas ferramentas analíticas capazes de resolver alguns dos problemas relacionados com o uso da partitura como método para descrever e visualizar a música. Estas ferramentas permitem fazer um tipo de análise que, até há algumas décadas atrás, era virtualmente impossível de realizar para a maioria dos etnomusicólogos. *Praat*, o conhecido programa de computação desenvolvido por Paul Boersma e David Weenink e desenhado para estudos fonéticos, pode também ser útil para a notação e para a análise do canto. Este artigo aborda alguns aspetos do uso musicológico desse programa com ênfase na relação entre o dado acústico e as interpretações musicais subjetivas, e a sua relevância para a análise e para a investigação da percepção.

**Palavras-chave:** análise de canto, transcrição musical, percepção do canto, *Praat*

---

Fecha de recepción / Data de recepção / Received: septiembre 2015

Fecha de aceptación / Data de aceitação / Acceptance date: noviembre 2015

Fecha de publicación / Data de publicação / Release date: febrero 2016



## Introduction

In the field of ethnomusicology, musical transcription has been a central and problematic issue for a long time. Different configurations of scales, intervals, ornaments, and rhythms, as well as the impossibility to realistically describe aspects which may be referred to as timbre, make the usual Western way of writing music not completely suitable, effective, and affordable in the case of music based on different codes (Stockmann 1989, Ellingson 1992, Nettle 2005, Baumann and Stock 2005). For this reason, musical transcriptions on the score are sometimes to be considered with caution or even suspicion. The digital revolution has dramatically changed the set of tools at hand for making music analysis. In particular, modern computer tools help to solve some of the problems related to the use of the score as a method for visualizing music melodies. In addition, it is now possible to provide a detailed analysis of aspects of the organization of musical sounds and, more generally, to reach a deeper understanding as regards the acoustic surface of musical performances<sup>1</sup>.

This article examines some issues concerning the transcription and analysis of singing voice intonation and discusses how the digital technologies may help in investigating the relation between the objective tracking of the fundamental frequencies of singing and the subjective perception and musical interpretation of a transcriber. In particular, I dwell on the possibilities provided by the *Praat* computer program (Boersma and Weenink 2015) in (i) giving an integrated representation of sung melodies in the form of a “mixed transcription” that shows both the fundamental frequency track and the musical notation given by the transcriber<sup>2</sup>, (ii) analyzing aspects of the melodic structure, and (iii) investigating the perceptual dimension of intonation in singing listening<sup>3</sup>. These issues are discussed taking a single sung verse as a case-example. The verse is in the Sardinian-Campidanese language and was performed by the

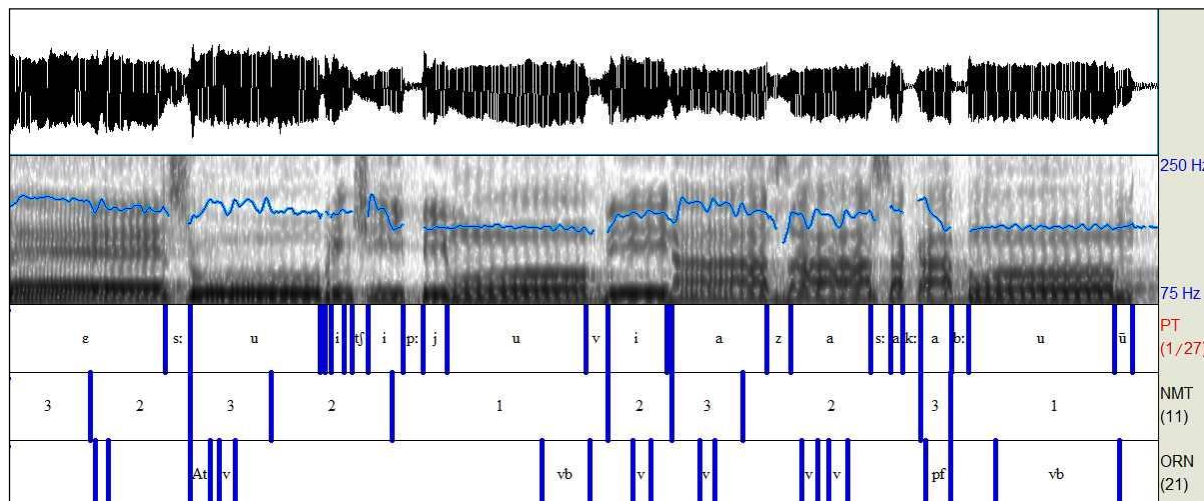
---

<sup>1</sup> The instrumental approach in music analysis clearly profits from the widespread diffusion of computers and the sound analysis software of today. However, the pioneering use of fundamental frequency analysers in the field of music psychology and ethnomusicology is time-honoured (Seashore 1938, Metfessel 1929, Seeger 1951, Gurvin 1953, Dahlback 1958). Nowadays, the number of tools dedicated to music analysis is huge and it is impossible (and perhaps useless) to even try to provide a thorough list. To mention just a few of the most popular and/or recent tools, in particular some of those explicitly devised to accomplish instrumental sound analysis, we can cite here *Sonic Visualizer* (2014), *Acousmographie* (2014), *Speech Analyzer* (2012), *iAnalyse4* (2014).

<sup>2</sup> In this article I shall distinguish between the acoustics of the concept of fundamental frequency (or  $f_0$ , according to the agreed terminology proposed in Titze *et alii* 2015) and the psychoacoustic one of pitch: “[p]itch is a subjective quality of sound which enables one to compare one sound with another; whereas frequency is a physical attribute, pitch is psychological. Thus, low pitches correspond to low frequencies, and high pitches correspond to high frequencies. However, there is no exact one-to-one correspondence between pitch and frequency. The relation depends on the observer, the frequency range, and loudness, to certain extent” (Askill 1979: 45).

<sup>3</sup> In particular, *Praat* will be used through a set of scripts from the SVAAT package, made by the author for [s]inging [v]oices [a]nalysis [a]nd [t]ranscription and designed to both help manual notation of sung melodies and to investigate the relation between the subjective musical interpretation of a transcriber and the acoustic “facies” of sung melodies. In the present paper, due to the lack of space, a number of aspects of the problem of singing notation and analysis, such as the issues regarding singing timbre, dynamic, agogic, and melodic profiles, vocal registers, laryngeal mechanisms, etc., as well as the issue of the automatic transcription of melodies, are not addressed despite their relevance to the topic. For more information on the SVAAT package, please contact the author at pa.bravi@tiscali.it.

Sardinian *cantadori* (improvising poet-singer) Roberto Zuncheddu during a poetic duel that took place in Sestu (Cagliari – Italy) on 8 September 2003, during the village patronal feast<sup>4</sup>. The sung line was annotated by the author using IPA symbols for the phonetic transcription and a numbered musical notation for the musical transcription, together with symbols designed to classify ornamental phenomena (*portamento*, *voltas*, *vibrato*, etc.; cf. Bravi 2010: 445-456). These annotations were made on a *Praat* TextGrid, where the annotations can be exactly aligned with the sound signal (figure 1).



**Figure 1.** The *Praat* Editor with the manually annotated sung verse taken as a case-example. From the top: waveform, wide-band spectrogram with pitch contour superimposed, TextGrid with three tiers: phonetic transcription, numbered musical notation, singing ornamentation.

### The transcriber's dilemma

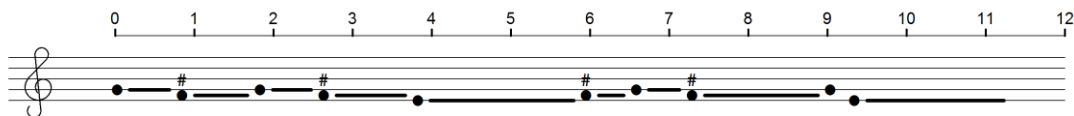
Transcription has been one of the core issues of ethnomusicology for many decades, and a number of different methods and adaptations have been used to describe musical sounds through staff notation. A series of conceptual oppositions has been used to describe the degree of “precision” of a music transcription –i.e. the accurateness of the correspondence between the recorded sounds and their notation– from different points of views: prescriptive *vs.* descriptive (Seeger 1958), general *vs.* specific (Hood 1971), synthetic *vs.* analytic or schematic *vs.* detailed (Stockmann 1989). In fact, the fineness of detail of a transcription may change, depending on the chosen level of analysis, the specific aims of the research, and ultimately on the subjective choice on the part of the transcriber.

