Analysis and Synthesis of Vibrato in Lyric Singers

Ixone Arroabarren, Miroslav Zivanovic, Alfonso Carlosena

Universidad Pública de Navarra
Dpt. Electrical and Electronic Engineering
Campus de Arrosadía
E-31006 Pamplona, Navarra, Spain
carlosen@unavarra.es

ABSTRACT

In this paper two aspects of the vibrato signal characterisation are presented. First, an analysis method, which breaks down the musical signal into its different partials, and decomposes them into two separate contributions: AM and FM. This allows an objective characterization of the vibrato according to musical parameters. Second, a procedure for vibrato synthesis is proposed, based on the previous analysis. Even though the results of the synthesis are not fully satisfactory, they shed new light on the vibrato modeling.

1. INTRODUCTION

The analysis, characterisation and modelling of many kind of acoustical signals (voice, music, ..) has been a topic of interest to many researchers, in the areas of phsycoacustic, musicology, and signal processing. One specific topic of study, within this broad area, is the characterisation of lyric singers, and in particular the analysis of the vibrato, which is a vocal effect widely used. The interest in the vibrato dates back to the 30´s, where Seashore established a pioneering work introducing novel techniques and even devices to measure such vocal effect.

As it is the case in many other human-related analysis, automatic measurements purport to reproduce, or simply support, results that are usually obtained by subjective appreciation from experts, in this case musicians and/or people having musically educated listening. Automatic measurements should come to the same conclusions but in a faster and more objective manner, avoiding other external effects (room acoustics, noise,...) and even biased opinions.

The vibrato is essentially, at least as perceived by a listener, an almost periodic modification in the fundamental frequency of a musical note. While this variation is more evident in the case of an stringed instrument, it is not as clear in the human voice since such modification is accompanied by changes in the intensity and instantaneous amplitude of the sound [1]. The phenomenon is even more complex taking into account that any sound, in particular the human voice, contains in addition to the fundamental frequency a number of partials whose behaviour follow a complex pattern as the fundamental frequency varies.

The above described perception drives naturally, from the point of view of the signal, to a description and modelling in terms of its AM and FM components. However, it is well known that the problem of decomposing a signal into its AM and FM parts is ill posed, in the sense that there exist an unlimited number of possible combinations of such components. The very limited literature on the topic has been focused on the search of the frequency pattern of the vibrato, and there is only one remarkable reference that attempts an amplitude description [1]. A common limitation found in this literature is that the way in which such parameters are obtained is intuitively satisfactory but, to our view, not very sound. Moreover, the lack of information makes it difficult to fully reproduce the results obtained and thus the comparison with other approaches.

To cope with this situation, we proposed in [2] a method to obtain the FM description of the vibrato signal (i.e. the instantaneous frequency function), and subsequent procedures to obtain from it the three subjective parameters perceived by educated listeners: *intonation*, *extent*, and *rate*. The application of the procedures on several different professional female singers, performing the same musical score, clearly show their differences, which in addition correlate well with human perceptions. The corresponding instantaneous amplitude was also calculated but not exploited in the above-mentioned work.

The aim of this paper is to extend our previous work by proposing a characterisation of the amplitude variations associated to the vibrato. In this way, the vibrato effect can be fully modelled and parameterised, paving the way to synthesis procedures. We should note that, according to musical literature, even though a vibrato is essentially a frequency-modulated signal, amplitude variations are crucial to produce a natural and friendly perception of the sound [3]. The results obtained from the analysis procedures are subsequently used to explore a possible method for synthesis. Preliminary results are also included in this paper.

2. VIBRATO SIGNAL ANALYSIS

In the reference [2], we fully disclosed two methods to obtain Instantaneous Frequency (IF) and Instantaneous Amplitude (IA) for a musical note with vibrato, assuming the following model for each harmonic (i.e. partial)

$$s(t) = A_i(t)\cos\phi_i(t), \qquad (2)$$

being $A_i(t)$ the instantaneous amplitude and

$$f_i(t) = \frac{\partial\phi_i(t)}{\partial t} \qquad (3)$$

the instantaneous frequency.

