INSTANTANEOUS FREQUENCY AND AMPLITUDE OF VIBRATO IN SINGING VOICE

Ixone Arroabarren, Miroslav Zivanovic, Xavier Rodet, Alfonso Carlosena*

Universidad Publica de Navarra
Dept. Electrical and Electronic Engineering
Campus de Arrosadia, E-31006 Pamplona, Spain
* IRCAM, 75004 Paris, France

ABSTRACT

In this paper we investigate the relationship between the Instantaneous Amplitude (IA) and the Instantaneous Frequency (IF) of vibrato Signals in singing voice. It is shown that this relationship is of great value to obtain information about the vocal tract model. However, to make this analysis possible it is necessary to cope with two basic limitations: reverberation in recordings, which shows up as multiharmonicity in each partial, and phase effect of vocal tract formants which distort the instantaneous frequency of some partials.

1. INTRODUCTION

Vibrato in singing voice is traditionally modeled as a frequency and amplitude modulated signal. Though it is not completely evident how both kind of modulations, related in some way by the vocal tract response, are perceived by humans, most research efforts have been concentrated on the analysis of the FM component only. Thus, assuming that partials produced in the singing voice are harmonically related, a number of procedures have been proposed to calculate the Instantaneous Frequency of the signal by tracking the time evolution of pitch or by making use of a variety of time-frequency techniques. From the information carried by the IF signal, a number of acoustic parameters are obtained, namely intonation, extent an rate, which are used to characterize the quality of the vibrato performed by a given singer [1, 2].

From this analysis viewpoint, the differences given by the distinct methods in the calculation of the instantaneous frequency, have a negligible influence in the musical parameters mentioned above. Moreover, it is not so clear to which extent an educated listener is able to resolve within one of those parameters values. However, things are different when the information, obtained not only from the IF but also the vibrato signal itself, is used for vocal tract modeling and thus for re-synthesis purposes. In these latter cases, it is mandatory to have a more precise information of the instantaneous frequency and in cases where it is ill-defined, what the causes are and then, how the signal needs to be modeled. At the same time, variations in the IA depend not only on the vocal tract response but also on the source intensity variation.

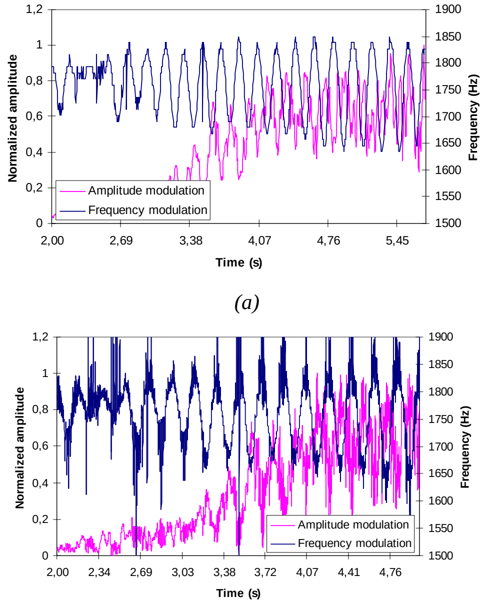
These two problems, i.e. the difficulties of obtaining a sensible FM and AM representation have been treated by authors in [2, 3]. In this paper we will seed new light on these problems, and their solution, and will show how the combined analysis of IA and IF may serve to obtain very useful information about the vocal tract response. Vibrato in singing voice will be seen as a paradigm of the fascinating topic of IF.

The organization of this communication is as follows: first, in section 2, we show the limitations of the AM/FM decomposition given by the application of time-frequency techniques to arbitrary recordings, and thus the difficulties to obtain from this information a sensible representation of the vocal tract. In section 3 the study of anaechoic recordings will be presented. Here the relationship between the IA-IF interplay and the vocal tract response will be analyzed again and shown to be more evident. The main conclusions of the work are summarized in section 4.

2 . SIGNAL ANALYSIS

From the signal processing point of view a sung vibrato is a multicomponent signal [4, 5], because it is composed by several harmonics, or partials that characterize the timber. Also, it is a non-stationary signal because its spectrum strongly varies along the time. The first idea was to study the non stationary nature of the signal, in terms of its instantaneous frequency; and so the signal was decomposed into its components, and each one of these was analyzed by Time Frequency tools [2]. This approach is different from the conventional employed in the additive synthesis literature [6] where, a Short Time Spectrum of the signal is calculated, and every harmonic of the sound is detected at a given time.