As far as the “accurateness” of representation of performed melodies is concerned, a plot

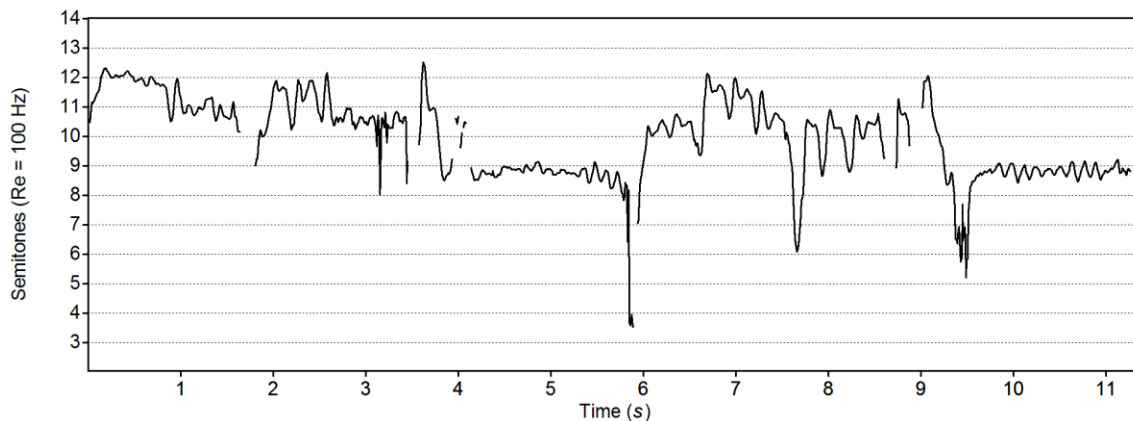
<sup>4</sup> In particular, this sung verse is extracted from the opening poem of the performance. The metrical-musical form used by the poet-singer is the *mutetu longu*, the most important genre in the Sardinian-Campidanese poetical tradition (Zedda 2009). The performance was recorded on location by the author. The audio file used as a case-example can be listened [here](#).

with the contour of the evolution of the fundamental frequency of a melodic line gives the most detailed description of a melody in visual form. There are various kinds of musicologists who favour melodic representations through  $f_0$  contours over standard musical notation: those who emphasize the importance of melodic details; those who are more interested in intonation than in rhythmic features; those who aim at excluding the “subjectivity” of human transcription; those who do not want to appear conditioned by Western musical theory and systems of representation; those who seek automatic methods of representation of melodies; those who plan to carry out research on very large corpora of melodies, etc. The emphasis on the advantages offered by  $f_0$  extractors with respect to standard musical transcription is common to all these categories of musicologists despite their different perspectives and objectives.

The two kinds of visual representation of singing –human transcription and the  $f_0$  contour– share some basic principles, but it is clear that the difference between the two is wide and fundamental. Both staff notation and a  $f_0$  profile give a representation of sound focussed on the temporal evolution of the intonation. In figure 2 this feature is represented in both ways. In the case of staff notation, intonation is graphically expressed in terms of notes of variable height and duration or –as in this case, which represents a free-rhythm sung line– by notes spaced according to their absolute duration; in the case of the  $f_0$  contour, intonation is represented by a line (or by a series of points) that moves up and down in a bidimensional space, where the X-axis corresponds to time and the Y-axis to the  $f_0$  values detected<sup>5</sup>. It is clear that the two representations are pretty different, so that if one compares a “basic” –or, to use one of the dichotomies cited before, “prescriptive”, “general”, “synthetic”– transcription of the sung line (as the one proposed here and made by myself) with the graph of the  $f_0$  contour, it is not even easily recognizable that they refer to the same musical event. Apart from any considerations regarding the ethnocentric character of the medium used for sound representation and the epistemological and cultural implications of this kind of scientific practice (*cf.* Marian-Bălașa 2005, Staniek 2014), it is clear that whereas the “basic” transcription drops many melodic subtleties recognized by a careful listening, the intricate  $f_0$  contour presents a number of peaks and valleys that, in many cases, are not even perceivable from listening. No matter how detailed a staff notation may be, when one deals with singing, the  $f_0$  profile shows a higher level of complexity.



<sup>5</sup> In figure 2 the unit of measure is the semitone with reference to 100 Hz. Other scale types based on psychoacoustic criteria are those expressed in mel, bark and ERB (Houtsma 1995).



**Figure 2.** Two graphical representations of the same sung verse in free rhythm: at the top, a synthetic transcription representing the basic melodic line; at the bottom, a bare  $f_0$  contour on a semitone scale.

On the one hand, the (ethno)musicologists who use staff notation as a means of representing and analyzing sung melodies are obliged to face the ultimately intractable arbitrariness of the choice of the accurateness of the transcription, as well as the frequent practical impossibility of being able to provide a notation that is at the same time both clear and reliable. On the other hand, however, those who use computer tools for fundamental frequency extraction in singing analysis are faced with the opposite problem, i.e. the fact that “[t]he profusion of detailed visual data [...] will have to be re-translated into musical reality and musical sense” (Herzog 1957: 73). In fact, the information provided by a  $f_0$  contour is redundant with respect to what the ear perceives and to what the human cognitive system can interpret on a musical basis. Therefore the problem is to pick out data that are perceptually significant and have a musical value within the overwhelming amount of acoustic data, and disregard the rest as not relevant. In this view, the following paragraphs examine some common issues arising during instrumental analyses of sung melodies.

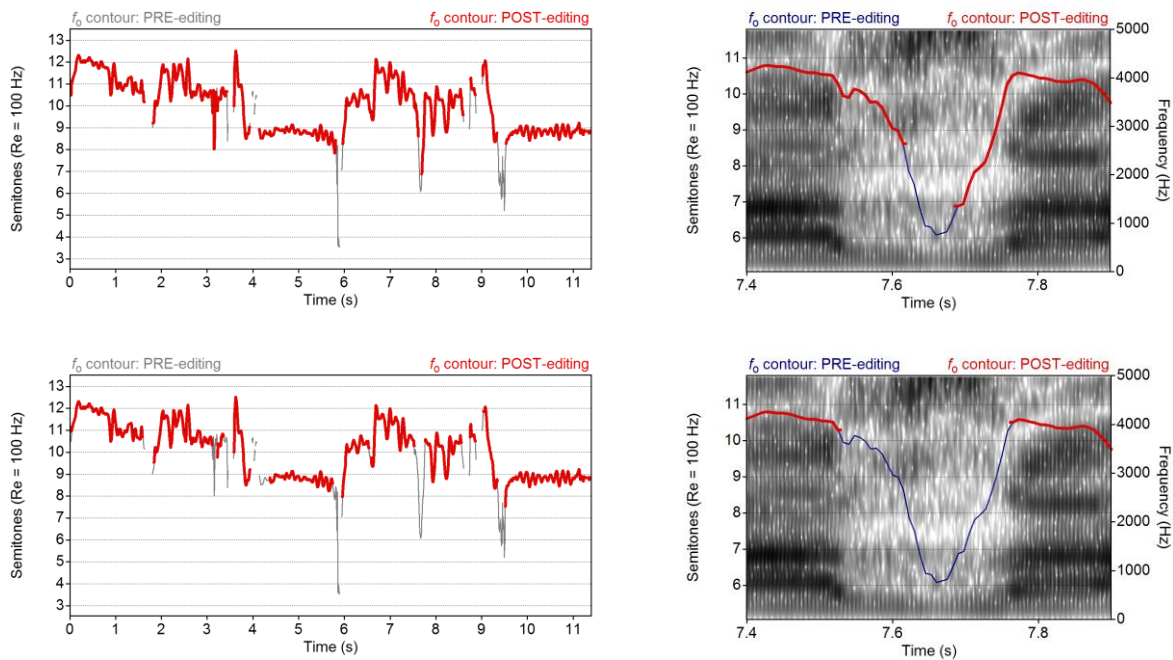
### Dealing with microintonation

When one listens to singing,  $f_0$  movements that do not occur in correspondence to vowels are not normally perceived and interpreted as musically significant. There may be wide  $f_0$  fluctuations on voiced consonants that are not intended by the speaker and which are the inescapable side effects of the mechanics of vocal articulation. Effects such as these may be referred to as “microintonation” and for most cases may be disregarded as not relevant for the listeners and for musical analysis (Patel 2008: 213). While such phenomena have been the object of phonetic studies<sup>6</sup>, no detailed study on this issue has been carried out until now in the field of singing perception. On the basis of impressionistic evaluations, one could say that the rapid movements of the fundamental frequencies occurring on voiced consonants, though clearly

<sup>6</sup> For example, a well-studied topic regards the effects of voiced vs. voiceless obstruents on the  $f_0$  in the following vowel (Löfqvist, Baer, McGarr, and Story 1989).

visible in the  $f_0$  curve, may only be perceived if the phono is listened to as isolated from the context or through delayed playback.