Such methods are based on well known time-frequency techniques [4, 5, 6, 7] and both produce similar results for the IA with some differences for the IF. In general IF obtained from the STFT is a smoothed version of that obtained from the analytic function. As a mean of example we show here the results for a La5 performed by a famous singer. Figure 1 shows the long term spectrum, while figure 2 shows the instantaneous frequency and amplitude obtained for the isolated second harmonic, via STFT analysis. Similar figures are obtained for the remaining harmonics.

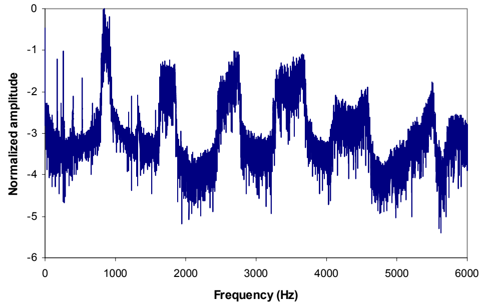

Fig. 1. Spectrum of the musical signal

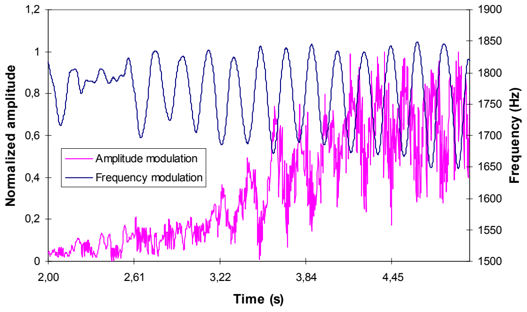

Fig. 2. AM-FM representation obtained from the Short Time Fourier Transform for the second harmonic

Concentrating now on the instantaneous frequency, it can be regarded as a nonstationary signal whose central value (*intonation* in a musical interpretation) should be equal to the frequency note. The peak amplitude is the maximum deviation from it (*extent*) and the frequency of the variation is identified as the *rate*. A typical, well performed vibrato and musically acceptable, is characterised by an extent of a quarter tone (i.e. 3% variation of pitch, or 50 cents) and a rate anywhere between 4 and 7 Hz, being more common between 5 and 6 Hz [8]. In order to identify these new three functions, the instantaneous frequency was modelled as:

$$f_i(t) = b(t) + a_v(t)\cos\phi_v(t) \qquad (4)$$

Looking at the long term spectrum of $f_i(t)$, shown in figure 3, $b(t)$ was associated to the low frequency content of IF, while the remaining spectrum was treated as a band-pass signal decomposed again in its instantaneous amplitude, $a_v(t)$, and frequency $f_v(t)$. Obviously $b(t)$, $a_v(t)$ and $f_v(t)$, correspond to the three musical parameters introduced above, *intonation, extent and rate* respectively and are shown, for the same example, in figures 4 and 5. The three are, in general, not constant, as corresponds to a non-stationary vibrato. We should stress here that these three functions are not unambiguously defined, and are thus application dependent. As long as we have obtained sensible results, our particular choice can be considered appropriate. Comparisons with previous works [9, 10], and informal listening support the validity of our approach. For more details on the calculation of these signals, readers are addressed to our paper [2].