Readers are addressed to reference [3] for the details of the calculation, and here we will restrict to some particular, but representative, examples to highlight the main limitations of the procedures



(a)

(b)

Fig. 1. IF and IA decomposition for using (a) STFT and (b) Analytic Signal

The results shown in Fig. 1a and 1b compare the IF calculated for a single harmonic calculated from the STFT and the Analytic Signal model. It is evident that the IA and IF given by the STFT are smoothed versions of those obtained by the analytic signal, being the origin of the difference unclear. In [3] this difference was investigated by re-synthesizing the vibrato from the extracted IA and IF. Thus, each component was calculated from IA and IF as:

$$s(t) = A(t)\cos\varphi(t) \qquad (1)$$

where

$$\varphi(t) = 2\pi\int_{-\infty}^{t} f(\tau)d\tau \qquad (2)$$

being *A(t)* and *f(t)* the IA and IF respectively.

The outcome was that when the analytic signal's results were used, the synthesized component was perceived as the original recording. Otherwise, when STFT's results were utilized the sound was different and less natural. From this observation it follows that something is missing in the analysis by the STFT, and that its inherent smoothing eliminates information which is relevant from the listener point of view.

According to our investigations, the reasons for the discrepancy is the fact that each harmonic (partial) is actually a bunch of (non-stationary) close tones originated from the reverberation of the original sound. It makes sense that each echo path produces an additional harmonic that, since instantaneous frequency varies with time, has a different instantaneous frequency than the original one. Thus, each method (STFT an Analytic Signal) treat in a different way this situation. The IF given by the STFT calculates a sort of smoothed mean value, while the IF defined from the analytic signal gives ill defined IF values which can be even out of the bandwidth bounded by the bunch of tones, as can be seen in Fig. 1.b. This last phenomenon has been described and analyzed in the literature on the instantaneous frequency in a multicomponent signal [7, 8].

In reference [9] Maher and Beauchamp showed a kind of local spectral envelope representing the vocal tract by an AM versus FM representation. However, and according to our findings, this is not so evident for an arbitrary recording with reverberation. We have tried this representation for the example signal, and the results, for a single harmonic, are shown in Fig. 2. It is possible to guess the shape of the vocal tract response for that frequency rate, but the typical artifacts described in the previous paragraphs disguise the representation. This means that for signals characterized by a bunch of tones around the partial frequency, whose origin is reverberation, it is not straightforward to obtain a clear representation for the vocal tract response.

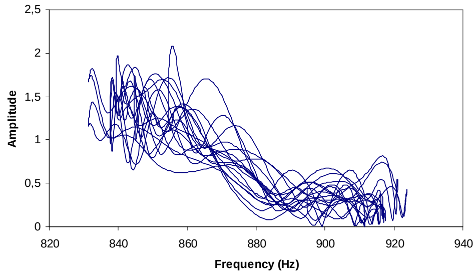


Fig. 2. IA versus IF representation for a commercial recording of a vibrato

3. ANAECHOIC RECORDING ANALYSIS AND AM FM RELATIONSHIP

The best way to avoid reverberations is to consider anaechoic recordings. Therefore, the same analysis proposed in section 2 has been applied to this kind of signals. In Fig. 3. IF is shown for the first harmonic of a tenor vowel, where it is possible to see that the two kind of analysis (method proposed in [3] and Sinusoidal Additive Synthesis) give almost the same results.

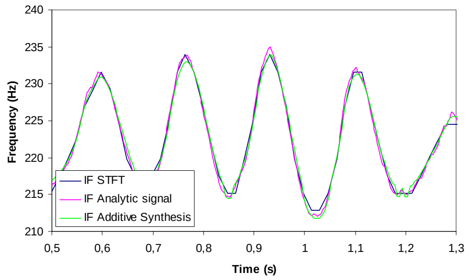


Fig. 3. IF of a single harmonic for an anaechoic vibrato recording

A similar conclusion follows from the IA analysis. There are slight differences at some time instants but they are due to resolution problems in the STFT. So, it is possible to claim that the IF of a component is the IF of the harmonic (partial), and it is also the same as the IF provided by the analytic signal.

Now, we are in the position to analyze the relationship between the IF and IA, and the extracted vocal tract response since it is dependent only on the voice and not on other external factors. The voice has traditionally been viewed as a linear source/filter system. This is, there are one or more sound sources, and a bank of filters that shape the spectrum of those sound sources. The voice is usually characterized as a periodic signal corresponding to the oscillation in vocal folds, or as non-periodic source corresponding to turbulent noise, or a mixture of these. The voice system filter properties are controlled by the shape of the vocal tract. In the case of voiced sounds, as is the case, the excitation is a periodic source with harmonically related components, and the filter is composed by a set of resonances, or formants. Thus, the location and shapes of formant resonances are strong perceptual cues that are used to identify vowels and consonants [10].

