



<http://www.diva-portal.org>

This is the published version of a paper presented at *Proceedings of International Computer Music Conference (ICMC)*.

Citation for the original published paper:

Einarsson, A., Friberg, A. (2015)

Using Singing Voice Vibrato as a Control Parameter in a Chamber Oper.

In: *Using Singing Voice Vibrato as a Control Parameter in a Chamber Opera*

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-179637>

# Using Singing Voice Vibrato as a Control Parameter in a Chamber Opera

**Anna Einarsson**

KMH Royal College of Music,  
Stockholm  
anna.einarsson@kmh.se

**Anders Friberg**

KTH Royal Institute of Technology,  
Stockholm  
afriberg@kth.se

## ABSTRACT

*Even though a vast number of tools exist for real time voice analyses, only a limited number of them focus specifically on singing voice, and even less on features seen from a perceptual viewpoint. This paper presents a first step towards a multi-feature analysis-tool of the singing voice in a composer-researcher collaboration. A new method is used for extracting the vibrato extent of a singer, which is then mapped to a sound generation module. This was applied in the chamber opera *Ps! I will be home soon*. The experiences of the singer performing the part with the vibrato-detection were collected qualitatively and analyzed through Interpretative Phenomenological Analyses (IPA). The results revealed some interesting mixed feelings of both comfort and uncertainty in the interactive setup.*

## 1. INTRODUCTION

The singing voice is one of the most versatile instruments, with an abundance of different parameters for pitch, timbre, timing and speech control. Potentially this makes it ideal to use as control signal in interactive music. However, when making an overview of accessible tools to use for analyses of singing voice in real time, and to be implemented in artistic work, some tools still seem to be missing, or at least are not so readily available. These are tools that mimic the multifaceted characteristics of the singing voice from a perceptual viewpoint. Thus to make new implementations from compositional needs seem almost mandatory in a creative process.

This paper is an offspring of one such collaborative effort, where the perceptual viewpoint is put in the foreground. The first step was to elaborate on relevant features to extract. They were chosen as to be easily integrated in a musical work realised live, and to be used by singers unfamiliar in working with live electronics. We envision a multi-featured "listening unit" that was able to extract several different parameters from the singing voice.

Copyright: © 2015 Anna Einarsson et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](https://creativecommons.org/licenses/by/3.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

In this work, we started with singing voice vibrato and developed an extraction method inspired by a previous model [1]. In the below described context of the chamber opera *Ps! Jag kommer snart hem!* [2] we then explored a beta-version of the vibrato extraction model, in relation to how it may be used as a mean for affecting subsequent electronic sounds in computer assisted composition, and how the performer experienced this part in both the cognitive and emotional realm.

## 2. VIBRATO AS SINGING VOICE FEATURE

Vibrato is one of the salient features that characterises the voice [3]. It is a useful feature also due to the relative ease with which singers can isolate and manipulate it. This said with the reservation that it does not apply to singers from all styles of music [4]. Depending on the tradition the singer emanates from different styles call for different vibrato utilisation, [5] but there are also individual differences; the vibrato is more or less an integral part of the singer's individual timbral identity [6]. Thus the amount of effort as well as comfort in manipulating this parameter may differ among singers. Interesting to note is then that technology in this way can work in two ways: both by being affected by the voice and by informing the voice of its (perhaps up to this point unexplored) resources and possibilities.

Having said that vibrato may be challenging to alter for the individual singer; Vibrato is not an altogether uncomplicated matter to make use of in composition either. Perceptually it is part of the affective prosody that reveals emotional states to the listener [7]. The amount of vibrato can be used as a deliberate cursor for style, or its absence a demarcation against undesired associations.

Previous research on singing voice vibrato detection has for example focussed on how to distinguish singing from speaking voice, [8] or identification of singers in order to form a so-called singer ID [9] [10].

## 3. SOFTWARE MODULES

The different software modules are implemented in MAX/MSP and consist of three parts: The vibrato extraction, the mapping, and the sound generation modules.