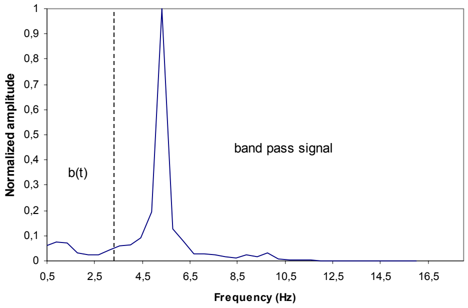

Fig. 3. Spectrum of the vibrato signal

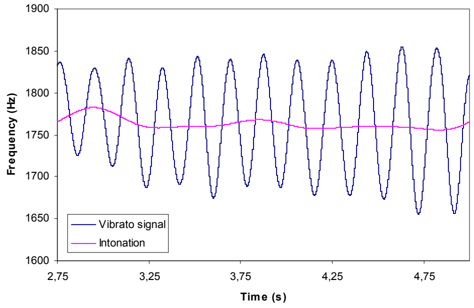

Fig. 4. Intonation of the vibrato signal

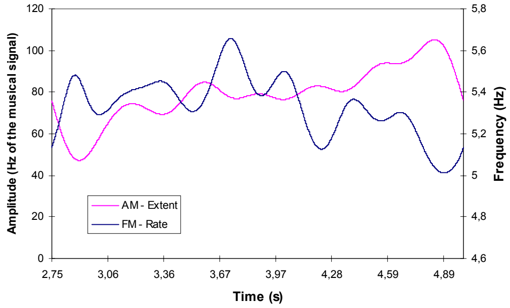

Fig. 5. Extent and rate of the vibrato signal

3. INSTANTANEOUS AMPLITUDE ANALYSIS

While vibrato is mainly characterized by an almost sinusoidal variation in the instantaneous frequency (see figure 2), the accompanying amplitude variation induced by the vocal tract is crucial in characterizing the musical signal. The main problem, from the analysis point of view, is that while for the frequency variation musically meaningful parameters can be easily identified (intonation, rate and extent), there is no evident connection regarding the amplitude variations.

Making use of the same example, it can be seen from figure 2 how amplitude variations of a single harmonic follow the same rate than the instantaneous frequency. A time frequency representation (STFT) of the first three harmonics, see figure 6, further shows that the relative phase between the two parameters is not the same for all harmonics. The reason for that is that each harmonic sweeps a different range of the vocal tract response. However, to establish a sensible connection between amplitude and frequency variations, the effect of signal energy must be considered too and eliminated. In figure 6 it is possible to see that besides the amplitude variation derived from the vocal tract response there is an energy increase with time, and that variation must be removed before relating amplitude and frequency oscillations.

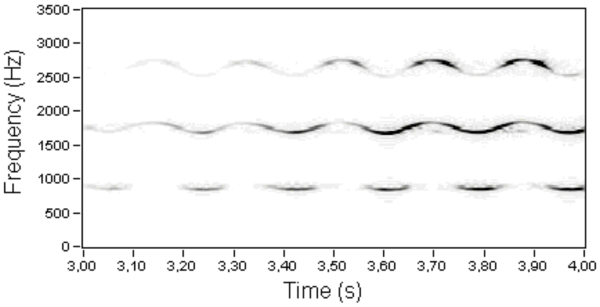

Fig. 6. Time frequency representation for the first three harmonics of the musical signal

To this end, the spectrum for the instantaneous amplitude, was calculated to see the low frequency contributions in order to subtract them from the signal. Instantaneous amplitude and its low frequency "pedestal" obtained in this way are shown in figure 7. This procedure has been repeated for each partial (harmonic), showing noticeable differences in the way the intensity of each harmonics grows with time. An example for a La5 is shown in figure 8.

All these considerations drive to a model where the instantaneous amplitude of (each harmonic of) the signal, $A_i(t)$, is represented as:

$$A_i(t)=\alpha_i\cdot\beta_i(t)\gamma_i(t) \qquad (5)$$

being α_i a constant representing the relative amplitude of the harmonic in comparison with the fundamental, β_i the time-dependent intensity and γ_i the amplitude variations induced by the vocal tract.