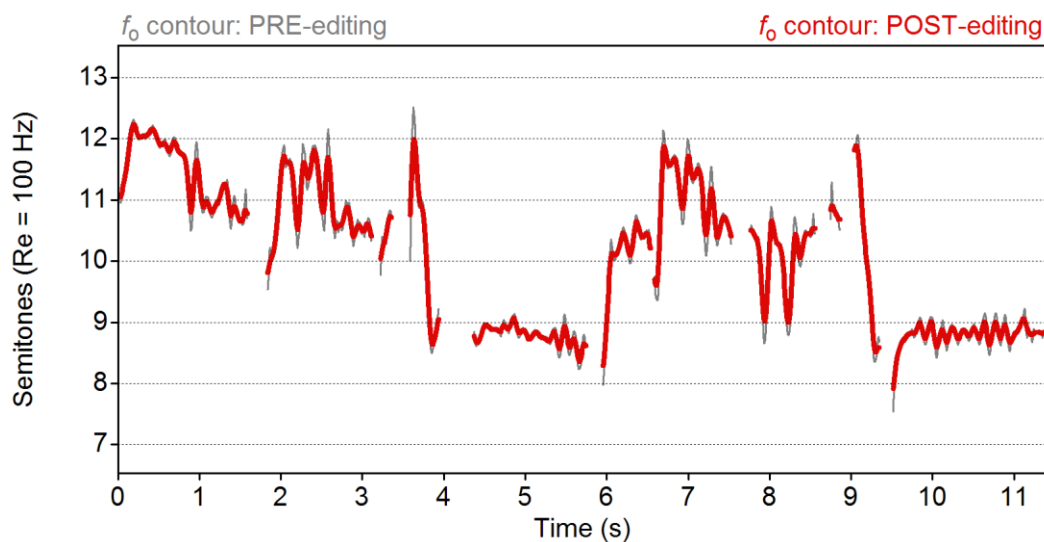
This topic may be investigated in an experimental way through *Praat*. In particular, one script in the SVAAT package (note 3) provides facilities that allow the sound to be manipulated in order to test the perceptual relevance of these acoustic features and also, if appropriate, the exclusion of these perturbations from the  $f_0$  profile (figure 3)<sup>7</sup>.



**Figure 3.** The effects of the two strategies for editing the  $f_0$  contours. The superimposed red contour is obtained either by changing the default voicing threshold from 0.45 to 0.80 (top-left panel) or by unvoicing segments relevant to consonantal phonos (bottom-left panel). The two panels on the right show a detail of the different effects in correspondence to the voiced fricative [z], showing a more precise deletion of voiced frames obtained by manual annotation of phonetic segments (bottom-right panel).

<sup>7</sup> This goal could be achieved in different ways. One option is that of changing the default value of the voicing threshold. By heightening the standard threshold of the *Praat* algorithm (0.45), the strength of the unvoiced pitch candidates changes and, as a result, the majority of the pitch values which we may suppose to be irrelevant in singing perception will be excluded from the pitch line. The advantage of this option is that it does not necessitate any preliminary sound annotation, and may therefore be useful in the analysis of many (non-annotated) sound files. If one needs to carry out a more precise control, SVAAT provides an option which allows pitch candidates corresponding to specific consonant categories to be unvoiced. This requires a preliminary phonetic transcription which may be obtained through automatic procedures of phonetic alignment (*cf.* Goldmine 2011) or –for most cases– via manual annotation of the sound file.

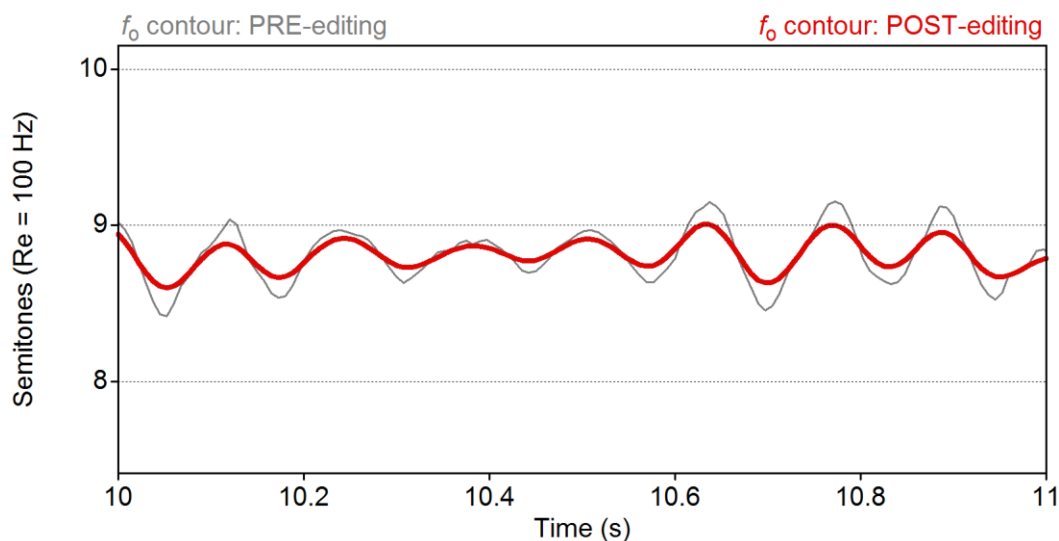
Another aspect of singing microintonation regards the fact that, in most cases, the evolution of a  $f_0$  line is characterized by a highly jagged curve. Frequency micro-perturbation (jitter) is, to a certain extent, a typical and inescapable phenomenon of vocal sounds. This kind of micro-perturbation of the sung melody is usually not perceivable as a proper intonation feature. When jitter is high, a listener may perceive a particular voice quality: s/he may hear the singing voice as hoarse, harsh or rough, but s/he does not hear a real voice modulation (Laver 1980). Therefore, insofar as one is interested in the melodic profile and not in the quality of the voice, a smoothing of the  $f_0$  curve aimed at excluding this kind of micro-perturbation of the  $f_0$  profile is useful. *Praat* provides a command to do this and allows the user to define the width of the smoothing bandwidth<sup>8</sup>. By applying a smoothing filter to the  $f_0$  contour, one obtains a clearer profile which, from the perceptual point of view, should be equivalent to the original profile (figure 4a). However, it has to be noted that this filter has to be used carefully, since in some cases this perceptual equivalence does not hold true. To quote an example, in figure 4b one can see the effect of smoothing (with the default bandwidth value of 10 Hz) in a short vibrato segment taken from the final part of the sung verse used here as a case-example. It is clear that the waving of the  $f_0$  curve is more regular after the editing obtained through smoothing, and that one of the parameters usually employed to describe *vibrato* features, i.e. the *vibrato* extent, is lower in the case of the smoothed line<sup>9</sup>.



<sup>8</sup> The default value of the smoothing bandwidth in *Praat* is 10 Hz. The *Praat* procedure for smoothing pitch tracks is briefly described here: <http://uk.groups.yahoo.com/group/praatusers/message/5529>.

<sup>9</sup> For that matter, a comparison between the original sound with a manipulated version of the same sound, in which the original  $f_0$  track is replaced with the smoothed one, shows that the difference between the two is faintly perceivable, due to the fact that the vibrato rate remains unchanged. *Praat* allows the pitch and the duration of sounds to be manipulated through “Manipulation” object types (see the *Praat* intro, chap. 8, for explanations and instructions). SVAAT provides various kinds of facilities designed to test –both visually and aurally– the perceptual effects of the  $f_0$  editing operations here described.





**Figures 4a and 4b.** The effect on the whole sung verse of smoothing the  $f_0$  contour (4a: top panel) and on a short part of the sung line where a vibrato is present (4b: bottom panel).