One of the goals of this work is to show how the formant information, and also a complete vocal tract model, can be extracted from the joint analysis of the AM-FM components of the vibrato. As depicted in the previous section, the most straightforward approach consist in represent the IA versus the IF for each harmonic, and using time as parameter. We saw the difficulties originated from the fact that each partial is composed of several closely located harmonics, caused by reverberation. Now, this effect is not present owed to the use of anaechoic recordings but we will have to deal with the phase effect of formants. In order to make clear this problem some simulations will be shown first.

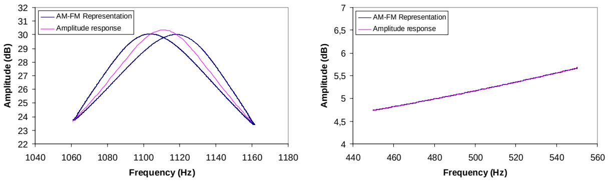
In many cases, particularly in singing synthesis, the vocal tract response has been modeled as an all pole filter, and its formants as second order filters [11, 12]. Thus, it is interesting to model the behavior of one frequency modulated harmonic exciting such formant model (i.e. filter). The filter output is then analyzed as in the real case (see section 2), and its IA and IF have been calculated.

The model for the vocal tract response is as follows:

$$H(z)=\frac{g}{1-a_1z^{-1}-a_2z^{-2}} \qquad (3)$$

being the bandwidth of the filter 50Hz and its central frequency 1000Hz. The excitation signal is FM modulated by a sinusoid whose amplitude is the 10% of its central frequency (100Hz peak to peak for a signal's central frequency f_{cf}*=1000Hz*), and whose rate is of 5,5 Hz [1, 2]. The values of the frequency modulation are typical values of the human vibrato in singing voice, and the bandwidth of the filter is also a common value for a real formant [13]. The effect of the phase response of the filter is evident in fig. 4a, where we represent IA versus IF at the output and compare it with the true filter shape.

Otherwise, if the excitation's central frequency were situated far from the central frequency of the filter (e.g. f_{sc}*=700Hz*), this effect would not be relevant, and the IF would be the same at the input and output of the filter, and the AM-FM representation would coincide with the magnitude response of the filter. This is represented in Fig. 4b.



(a) *(b)*

Fig. 4 AM-FM representation for (a) tone within a formant and (b) far from a formant

In real signals neither the original IF nor the magnitude response of the vocal tract are available. Also the source amplitude variation has to be taken into account [3], while in the simulation this effect was not considered. Anyway, it is possible to see if this phase effect appears in the real signal: if there is a harmonic whose frequency is close to the central frequency of a formant, the phase effect in its IF will be stronger than in the case where this frequencies are far from each other. In a low pitched signal it is easy to know if this happens or not. As an example, in Fig. 5, the AM-FM representation for twenty harmonics of an anaechoic recording is represented.

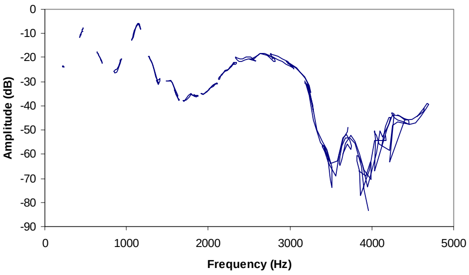
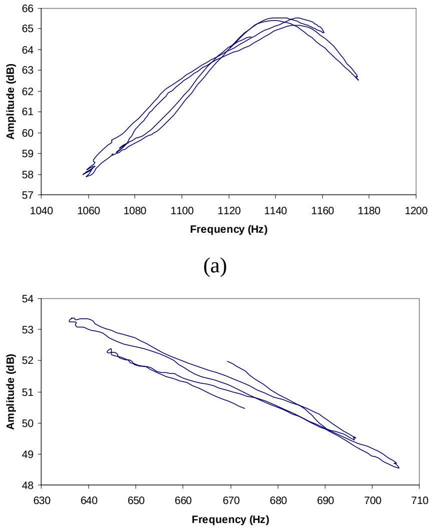


Fig.5. Vocal tract representation through AM-FM representation (Anaechoic recording)

To avoid the source amplitude variation, while preserving the formants phase effect, only a cycle of vibrato is considered, assuming that during that time slot the intensity of the sound is almost constant. In this signal the fifth harmonic is very close to the central frequency of the second formant, while the third harmonic is far from this frequency, and from the central frequency of the first formant. So, it is worthwhile to study the IA and the IF of these two harmonics.