### 3.1 Vibrato Extraction

In this module the vibrato extent (amplitude) is extracted from audio, vibrato rate omitted. The input is the monophonic audio signal from the singer. (1) In the first step pitch and sound level are extracted using the YIN module provided by IRCAM [11]. (2) A determination of a “proper” tone is done using a combination of pitch range, sound level limits, and the quality estimation factor from YIN. Only when all requirements are met, the extracted pitch is passed on for analysis (gated). (3) The pitch signal is divided in two parts. The first part is slightly low-pass filtered. The second part is median filtered using a window corresponding to one vibrato cycle assuming a vibrato rate of 6 Hz. This will cancel out any regular vibrato around this frequency. (4) The two parts are subtracted. The resulting signal will be zero when no vibrato is present. (5) Finally the absolute value of the signal is averaged using a median filter twice the length of a vibrato cycle. The resulting signal is proportional to the vibrato extent in cents.

### 3.2 Mapping

The incoming vibrato control signal was divided into four intervals (0-0.2, 0.2-0.25, 0.25-0.3, 0.3-0.45) and each interval called on a different synthesizer, all run by a master clock. Smoothing was applied to the control signal calculating the median of every 5 incoming control values. Also latency was built in of 600 ms to filter out unwanted triggers.

### 3.3 Sound generation

The patch for synthesis consisted of four polyphonic synthesizers, each containing resonance-filters and oscillator banks, excited by noise. Tuning was set from data stored as lists. Small random variations were added to both the amplitude and sustain in order to make the resulting synthesis more dynamic.

## 4. USAGE IN THE CHAMBER-OPERA

The application was implemented into an artistic work that was a commission from the Malmö Opera to composer <first author> called *Ps. Jag kommer snart hem!* (Eng. Ps. I will be home soon!). It is a chamber opera for 6 chamber musicians, 4 vocalists and electronics (live performed, tape and/or interactive), libretto by Maria Sundqvist [3]. The libretto takes the life of the famous writer Astrid Lindgren’s assumed role model for Pippi Longstocking’s father, Calle Pettersson, as a point of departure. The part where vibrato extraction was implemented is called *Vykort*. Its instrumentation is mezzo-soprano, tenor saxophone and live interactive electronics as a stand-alone patch.

### 4.1 Singer

The part *Vykort* was sung by a mezzo-soprano that had no previous experience of working with live electronics. She was classically trained, and also a trained conductor, thus accounted for part of the musical studying during rehearsals. She seemed modestly comfortable in manipulating her vibrato.

### 4.2 Compositional considerations

#### 4.2.1 Mapping

Fels, Gadd and Mulder [12] suggest the metaphor as a model for designing a device that also would be understandable to an audience. Since metaphors are culturally shared, shared knowledge is presumed. As Simon Emmerson [13] has described in his seminal work *Living electronics*, many composers have a history of implementing models, analogies and metaphors as bases for form as well as sound material, yet varying in the need for the model to be perceived by the audience.

In *Ps!* part one *Vykort*, the approach of metaphorical mapping as described by Mark Johnson [14] was applied. In short, according to the *structured mapping approach*, what is transferred are not the attributes but the relations between domains: the source domain and the target domain. In this work the metaphor was one of *force*. Density was the expression of force in the target domain, where the target was an accompanying dynamic cluster structure. Vibrato was the expression of forcefulness in the source domain, the source being the singing voice. Thus as the vocalist sung with more vibrato, the cluster included more notes and thereby became more dense. With less use of vibrato the cluster was thinning out, i.e. notes were dropped. The idea was based on Jazz Big Band composition technique. The change of clusters were controlled to take place metrically in time, so changes in cluster size were made with each new bar.

#### 4.2.2 Vocal composition

The compositional approach was to alternate between improvisation to generate material and a well-considered carving out of the final gestalt of the melody. At the time of composing the piece the singers were not yet casted. The score did not contain directions for where or how to vary the vibrato. Rather these were degrees of freedom left for the singer to decide upon.

### 4.3 Procedures

The application was tested during rehearsals and during the month of scheduled performances. The computer was placed with the sound engineer on a balcony so there was no visible technology in the concert space together with the vocalist.

The composer took autobiographical notes during the process of developing the tools, and from composing the piece. The performance was also documented by audio and videotape. Later on a written semi-structured interview with the singer was conducted. The sheet contained 14 open-ended questions that focused on the interplay with the computer in the domains of score, responsivity from the computer, the self and the other, the singing voice, time and listening. Examples of items are: “How was your first encounter with the score?” “Were the relationships to other fellow musicians affected and in that case, how?” How well did the responsivity work according to your experience?” The word “responsivity” was clarified on demand in relation to “interactivity”. The questions had a clear phenomenological emphasis. An advantage of the written interview was that answers would presumably be more thoroughly reflected upon.