### Sung notes

For centuries the connection between music and language has been a crucial point in a huge number of studies (some recent and comprehensive publications on this topic are Patel 2008, Arbib 2013). Here I refer to this vast area of studies only as regards the part that concerns the parallel between the concept of “note” in music and that of “phoneme” in phonology, which has all too often been put forward and employed (Springer 1953, Nettl 1958, Hood 1971, Bernstein 1976, Sloboda 1985, Aiello 1994). The analogy between the two concepts has also been discussed or rejected. For example, Nash Rose observes that the note-phoneme analogy does not hold true for various reasons:

The term “phoneme” is an analytical concept based upon opposition of features, and has variant realizations specifiable by rule. A musical note is a physical entity in a gradation of physical entities. Notes cannot be put into a distinctive feature matrix, as one note differs from another note only in pitch. Yet, two notes having the same pitch may have completely different musical functions in the same melody (Rose 1973: 42).

In parallel, it has to be observed that in the same field of phonology, the existence of an entity such as the phoneme, the usefulness of this concept and the acceptability of the segmental paradigm have also been questioned (Albano Leoni 2009).

As far as we are concerned, when dealing with singing analysis, we may be helped by the idea that a musical “note” is a form of a culturally established categorization of sounds and an abstraction and a simplified view with respect to the acoustic reality of sung melodies. This kind

of reduction and musical interpretation can be the subject or the starting point for further research into both the perception of singing and the characterization of different singing styles.

Translating the complexity of a  $f_0$  contour of a song into sequences of notes, i.e. series of stable and predefined pitch levels, does however pose several problems. Firstly, musical perception and interpretation undergo a culture influence. As Seashore put it almost eighty years ago, “the matter of hearing pitch is largely a matter of conceptual hearing in terms of conventional intervals” (1938: 269) and, therefore, any decision to reduce sung melodies in terms of notes is associated to the “chunking” that is necessary for the musical information processing (Dowling and Harwood 1986), but may be affected by a (Western) culture bias (Will 1998, Burns 1999, Stadler Elmer and Elmer 2000, Trehub 2013). As music cognition happens within emic systems of pitch categorization, the (outsider) transcriber may be the victim of “aural ghosts” (Ambrazevičius 2004: 108). Secondly, there is room for subjectivity in music perception and interpretation, and this may lead to differences in transcriptions carried out by different transcribers (England, Garfias, Kolinski, List, Rhodes, and Seeger 1964, List 1974, Jairazbhoy 1977). Thirdly, there are melodic lines that cannot be properly described in terms of “note”, but which could instead be seen as “paths” –i.e. dynamic movements either top-directed or bottom-directed, such as the “tumbling strains” described by Curt Sachs (1962); the up-and-down oscillations of the voice through rapid melismatic movements (Hurtado Torres and Hurtado Torres 2005); the passages between wide “pitch zones” (Kondrat’eva 2009: 28), or also the random movements between “moving anchors” (Carpitella 1966: 270) or within a “diastematic area” (Sorce Keller 1990: 215).

The reverse side of the question is that when we are dealing with singing performance coming from Western written scores, it is not always easy to identify the “prescribed” melody in the song. For example, it is quite impossible (or arbitrary) in many cases to distinguish a rapid succession of notes from the linear or winding movements based on the acoustic surface of the singing performance. As Nicholas Cook asks himself:

[a]n opera singer sings a rapid scale: does she actually sing every note? Is there a clear transition between each note and the next, or does one slide into the other? If so, how do you tell where one note stops and the next one starts? Or is there just a rather uneven glissando? (1994: 79).

Be it as it may, the idea that notes form the “basic bricks” of music is central in classical Western (and not only) music theory. The possibility of delimitating the limits of this postulate when dealing with vocal music and possibly putting this venerable axiom under verification and discussion can clearly profit from formal analyses of singing which are not limited to the standard transcription of melodies.

*Praat* can be used to this end for investigations on the relation between notes and  $f_0$  tracks, i.e. between the subjective interpretation and the physical reality of singing, and possibly to compare interpretations given by different transcribers<sup>10</sup>. The two forms of representation may

---

<sup>10</sup> One script in the SVAAT package provides a way of checking the transcription by ear, comparing it with the

also be visualized together in the form of a “mixed transcription” (figure 5). The plot comprises both the  $f_0$  track and the transcribed notes (indicated in the form of grey rounded rectangles, with vertical lines separating sung syllables), together with symbols representing ornamental figures placed above the plot. This allows a direct visual representation of the intervals actually sung in the performance. In this case, for example, it can be observed that the height of the second degree is lower than that of the correspondent degree of the equal-tempered scale<sup>11</sup>. This kind of graphical rendering allows a clear representation of the relation between the  $f_0$  track and the musical interpretation of the song in terms of notes and ornaments.

A few further considerations can be made starting from this case-example. Firstly, the superimposition of notes and the  $f_0$  track allows one to observe –and possibly to measure and analyse – what happens in the points of transition between adjacent notes. In many cases, the attack of the note is characterized by a rapid rise, since the note is reached starting from below. The singer usually needs what Nicole Scotto Di Carlo calls “modulation time”<sup>12</sup> to get to the desired height. This is in part a physiological fact, but it may also be a part of the style of a particular singer or a feature of a specific performance (*cf.* Bravi 2015). Secondly, one can observe that in some cases  $f_0$  heights are above or below the medium value of the scale degree. These kind of deviations –overshootings and undershootings– can be analysed as far as their distribution is concerned, for example, either in correspondence to upward or downward melodic movements (Pfordresher, Brown, Meier, Belyk, and Liotti 2010) or in relation to particular vowels (Sundberg 1987: 143-144). Thirdly, the graphic representation shows that one characterizing feature of this singing style is the inner movement within the note: not only are there many ornaments (particularly voltas and vibrato sections), but also an overall melodic instability is present. This “capillary trembling”<sup>13</sup> of singing is a feature that characterizes many different singing styles and which may only be visualized and properly analyzed through instrumental  $f_0$  extraction.

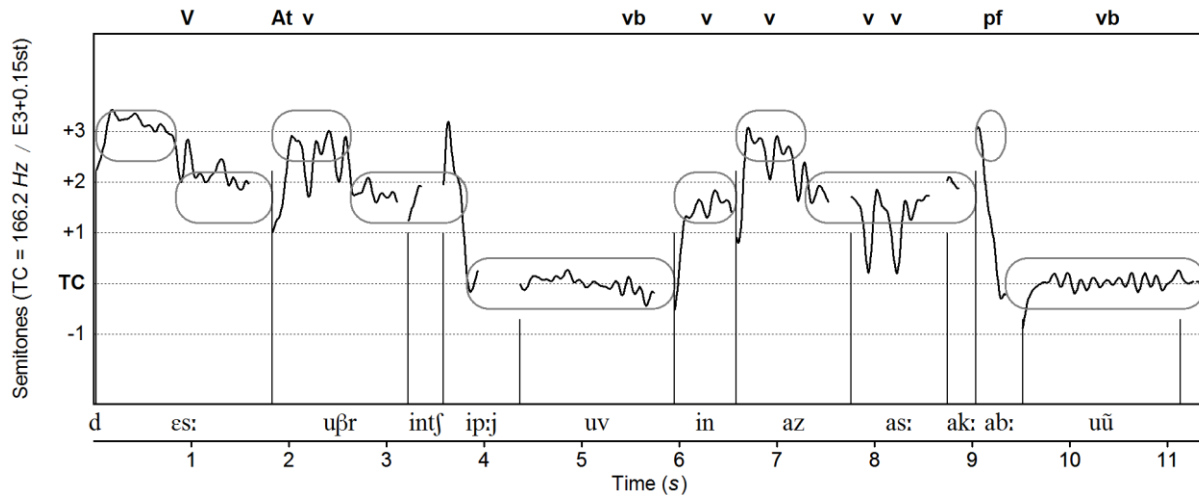
---

original sounds. This is possible by either listening to the original recording (or hummed rendering of the  $f_0$  track) superimposed on a rendering of the notes as sine sounds (sound examples [here](#) and [here](#)) or confronting the recording with its manipulated version, where the original pitch tier is replaced with a new construed one based on the notes as transcribed. In order to listen to this file you have to download *Praat* from <http://www.fon.hum.uva.nl/praat/> and install it in your machine. Then, you can automatically download the compressed file by clicking [here](#). (For explanation on how to read Manipulation files in *Praat*, see [http://www.fon.hum.uva.nl/praat/manual/Read\\_from\\_file\\_\\_.html](http://www.fon.hum.uva.nl/praat/manual/Read_from_file__.html); on how to view them, see <http://www.fon.hum.uva.nl/praat/manual/Editors.html>; on how to use them, see <http://www.fon.hum.uva.nl/praat/manual/ManipulationEditor.html>).