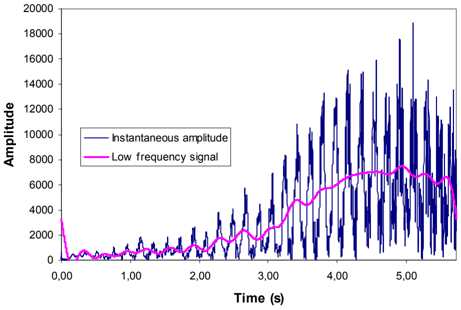

Fig. 7. Low frequency variation of the instantaneous amplitude

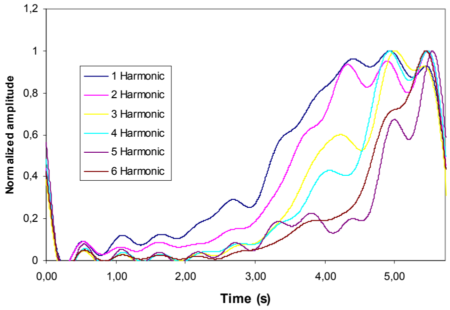

Fig. 8. Low frequency signal for the first six harmonics

We are now in the position to relate amplitude and frequency variations, that is to say, AM and FM components of the vibrato signal represented by *Ai(t)* and *fi(t)*. Since we assume that instantaneous frequency is the same for all harmonics, except for an integer scale factor, we have taken as reference for all harmonics the instantaneous frequency of the harmonic that has been obtained with the best resolution (the second harmonic in our example). Figure 9 shows the plain AM versus FM curves obtained for a number of harmonics, using time as parameter. A sort of spectral envelope results, which can be identified as the frequency response for the vocal tract when singing the vibrato. From this envelope, a linear model for the vocal tract, approximating the frequency response, may be obtained. In this way, each harmonic conveys information of an interval of the vocal tract response. This approach is different from previous works on spectral envelopes [11, 12], because of two reasons: the signal here considered is a FM modulated signal, what widens the bandwidth. On the other hand, the harmonics are spaced far apart which is worse for the characterizing methods explained in [11, 12].

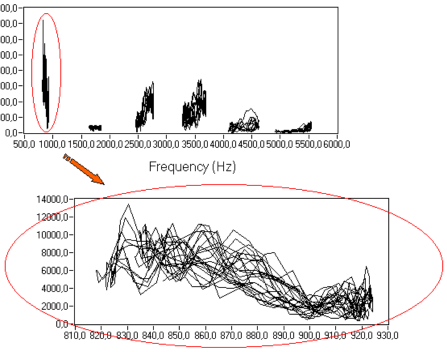

Fig. 9. AM versus FM for the first harmonic of the musical signal

4. AN APPROACH TO SYNTHESIS

The results shown in the paragraphs before allow us to envision a model for the AM component as a result of the amplitude variations induced by the vocal tract (modelled as a linear filter), excited at a given instantaneous frequency. In this way, the AM component is calculated by filtering a FM modulated sinusoid, whose modulating signal is generated according to the proposed model for the vibrato signal. In practical cases, where vibrato rate is small in comparison with local bandwidth of tract response, a simple AM-FM dependence (see figure 9) can be applied, instead a more complex convolution, to generate the Instantaneous Amplitude from the FM. All harmonics are properly weighted and added afterwards. This procedure can be regarded as a sort of Additive Synthesis [13, 14, 15], but including two remarkable difference*s*: First, all of the functions involved (equations 4 and 5) can be generated either from the analysis procedure (i.e. re-synthesis) or can be modelled as simple mathematical functions: for instance a pure sine for the instantaneous frequency. And second, all parameters that can be modified have a musical meaning, especially those of the instantaneous frequency. Either way, the synthesis procedure described above is schematically shown in figure 10.

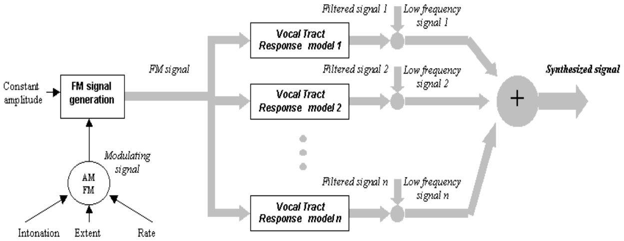

Fig. 10. Block diagram of the synthesis method

We have synthesised, according to the above procedure, several artificial vibratos by i) approximating the signals obtained from the analysis by simple trigonometric functions, ii) and also by using exactly the signals resulting from the analysis. In both cases the subjective perception of the vibrato (informal listening) is very similar, and quite close to the original recording. In other words, a listener can appreciate all the vibrato parameters in both cases. However a lack of realism is appreciated in the synthesised signals particularly at the end of the score.

5. DISCUSION OF RESULTS AND CONCLUSIONS

From the results shown above, it is obvious that if the signal resulting from the re-synthesis of the components (intonation, rate, vocal tract response…) arising from the analysis is significantly different from the original record, then something is missing in the analysis procedure. In other words, all of the parameters calculated for the musical signal, (for the instantaneous amplitude and instantaneous frequency) can be identified by a listener in a real musical signal and in a synthesized signal. But in the case of the synthesized signal they are not enough to build from them a convincing musical signal with vibrato.