#### 4.4 Analyses of interviews

According with Interpretative Phenomenological Analyses (IPA) procedure, after repeated close reading the interview-data was reduced and placed into categories in an iterative process between the parts and the whole. Different interpretations have been attempted in previous lectures and artist talks, where after the material again has been revisited. The result was also resent to the singer for feedback.

## 5. RESULT AND DISCUSSION

### 5.1 The singer’s account

The encounter with a new score may be regarded as an expression for the expectations that a musician brings into a musical project. Core elements of this encounter, which interestingly also could be traced throughout the work was that it seemed *exciting, challenging and uncertain*. The uncertainty was of course in part due to the electronic layer being at that point yet unknown to the vocalist; the score only contained notated clusters with possible sound material for each bar, but no timbres were known.

The relationship to the computer is interestingly ambiguous and paradoxical. The challenges of performing against a fixed beat are a recognised fact in the literature [15] [13]. This was also evident in this work and resulted in a sensation of being *controlled by the computer*. Another recurrent theme is that the *relationship to the computer evolved over time*, from being one denoted by reluctance to a rewarding one. The singer says:

“Jag upplevde den [datorn] som mot mig till en början. Men när jag vant mig och kommit in i det så kändes det mycket spännande och som att det gav mig något extra” (In the beginning I experienced the computer as going against me.

But once I got used to it and became involved it felt very exciting, and as if it provided me with something special). In this statement both the initial excitement and the sense of challenge can be traced.

Two other themes emerge dealing with the stance towards the computer. On one hand the computer is indeed experienced as a separate thing *detached from the performer*, with reference to the concept of the computer as the disembodied other (see for example [16] p. 168). This is well captured in the following:

”Ja, när vi gick från rep-situation till föreställning med publik upplevde jag att tempot gick ner. Jag tror att det beror på att när den nerv som har med nervositet att göra infann sig hos mig för att det blev skarpt läge, då behöll datorn sitt tempo och jag upplevde att tempot var långsamt, att det gick ner. Egentligen hade det bara med mig att göra. Det var en intressant upplevelse” (Yes, when we went from rehearsals to performing in front of an audience I experienced the tempo dropped. I believe it depends on that as the show was on, the nerve related to nervousness appeared, and the computer kept its tempo but I experienced the tempo as slow, as if it had decelerated. And still it only had to do with me. It was an interesting experience). The singer also describes how she feels uncertain and perceives the computer as unpredictable; again a theme that goes back to the expectations arisen when first encountering the score.

“Jag kände mig osäker på vad som kunde komma (i form av toner och kluster) och det ställde större krav på mig som sångare” (I felt uncertain about what was to appear (regarding notes and clusters) and it put higher demands on me as a singer).

She reports her listening strategies were affected with a heightened focus due to the uncertainty. On the other hand there is a strong *feeling of connectedness and a sense of cohesion*. For instance when she says the following:

[Datorn] “Inte en förlängning av min identitet men som att vi blev mer utav ”ett”. Och att den gav mig trygghet”. Och: “Jag var en del i ett sammanhang på ett starkare sätt än vid andra tillfällen. Dvs jag kände mig mindre ensam och mindre ”utsatt”” ([The computer] Not as an extension of my identity but as if we became more “one”. And that it provided me with comfort. Or: I was part of a context much more strongly than at other occasions. I.e. I felt less alone and less “exposed”). So the computer was a source of both comfort and uncertainty. Also the response was experienced as slow. This underlines an experience of the computer as separated, but also relates to the theme of being controlled by the computer.

### 5.2 Technical account

The application worked flawless during the whole month of performances and technical staff that did not have any knowledge in MAX/MSP managed its turning off and on. The vibrato extraction worked well and was robust after introducing the extra delays as described above.