<sup>11</sup> The interval is calculated here as the median value of the  $f_0$  in the segments identified in the transcriber’s ciphered notation as “2”.

<sup>12</sup> “[T]emps d’ajustement” (Scotto Di Carlo 2007: 159).

<sup>13</sup> “[C]apillare vibratilità” (De Natale 1990: 92).



**Figure 5.** Visual representation of the sung verse taken as an example in the form of a “mixed transcription” displaying both the  $f_0$  track and the transcriber’s musical interpretation.

### Scales and intervals

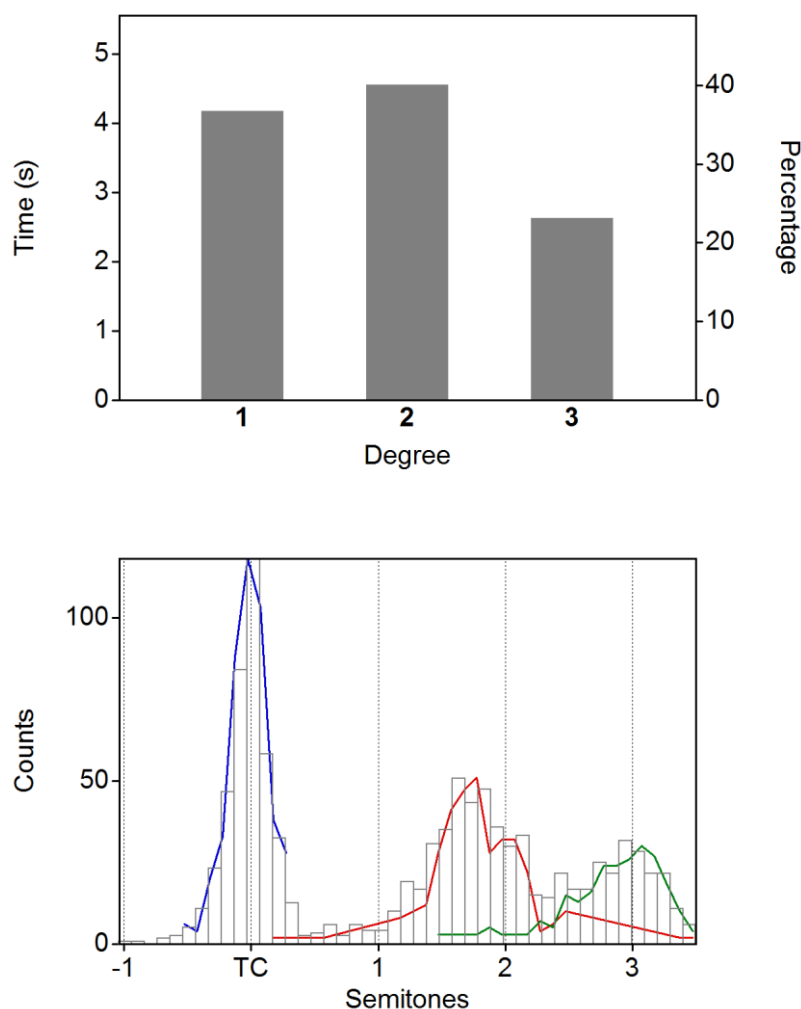
After the well-established postulate that music is made of notes, a second deeply-rooted axiom in music theory is that the musical organisation of sounds is based on scales, i.e. abstract models of ordered intervals based on the discretisation of the frequency continuum. The concept of scale has been at the centre of a different epistemological debate in comparison with the discussion arisen on the concept of note. The idea that “the definition of the scale is an inescapable assumption for whatever study on a given musical culture”<sup>14</sup> is not usually put under discussion. *Per se*, the existence or, at least, the usefulness of the concept of scale –and of related concepts like *mode* (Powers 1980) or *gamut / modulo* (Roberts and Jennes 1925, Carpitella and Biagiola 1978)– as a tool of the trade for music analysis is normally taken for granted. The stabilization of heights is, in fact, one of the acoustic traits that distinguish singing from speech (List 1963, Giannattasio 2005). However, what has instead been observed in a great number of ethnomusicological publications, starting from Alexander J. Ellis’s milestone essay focussing “On the Musical Scales of Various Nations” (Ellis 1885), is that the scales and the intervals from which they are made are variable and not a universal or natural thing.

This kind of issue is more easily examined when dealing with instruments with fixed intonation than with singing or with instruments, such as the violin and, to a certain extent –wind instruments. In the case of the sung verse examined here through *Praat*, two different scale analyses may be carried out. The first one is a quantitative analysis of the presence of the scale degrees (as notated by the transcriber) in the song. In this case, as shown in figure 6a, the sum of the durations of the notes shows that the third degree is less present than the first and second

<sup>14</sup> “[L]a definizione della scala costituisce un presupposto inevitabile per qualsiasi studio su una data cultura musicale” (Giuriati 1991: 90).

degrees. The second one regards the distribution of the  $f_0$  values. In this case, as shown in figure 6b, it can be observed that the degree classified by the transcriber as the tonal centre is more clearly defined than the second and the third degrees, and that the second degree in this sung line is lower than the correspondent degree of the equal tempered scale, at least as far as its most prominent peak is concerned.

Of course, an analysis based on just one single sung line –as in this case-example– does not allow any generalization. A study of the distribution of scale degrees and a thorough examination of the width of the intervals require an analysis of a long performance or of a corpus of melodies (*cf.* Van der Meer 2000, Moelants, Cornelis, and Leman 2009).



**Figures 6a and 6b.** Gamut in the form of a bar chart (6a: top panel) and  $f_0$  histogram (6b: bottom panel) of the example sung verse (class interval = 10 cents). Bins (in grey) come from the  $f_0$  track of the entire verse; coloured lines come from segments classified by the transcriber as degree 1 (blue line), degree 2 (red line), degree 3 (green line). Values lower than the first quartile have been removed.

### Exploring ornamentation

Melodic ornamentation is one of the main features that characterize a vocal style. To some extent, what can be defined as an “ornament”, with respect to what is considered to be the skeleton of the melody, is a subjective matter. For example, why (or when) can one consider (and transcribe) a melodic succession such as C-D-C-B-C as a turn and not as a sequence of principal notes? This is certainly connected to various factors (e.g., the rapidity of the melodic movement, its presence/absence in different variants of the melody, the relation with other parts, etc.), but the transcription and classification of a melodic movement as an ornament –i.e. the decision whether it is a part of the skeleton or merely a temporary deviation from it, or (to use the venerable Aristotelian terminology) if it is *substantia* or *accidens*– is, beyond historically determined conventions, a subjective decision on the part of the transcriber<sup>15</sup>.

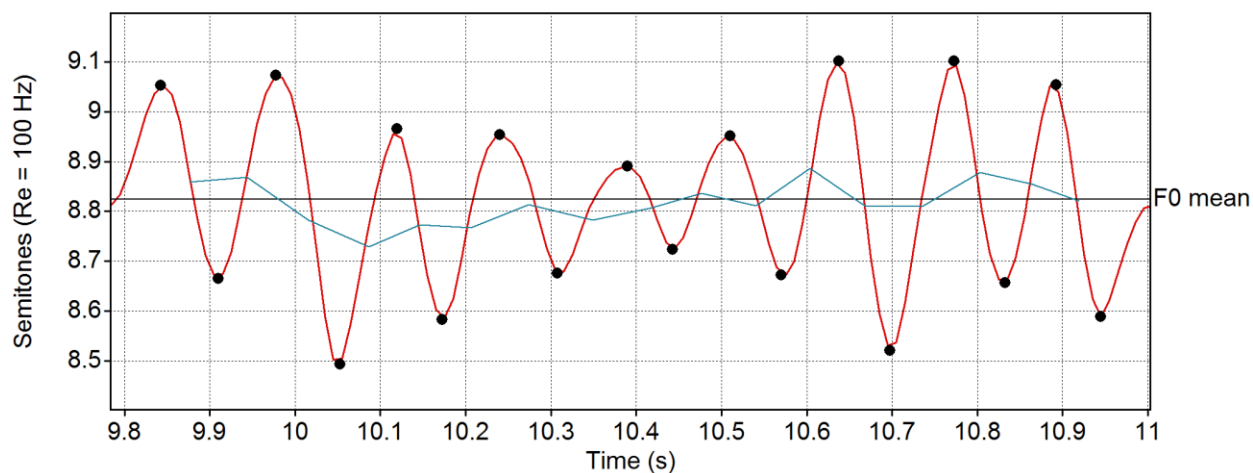
In ethnomusicological transcriptions, embellishments are usually written as “grace notes”. Printed smaller on the score and with no definite time values, these notes indicate quick melodic movements that are considered of less structural value than the “basic” notes. As a matter of fact, ornaments of vocal melodies are anything but irrelevant aspects of musical styles: they “color the song in a special way” (Lomax and Grauer 1968: 66); if omitted, “the melody would lose much of its atmosphere and feeling” (66). Not only is ornamentation often the most evident acoustic feature of particular voices and singing styles, but in some cases it also has a prominent role in the overall characterization of a piece of music or a music style, whereas the main melody –i.e. the sequence of principal notes– is secondary, to the point that it can barely be recognized *per se* (for example, by trying to play it on a piano) if it is deprived of its embellishments. The aesthetic value of a vocal piece of music may not rely on the main melody, but mostly on the expressive means that it employs.