In Fig. 6.a and Fig.6.b, the details of the AM-FM representation for the fifth and third harmonic are shown. Although the intensity of the harmonic is not constant, it seems clear that in the fifth harmonic the phase effect appears, because in Fig.6.b the IA belonging to each represented semicycle of vibrato are parallel, while in Fig. 6.a this does not happen.



(a)

(b)

Fig. 6. Vocal tract representation details for the fifth (a) and third (b) partial (Anaechoic recording)

In order to remark the effect, we show in Fig. 7, the normalized IF of the first, third and fifth harmonic. In this case the first two are almost identical whilst the IF of the fifth harmonic is different.

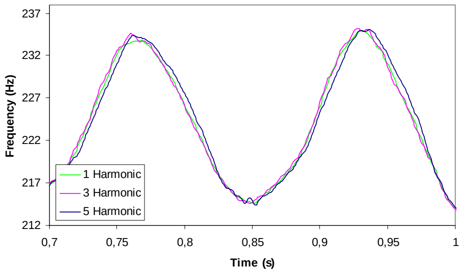


Fig. 7. Phase effect in the IF of the harmonics

Considering that the first and the third harmonics are far from the central frequency of the formants we can conclude that the phase of the vocal tract response affects to the IF of the signal's harmonics, but only when the frequency of the harmonic is close to the central frequency of the formant. Anyway, this effect is not very relevant because the rate of vibrato is slow compared with the bandwidth of the formants. This means that, provided that anaechoic recordings are available, the analysis of vibrato allows a sensible extraction of the vocal tract response thanks to the frequency modulated nature of that vocal effect, as demonstrated by the results in Fig. 5.

4. CONCLUSIONS

The use of time-frequency techniques for the analysis of vibrato in lyric singing has made clear that the decomposition of such vibrato in its AM-FM components may suffer from some limitations, when used to model the vocal tract response, and thus for resynthesis purposes. The use of anaechoic recordings is mandatory to avoid reverberation which is manifested in a multiharmonicity of partials, and thus in a meaningless calculation of the instantaneous frequency. Once anaechoic recordings are available, some other artifacts need to be corrected such as source amplitude variation (not analyzed in this paper) and the phase effect of formants that distorts the calculation of the instantaneous frequency. Taking into account all this facts, the main conclusion of the paper is that vibrato in human voice can be advantageously used to obtain a satisfactory representation of the vocal tract response.

6. REFERENCES

[1] 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

[2] I. Arroabarren, M. Zivanovic, J. Bretos, A. Ezcurra , A. Carlosena, "Measurement of Vibrato in Lyric Singers", *IEEE Transactions on instrumentation and meassurement*, Vol. 51, nº 4, August, 2002

[3] I. Arroabarren, M. Zivanovic, A. Carlosena, "Analysis and Synthesis of vibrato in Lyric singers", *Proceedings of the European Signal Processing Conference*, September 3-6, 2002, Toulouse, France

[4] B. Boashash, "Estimating and Interpreting The instantaneous Frequency of a signal. Part 1: Fundamentals", *Proceedings of the IEEE*, Vol 80, nº4, pp. 519-538, April 1992

[5] B. Boashash, "Estimating and Interpreting The instantaneous Frequency of a signal. Part 2: Algorithms and applications", *Proceedings of the IEEE*, Vol 80, nº4, pp. 539-568, April 1992

[6] X. Serra, "Musical Sound Modelling with Sinusoids Plus Noise", *Musical Signal Processing*, 1997

[7] P. Loughlin, B. Tacer, "Comments on the interpretation of Instantaneous Frequency", *IEEE Signal processing Letters*, Vol 4, nº 5, pp. 123-125, May 1997

[8] P. M. Oliveira, V. Barroso, "Instantaneous Frequency of multicomponent signals", *IEEE Signal Processing Letters*, Vol. 6, nº 4, pp. 81-83, April 1999

[9] R. Maher, J. Beauchamp, "An investigation of vocal vibrato for synthesis", *Applied Acoustic*, 30, pp. 219 - 245, 1990

[10] P. Cook, "Toward the Perfect Audio Morph? Singing Voice Synthesis and Processing", *Proceedings of the Digital Audio Effects Workshop*, November 19-21, 1998, Barcelona, Spain

[11] L. Rabiner, "Digital Formant Synthesizer", *Journal of the Acoustical Society of America* , Vol. 43, n °4, pp. 822-828, 1968

[12] G. Bennett, X. Rodet, "Synthesis of the singing voice", *Current directions in computer music research*, pp. 19-44, 1991

[13] X. Rodet, "Time domain formant wave function synthesis", *Computer music journal*, Vol 8, n°3, pp. 9-14, 1984