## 6. CONCLUDING REMARKS

The *unpredictability* in the relationship to the computer the singer reports in the interview demands a closer inspection. First, it may be due to an arbitrary choice of scale in the mapping. In voice research the range of a classical singer's vibrato extent is on average about +/- 70 cent around the intended note [1]. Yet, as an integrated principle in use in a piece, which scaling should be used between in this case the extracted vibrato extent and the synthesis control? Is it linear? This also relates to the discussion regarding the extent to which a singer actually can vary the span of the vibrato [5]. Second, it may relate to the preconceptions of the singer. What formal training she or he has had and what musical culture he or she stems from heavily influences how the music is perceived and what one needs to be in control of in the moment. In music that has elements of improvisation, to not know how things will sound or have them different each time is part of the presuppositions for performing.

The two-sided experience of the computer is an interesting finding since in part it combines two different strands of thinking about the relationship to the computer in mixed works, i.e. the computer as separated from or the computer as a prolongation of the musician. In the current case it seems to provide comfort and uncertainty, detachment and cohesion simultaneously. Or it may be that it is a story told from the perspective of otherness, where the separation is a condition for feeling the sense of unity as reported. Furthermore the singer reported that the response from the application was experienced as slow: Here there was a negotiation between the compositional desire to have a stable shift of chords with each new bar and the performers need to have a more direct response. A different structure in the synthesis would have made the response more flexible and possibly more satisfactory.

We can conclude that vibrato detection is a powerful tool to use. What sounds are to be affected by the vibrato detection and even more importantly how they are affected in the time domain is yet an object of further inquiries. Preliminary results from an on-going project hints at the possibility of the structure of the sound and the length of the response to be important elements for determining how the relationship to the response is perceived.

## 7. REFERENCES

- [1] Friberg, A., et al. (2007). "CUEX: An algorithm for automatic extraction of expressive tone parameters in music performance from acoustic signals." *Acta acustica united with acustica*, 93(3): 411-420.
- [2] Einarsson, A., Sundqvist, M. (2012) *Ps. Jag kommer snart hem!*, Svensk Musik, Stockholm.
- [3] Prame, E. (1997). Vibrato extent and intonation in professional Western lyric singing. *Journal of the Acoustical Society of America*, 102(1), 616-621.
- [4] Sundberg, J. (1994). Acoustic and psychoacoustic aspects of vocal vibrato, Speech Transmission Laboratory. *Quarterly Progress and Status Reports*, vol. 35, no. 2-3, pp. 045-068, 1994.
- [5] Zangger Borch, D. (2008). *Sång inom populärmusikgenrer: konstnärliga, fysiologiska och pedagogiska aspekter*. Diss. Luleå Univ., 2008 Piteå.
- [6] Mitchell, H. F., Kenny, D. T. (2010; 2009). Change in vibrato rate and extent during tertiary training in classical singing students. *Journal of Voice*, 24(4), 427-434. doi:10.1016/j.jvoice.2008.12.003
- [7] Peretz, I. (2010). Towards a neurobiology of musical emotions. In Juslin, Patrik N. & Sloboda, John A. (Eds.), *Handbook of music and emotion: theory, research, and applications*, Oxford University Press, Oxford, 2010.
- [8] Regnier, L. and G. Peeters (2009). Singing voice detection in music tracks using direct voice vibrato detection. Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. *IEEE International Conference on, IEEE*.
- [9] Nwe, T. L., Li, H. (2007). Exploring vibrato-motivated acoustic features for singer identification. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(2), 519-530. doi:10.1109/TASL.2006.876756
- [10] Bartsch, M.A., Wakefield, G.H.: Singing Voice Identification Using Spectral Envelope Estimation. *IEEE Transactions, Speech and Audio Processing*, 12, 100-109 (2004)
- [11] de Cheveigné, A., & Kawahara, H. (2002). YIN, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4), 1917-1930. doi:10.1121/1.1458024
- [12] Fels, S., et al. (2002). "Mapping transparency through metaphor: towards more expressive musical instruments." *Organised Sound* 7(2): 109-126.
- [13] Emmerson, S. (2007). *Living electronic music*, Ashgate Publishing Company.
- [14] Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*, University Press.
- [15] McNutt, E. (2003). Performing electroacoustic music: a wider view of interactivity. *Organised Sound* 8(03): 297-304.
- [16] Emmerson, S., *Combining the acoustic and the digital: Music for instruments and computers or prerecorded sound*. The Oxford handbook of computer music ed. 2009: Oxford University Press.