An analysis of the  $f_0$  track is highly profitable for investigations of vocal ornamentation. Figures 7 and 8 show the graphical outputs of two kinds of automatic analyses of one of the most important vocal line ornaments, namely, the vibrato, as can be carried out using *Praat*. In this case, the vibrato rate (near to 8 cycles per second) and the standard deviation of the vibrato extent are far from the values that are common in the *vibrato* of most opera singers (Sundberg 1999: 195-197) and identify a particular realization of this ornament that is typical in Sardinian song (Bravi 2012)<sup>16</sup>.

---

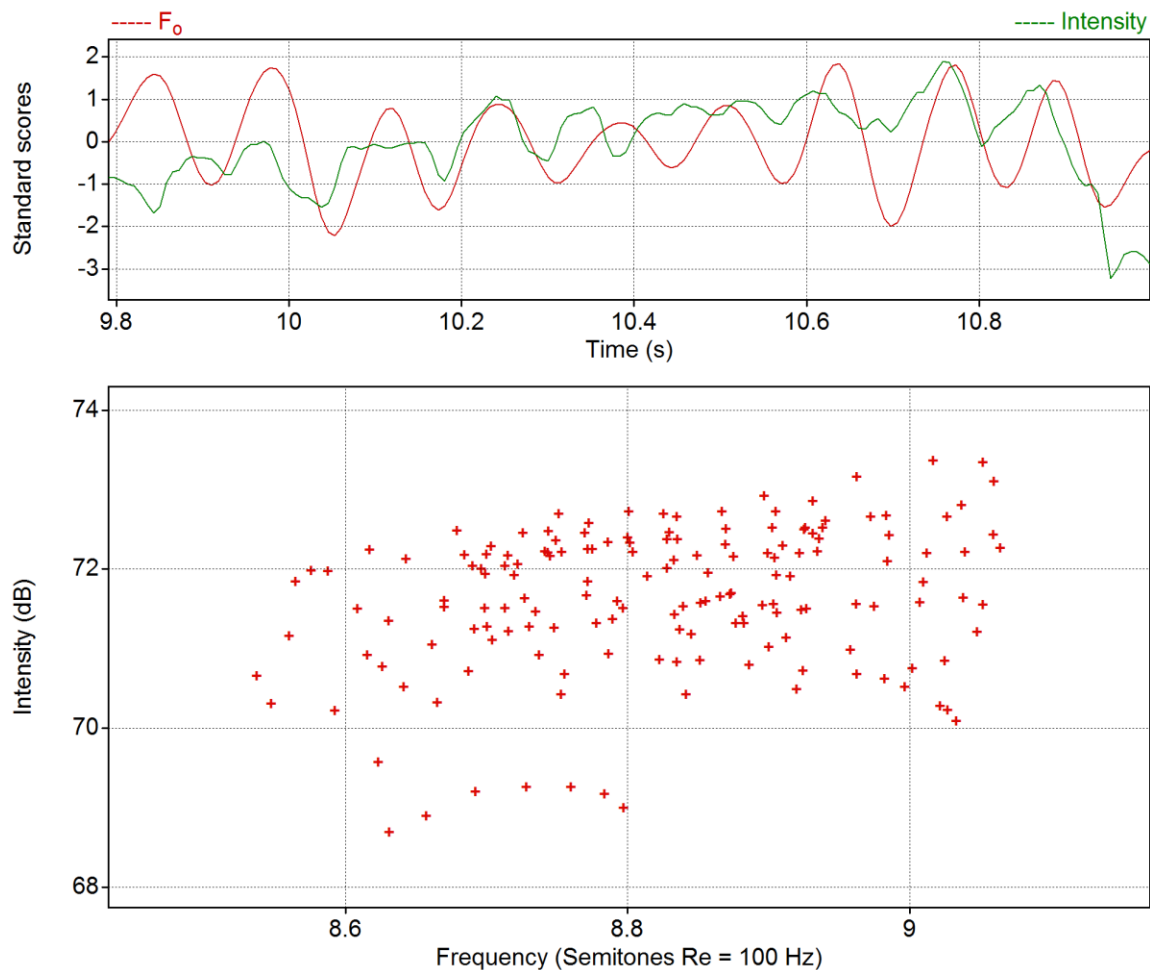
<sup>15</sup> The denomination, definition, and classification of musical ornaments are not exactly shared and have sometimes evolved in time. In this case a wide notion of ornament has been used, related to perceptually significant micro-melodic movements, not transcribed as major movements from one scale degree to another (*cf.* Fussi and Magnani 2010: 7-8).

<sup>16</sup> The relevant acoustic parameters of the *vibrato* segment are the following: *Vibrato Rate* (mean) = 7.718 – *Vibrato Rate* (sd) = 0.606 – *Vibrato Extent* (mean) = 0.430 – *Vibrato Extent* (sd) = 0.131 – *Moving mean  $f_0$*  (mean) = 0.035 – *Moving mean  $f_0$*  (sd) = 0.024. The  $f_0$  track is smoothed with bandwidth equal to 20 Hz.



**Figure 7.** A graphical representation of the  $f_0$  track identified in one segment by the transcriber as vibrato in the sung line of the case-example. The moving mean (blue line), the peaks and valleys (black points) and the overall mean (black horizontal line) are superimposed.

Another aspect that may be of interest in investigations aimed at characterizing vibrato is the relation between  $f_0$  and intensity (Seashore 1938: 44, Fussi and Magnani 2010: 264-266). The graphical output of this kind of analysis is presented in the two plots in figure 8, showing a mild correlation between the two variables.



**Figures 8a and 8b.** The relation between  $f_0$  and intensity in the vibrato segment: the top panel (8a) shows the  $f_0$  and intensity tracks in standard scores; in the bottom panel (8b), a scatter plot of the values of the two variables.

### The interface between acoustic and musical perception

If one skims ethnomusicological publications from recent decades, it is quite common to find some kinds of graphical representations of sounds obtained through instrumental analysis, such as  $f_0$  profiles or spectrograms. Although it is rare that a proper training in the empirical study of music is part of an ethnomusicologist's methodological background, many scholars in this field are in fact aware that staff notation represents –particularly in the case of vocal music– a mere compromise. That is to say, they know that it may be suitable for some kinds of music analysis and basic visual representations, whatever its intrinsic limitations and its cultural bias, but they also know that it is far from being demonstrated that pitch and rhythm are always the most significant and characteristic part in a vocal performance. And they know that even if one chooses to focus his/her analysis on these features alone, there are a number of cases where a realistic and detailed analysis of singing cannot rely on the subjective and rather coarse representation of sounds offered by staff transcription, whatever the adjustments s/he may decide to adopt. At the same time, however, they realise that the instrumental analysis of singing



provides an unmanageable amount of information that does not seem to correspond to perception. In fact, if the staff does not tell us everything about a sung intonation (and sometimes does it even weirdly or badly), a  $f_0$  contour says more than what the ear can perceive and may be musically intelligible.