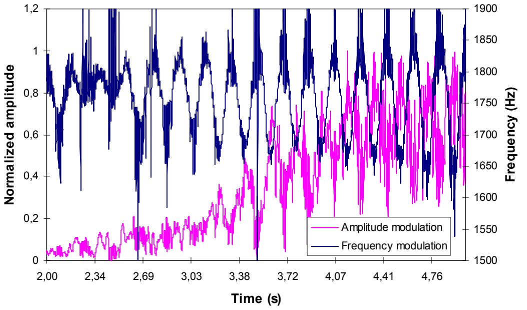

Fig. 11. AM-FM representation obtained from the analytic signal for the second harmonic

In an attempt to find where relevant information has been lost, a re-synthesis of the signal has been made, but recomposing the AM and FM components obtained from the two methods mentioned in the Introduction (the STFT and the analytic signal). It has been seen that in contrast to peak detection based STFT procedures, the analytic signal contains all the information of the original signal. Actually, by comparing figure 11 with figure 2, it is obvious that while the AM are indistinguishable, the FM in figure 11 shows a more rugged profile. This can be interpreted as an ill-defined instantaneous frequency (particularly at some instants), which is typical of signals composed of more than a single harmonic [7, 16, 17]. This is the reason why research is now in progress to define a more accurate model of the FM component, which takes into account a multiplicity of harmonics for each partial.

6. ACKNOWLEDGEMENTS

The Gobierno de Navarra and the Universidad Pública de Navarra are gratefully acknowledged for financial support.

7. REFERENCES

[1] R. Maher, J Beauchamp, "An investigation of vocal vibrato for synthesis", Applied Acoustic , 30, pp. 219 - 245, 1990
[2] I. Arroabarren, M. Zivanovic, J. Bretos, A. Ezcurra , A. Carlosena, "Measurement of Vibrato in Lyric Singers", IEEE Instrumentation and measurement technology Conference, pp. 1529-1534, May 21-23, 2001, Budapest, Hungary
[3] S. Mc Adams, X. Rodet, "The role of FM-induced AM in dynamic spectral profile analysis", http://mediatheque.ircam.fr /articles/textes/McAdams88a
[4] B. Boashash, "Estimating and Interpreting The instantaneous Frecuency of a signal. Part 1: Fundamentals", Proceedings of the IEEE, Vol 80, nº4, pp. 519-538, April 1992
[5] Lovell, Williamson, Boashash, "The relationship Between Instantaneous Frecuency and and Time Frecuency Representations", IEEE Transactions on signal Processing, Vol 41, nº 3, pp. 1458-1461, March 1993
[6] L. Cohen, "Time-Frecuency analysis", Prentice Hall, 1995
[7] P. J. Loughlin, B. Tacer, "On the amplitude and frecuency modulation decomposition of signals", Journal of the Acoustical Society of America, Vol. 100, pp. 1594-1601, September 1996
[8] I. Titze, "What's in a voice", New Scientist, 23 September 1995
[9] E. Prame, "Vibrato extent and intonation in professional Western lyric singing", Journal of the Acoustical Society of America , Vol. 102, nº 1, pp. 616-622, July 1997
[10] E. Prame, "Measurements of the vibrato rate of ten singers", Journal of the Acoustical Society of America, Vol. 96, nº 4, pp. 1979-1984, October 1994
[11] D. Schwarz, X. Rodet, "Spectral envelope estimation and representation for sound analysis-synthesis", Proceedings of the International Computer Music Conference 1999
[12] M. Campedel-Outdot, O. Cappé, E. Moulines, "Estimation of the spectral envelope of voiced sounds using a penalized likelihood approach", IEEE Transactions on speech and audio processing, Vol. 9, nº 5, pp. 469-481, July 2001
[13] X. Rodet, P. Depalle, "Spectral envelopes and inverse FFT synthesis", Convention of the audio engineering society. San Francisco, AES 1992
[14] X. Serra, "Musical Sound Modeling with Sinusoids plus noise", http://www.iua.upf.es/~xserra/articles/msm/, 1997
[15] X. Rodet, "Musical sound signal Analysis/Sinnthesis: Sinusoidal+Residual and elementary waveform models", http://mediatheque.ircam.fr/articles/textes/Rodet97, 1997
[16] W. Nho, P. J. Loghlin, "When is Instantaneous Frequency the average frequency at each time?", IEEE Signal Processing Letters, Vol. 6, nº 4, pp. 78-80, April 1999
[17] P. M. Oliveira, V. Barroso, "Instantaneous Frequency of multicomponent signals", IEEE Signal Processing Letters, Vol. 6, nº 4, pp. 81-83, April 1999