Much work has still to be done to fill this gap. At least under some respects, this issue seems to provide a parallel in the musical field for the relation between phonetics and phonology found in linguistics. However, while the latter has plenty of studies focussing on the problem of the ‘interface’ between the two aspects of the sonic shape of languages (Kingston 2007), the theme of the relations between etic and emic data in music –and the relevant perspectives in the fields of both ethnomusicology and music psychology– has not been properly addressed yet, in spite of the recurring claims. On the one hand, the science of singing perception is still a relatively young discipline and many topics have not been thoroughly investigated; on the other, –and this is a major epistemological issue of much research carried out within the realm of music psychology– these kinds of studies usually lack an intercultural view of the problems of music perception. In fact, in the vast majority of cases, the cognitive approach in musical studies takes into account aspects of Western music (or Western musical categories) and investigates them using Western (or Western educated) subjects. The overall assumption is –even if not explicitly declared– that the basic mechanisms of music perception are universal (but is there scientific evidence to prove this? And such a position sends a shiver down the spine of most ethnomusicologists). On this basis, the findings of experiments conceived and carried out within the scope of the current Western music paradigms are considered universally valid, in the implicit presumption that the same findings would be obtained regardless of the time and place in which the experiment are conducted. Cognitive studies in the field of music often take into account different prospective factors of variability, and particularly the fact that “trained” musicians may have different (supposedly more akin) abilities when listening to music with respect to “non-trained” subjects. But the fact that different musical “training” might entail different “ways of listening” –to deliberately use a very generic expression that covers a number of aspects, such as perception, interpretation, evaluation, expectation, and so on– is rarely taken into account and tackled. The assumption that the physiology of hearing and the basic machinery of hearing perception is an unvarying element in humankind is implicitly converted into one where it is irrelevant as to whether one has been “trained” in different musical cultures or simply been accustomed to listening to different kinds of music and giving them musical interpretations and evaluations based on different criteria. In this respect, there has been a kind of walleye in most psychology of music studies, a substantial cultural bias which casts a shadow on the generalizations regarding the “musical mind” coming from studies exclusively internal to the Western world (Becker 2009, Morrison and Demorest 2009).

On the contrary, what listeners perceive and hence interpret as musically significant seems to be deeply affected by their cultural background, that is, by their musical experience, habits, and practice. Musical evaluations and judgements start from the different aesthetic and cultural values relevant to specific musical cultures. These values also drive the attention of both the singers/players and the listeners towards particular musical aspects of the performances, which

are not always shared, at least in the same terms, by all music cultures.

Taking this epistemological point as the needle for any systematic music investigation, musicologists may profit from the computer assisted analysis of singing and from the facilities provided by programs such as *Praat* for the investigation of the relation between acoustic data and subjective –and culturally affected– perception. Exploring the interface between the two dimensions is a necessary step for obtaining a more comprehensive view of the qualities of singing voices.

## Bibliography

- Acousmographie*. 2014. <http://www.inagrm.com/accueil/outils/acousmographie> [accessed 15.09.2015].
- Aiello, Rita. 1994. “Music and Language: Parallels and Contrasts”. In: Aiello, Rita and John A. Sloboda (eds.), *Musical Perceptions*, pp. 41-63. New York-Oxford: Oxford University Press.
- Albano Leoni, Federico. 2009. *Dei suoni e dei sensi. Il volto fonico delle parole*. Bologna: Il Mulino.
- Ambrazevičius, Rytis. 2004. “Scales in Traditional Solo Singing”. In: Lipscomb, S.D., R.Ashley, R.O. Gjerdingen, and P. Webster (eds.), *Proceedings of the 8th International Conference on Music Perception & Cognition, Evanstone*, pp. 105-108. Adelaide: Causal Production.
- Arbib, Michel A. (ed.). 2013. *Language, Music, and the Brain. A Mysterious Relationship*. Cambridge MA-London: The MIT Press.
- Askill, John. 1979. *Physics of Musical Sounds*. New York-Cincinnati-London-Toronto-Melbourne: Litton Educational Publishing, Inc.
- Baumann, Max Peter and Jonathan P. Stock (eds.). 2005. *The World of Music* 47 (2): “Notation, Transcription, Visual Representation”.
- Becker, Judith. 2009. “Crossing Boundaries: An Introductory Essay”. *Empirical Musicology Review* 4 (2): 45-48.
- Bernstein, Leonard. 1976. *The Unanswered Question: Six Talks at Harvard*. Cambridge MA: Harvard University Press.
- Boersma, Paul and Weenink, David. 2015. *Praat: Doing Phonetics by Computer*. <http://www.fon.hum.uva.nl/praat/> [accessed 15.09.2015].
- Bravi, Paolo. 2010. *A sa moda campidanese. Pratiche, poetiche e voci degli improvvisatori nella Sardegna meridionale*. Nuoro: ISRE.
- \_\_\_\_\_. 2012. “Measurements of Vibrato Parameters in a Performance of Sardinian Traditional Sung Poetry”. In: Falcone, Mauro and Andrea Paoloni (eds.). *La voce nelle applicazioni. VIII Convegno AISV (Associazione Italiana Scienze della Voce), Roma 25-27 Gennaio 2012*, pp. 67-78. Roma: Bulzoni.
- \_\_\_\_\_. 2015. “Target Reaching in Melodic and Prosodic Analysis: Some Observations following the Examination of a Performance of Sardinian A Mutetus Longus Poetry”. In: Romano,

- Antonio; Matteo Rivoira, and Ilario Meandri (eds.). *Aspetti prosodici e testuali del raccontare: dalla letteratura orale al parlato dei media - X Convegno Nazionale AISV - Associazione Nazionale Scienze della Voce -Torino, 22-24.01.2014*, pp. 89-104. Alessandria: Edizioni Dall'Orso.
- Burns, Edward M. 1999. "Intervals, Scales, and Tuning". In: Deutsch, Diana (ed). *The Psychology of Music*, pp. 215-264. San Diego: Academic Press.
- Carpitella, Diego. 1966. "Studio etnomusicologico della documentazione sonora" In: Biocca, E., *Viaggio tra gli Indi. Alto Rio Negro. Alto Orinoco. Appunti di un biologo*, pp. 267-291. Roma: CNR.
- Carpitella, Diego and Sandro Biagiola. 1978. "Note per la compilazione della scheda FKM". In: Ministero per i Beni Culturali ed Ambientali, *Ricerca e catalogazione della cultura popolare*, pp. 47-50. Roma: Museo Nazionale della Arti e Tradizioni Popolari-Istituto Centrale per il Catalogo e la Documentazione.
- Cook, Nicholas. 1994 . "Perception: A Perspective from Music Theory". In: Aiello, Rita and John A. Sloboda (eds). *Musical Perceptions*, pp. 64-98. New York-Oxford: Oxford University Press.
- Dahlback, Karl. 1958. *New Methods in Vocal Folk Music Research*. Oslo: Oslo University Press.
- De Natale, Marco. 1990. *Analisi della struttura melodica*. Milano: Guerini e associati.
- Dowling, W. Jay and Dane L. Harwood. 1986. *Music Cognition*. Orlando: Academic Press.
- Ellingson, Ter. 1992. "Transcription". In: Myers, Helen (ed.), *Ethnomusicology: An Introduction*, pp. 110-152. New York: Norton.
- Ellis, Alexander John. 1885. "On the Musical Scales of Various Nations". *Journal of the Society of Arts* 33: 485-527.
- England, Nicholas M., Robert Garfias, Mieczyslaw Kolinski, George List, Willard Rhodes, and Charles Seeger. 1964. "Symposium on Transcription and Analysis: A Hukwe Song with Musical Bow". *Ethnomusicology* 8 (3): 223-277.
- Fussi, Franco and Silvia Magnani. 2010. *Le parole della scena. Glossario della voce del cantante e dell'attore*. Torino: Omega Edizioni.
- Giannattasio, Francesco. 2005. "Dal parlato al cantato". In: *Enciclopedia della musica. Vol. 5: L'unità della musica*, 1003-1036. Torino: Einaudi.
- Giuriati, Giovanni. 1991. "Scala". In: Agamennone, Maurizio, Serena Facci, Francesco Giannattasio e Giovanni Giuriati, *Grammatica della musica etnica*, pp. 83-112. Roma: Bulzoni.
- Goldmine, Jean-Philippe. 2011. "EasyAlign: an Automatic Phonetic Alignment Tool under Praat". *12th Annual Conference of the International Speech Communication Association*, 27-31.08.2011 - Firenze, Italy.
- Gurvin, Olav. 1953. "Photography as an Aid in Folkmusic Research". *Norveg* 3: 181-196.
- Herzog, George. 1957. "Music at the Fifth International Congress of Anthropological and Ethnological Sciences". *Journal of International Folk Music Council* 9: 71-73.
- Hood, Mantle. 1971. *The Ethnomusicologist*. New York: McGraw-Hill.
- Houtsma, Adrianus J. M. 1995. "Pitch Perception". In: Moore, Brian C. J. (ed.), *Hearing*, pp.

- 267-296. San Diego: Academic Press.
- Hurtado Torres, Antonio and David Hurtado Torres. 2005. "Nuevos horizontes en la transcripción musical del flamenco". *Música oral del sud* 6: 251-267.
- iAnalyse4*. 2014. [http://logiciels.pierrecouprie.fr/?page\\_id=672](http://logiciels.pierrecouprie.fr/?page_id=672). [accessed 15.09.2015].
- Jairazbhoy, Nazir. 1977. "The 'Objective' and Subjective View in Music Transcription". *Ethnomusicology* 21 (2): 263-273.
- Kingston, John. 2007. "The Phonetics–phonology Interface". In: De Lacy, Paul (ed.), *The Cambridge Handbook of Phonology*, pp. 401-434. Cambridge: Cambridge University Press.
- Kondrat'eva, Nataliya. 2009. "Pitch Structures in the Traditional Music of the Peoples of Siberia". In: Niemi, Jarkko (ed.). *Perspectives on the Song of the Indigenous Peoples of Northern Eurasia: Performance, Genres, Musical Syntax, Sound*, pp. 23-34. Tampere: Tampere University Press.
- Laver, John. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
- List, George. 1963. "The Boundaries of Speech and Song". *Ethnomusicology* VII (1): 1-16.
- \_\_\_\_\_. 1974. "The Reliability of Transcription". *Ethnomusicology* XVIII (3): 353-377.
- Löfqvist, A., T. Baer, N.S. McGarr, R.S. Story. 1989. "The Cryothyroid Muscle in Voicing Control". *Journal of Acoustical Society of America* 85: 1314-1321.
- Lomax, Alan and Victor Grauer. 1968. "The Cantometric Coding Book". In: Lomax, Alan. *Folk Song Style and Culture*, pp. 34-74. New Brunswick-London: Transaction Publishers.
- Marian- Bălașa, Marin. 2005. "Who Actually Needs Transcription? Notes on the Modern Rise of a Method and the Postmodern Fall of an Ideology". *The World of Music* 47 (2): 5-30.
- Metfessel, Milton. 1929. *Phonography in Folk Music: American Negro Songs in New Notation*. Chapel Hill: University of North Carolina Press.
- Moelants, Dirk, Cornelis Olmo, and Marc Leman. 2009. "Exploring African Tone Scales". *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*: 489-494.
- Morrison, Steven J. and Steven M. Demorest. 2009. "Cultural Constraints on Music Perception and Cognition". *Progress in Brain Research* 178: 67-77.
- Nettl, Bruno. 1958. "Some Linguistic Approaches to Musical Analysis". *Journal of the International Folk Music Council* 10: 37-41.
- \_\_\_\_\_. 2005. *The Study of Ethnomusicology: Thirty-one Issues and Concepts*. Champaign, Illinois: University of Illinois Press.
- Patel, Aniruddh. 2008. *Music, Language, and the Brain*. Oxford: Oxford University Press.
- Pfordresher, Peter Q., Steven Brown, Kimberly M. Meier, Michel Belyk and Mario Liotti. 2010. "Imprecise singing is widespread". *Journal of the Acoustical Society of America* 128: 2182-2190.
- Powers, Harold S. 1980. "Mode". In: Sadie, Stanley (ed.), *The New Grove Dictionary of Music and Musicians - Vol. 12*, pp. 376-450. London: Macmillan.
- Roberts, Helen H. and Diamon Jennes. 1925. *Report of the Canadian Arctic Expedition (1913-1918) - Vol. 14: Eskimo Songs. Songs of the Copper Eskimos*. Ottawa: F.A. Acland.

- Rose, Nash. 1973. *Turkish Intonation: an Instrumental Study*. The Hague: Mouton & Co.
- Sachs, Curt. 1962. *The Wellsprings of Music*. The Hague: Martinus Nijhoff.
- Scotto Di Carlo, Nicole. 2007. "Les Dysfonctionnements de la Voix Chantée". *TAPA* 26: 157-183.
- Seashore, Carl E. 1938. *Psychology of Music*. New York: McGraw-Hill.
- Seeger, Charles. 1951. "An Instantaneous Music Notator". *Journal of the IFMC* 3: 103-107.
- \_\_\_\_\_. 1958. "Prescriptive and Descriptive Music Writing". *Musical Quarterly* 44 ( 2): 184-195.
- Sloboda, John. 1985. *The Musical Mind: The Cognitive Psychology of Music*. Oxford: Oxford University Press.
- Sonic Visualizer*. 2014. <http://www.sonicvisualiser.org/>. [Accessed 15.09.2015].
- Sorce Keller, Marcello. 1990. "Alcune considerazioni per uno studio analitico della melodia nelle trascrizioni popolari e in quelle extraeuropee". In: De Natale, Marco, *Analisi della struttura melodica*, pp. 209-230. Milano: Angelo Guerini e Associati.
- Speech Analyzer*. 2012. <http://www-01.sil.org/computing/sa/index.htm>. [accessed 15.9.2015].
- Springer, George. 1953. "Language and Music: Parallels and Divergences". In: *For Roman Jakobson*, pp. 504-513. The Hague: Mouton.
- Stadler Elmer, Stefanie and Franz-Joseph Elmer. 2000. "A New Method for Analysing and Representing Singing". *Psychology of Music* 28: 23-42.
- Staniek, Jason. 2014. "Forum on Transcription". *Twentieth-Century Music* 11: 101-161.
- Stockmann, Doris. 1989. "Il problema della trascrizione nella ricerca etnomusicologica". In: *Ethnomusicologica. Seminari internazionali di etnomusicologia*, pp. 209-238. Siena: Accademia Musicale Chigiana.
- Sundberg, Johan. 1987. *The Science of the Singing Voice*. DeKalb, Illinois: Northern Illinois University Press.
- \_\_\_\_\_. 1999. "The Perception of Singing". In: Deutsch, Diana (ed.), *The Psychology of Music* [2<sup>nd</sup> edition], pp. 171-214. San Diego: Academic Press.
- Titze, Ingo R., et alii. 2015. "Toward a Consensus on Symbolic Notation of Harmonics, Resonances, and Formants in Vocalization". *The Journal of the Acoustical Society of America* 137 (5): 3005-3007.
- Trehub, Sandra E. 2013. "Communication, Music, and Language in Infancy". In: Arbib, Michael A. (ed.), *Language, Music, and the Brain. A Mysterious Relationship*, pp. 462-479. Cambridge MA-London: The MIT Press.
- Van der Meer, Wim. 2000. "Theory and Practice of Intonation in Hindustani Music". In: Barlow, Clarence. (eds.). *The Ratio Book*, pp. 50-71. Köln: Feedback Papers.
- Will, Udo. 1998. "An Ethnomusicologist's Dilemma: Technology and the Analysis of Music". *Ethnomusikologian Vuosikirja – Finnish Yearbook of Ethnomusicology* 10: 5-12.
- Zedda, Paolo. 2009. "The Southern Sardinian Tradition of the Mutetu Longu: A Functional Analysis". *Oral Tradition* 24 (1): 3-40.



---

**Biography / Biografía / Biografia**

Currently a contract professor at the Conservatorio di Musica “Palestrina” in Cagliari (Italy), Paolo Bravi has two Ph.D.s in “Methodologies of Anthropological Research” (University of Siena, 2008) and in “Theories and History of Languages” (University of Sassari, 2013). His research focuses on the formal features and cultural values of singing and spoken voices in oral traditions and adopts models, techniques and research methods from both ethnomusicology and phonetics.

---

**How to cite / Cómo citar / Como citar**

Bravi, Paolo. 2016. “Manual Transcription and Instrumental Analysis of Singing through *Praat*”. In: Bratus, Alessandro and Marco Lutz (eds.). Dossier: “Analysis beyond Notation in XX<sup>th</sup> and XXI<sup>st</sup> Century Music”. *El oído pensante* 4 (1). <http://ppct.caicyt.gov.ar/index.php/oidopensante> [Web: DATE].