MODELLING OF VIBRATO PRODUCTION

Ixone Arroabarren, Alfonso Carlosena

Universidad Publica de Navarra
Dept. Electrical Engineering and Electronic
Campus de Arrosadia, E-31006 Pamplona, Spain
ixone.arroabarren@unavarra.es, carlosen@unavarra.es

ABSTRACT

In this paper, a novel Non-Interactive Source-Filter model for singing voice, incorporating vibrato, is proposed. In this way the periodic frequency variation, peculiar to vibrato is included in a simplified signal model for singing voice production. The inclusion of vibrato will further allow to relate two distinct signal models for the voice: The Non Interactive Source-Filter and the Sinusoidal Models. In this way, the Instantaneous Amplitude and Instantaneous Frequency of the harmonics, which are intrinsically Sinusoidal Model parameters, will be related to the glottal source and vocal tract response, typical of the Source-Filter model.

Thanks to the proposed model, the most relevant acoustic features of vibrato will be related to the voice production mechanisms.

1. INTRODUCTION

Singing voice is one of the less studied vocal expressions, because it is far apart from traditional speech applications. Conversely, it represents an interesting challenge to the study of voice quality because of its differences from every day speech. Among the differences between singing voice and speech production, the most interesting one is perhaps the vocal vibrato. This is an specific musical feature not present in speech, and has been by itself the topic of interest of many researchers in the areas of physiology and musicology.

It is quite easy to describe vocal vibrato from the acoustical point of view, according to Sundberg's definition [1]: "vibrato is a regular fluctuation in pitch, timbre and/or loudness"; however, some basic aspects remain enigmatic still today. From this definition, it is clear that vibrato is a regular fundamental frequency variation, being this feature the most studied aspect. The most relevant works are those related to the parameterisation of the fundamental frequency variation [2, 3].

Going back to the former definition, it does not explain what happens with timbre and loudness during vibrato. The position of the sound harmonics (partials) is harmonically related to the fundamental frequency, and all of them will show the same regular fluctuation. Additionally, the amplitude of the partials also shows temporal variations during vibrato, being its origin and characterization not completely identified. Several works have addressed this problem [4, 5, 6], but apart from slight differences, almost all of them propose means to measure these amplitude variations and, at best, argue about their perceptual relevance. However, their relationship with voice production feature remains unclear.

In this context, the main goal of this work is to propose a model for vibrato, where the most relevant acoustic features (frequency and amplitude variations) will be linked to voice production mechanisms. Therefore, besides to measuring the most relevant acoustical effects of vibrato, the cause of these effects will be tackled through the voice production modelling. To that end, the Source-Filter model [7] and its environment will be taken as the voice production modelling, because it is simpler than physical models and it will make easier to link the voice production features with the amplitude and frequency variations, as obtained by a Sinusoidal Model.

The organisation of the work is as follows, in section 2 two signal models will be reviewed: the Sinusoidal and the Source-Filter models, which will characterise basically the effect and the cause of vibrato respectively. In section 3 vibrato production will be tackled, and a vibrato production model will be proposed. Finally, in section 4, this model will be evaluated, and the conclusions that follow will be given.

2. SIGNAL MODELS

2.1. Non Interactive Source Filter model

This is a simplified voice production model, represented by the block diagram in Fig.1. Singing voice production is modelled by a glottal source excitation that is linearly modified by the Vocal Tract Response (VTR) and the lip radiation diagram. Typically, the VTR is modelled as an all-pole filter, and relying on the linearity of the model, the lip radiation system is usually combined with the glottal source, in such a way that the Glottal Source Derivative (GSD) is considered as the vocal tract excitation.

The glottal source is periodic in voiced sounds, and it is characterized by five independent parameters, fundamental frequency, F_o, amplitude of voicing, Av, open quotient, O_q, asymmetry coefficient, α, and return phase interval, which is related to the spectral tilt, f_c, [8]. This signal will control the vocal texture while the VTR determines the sung vowel.

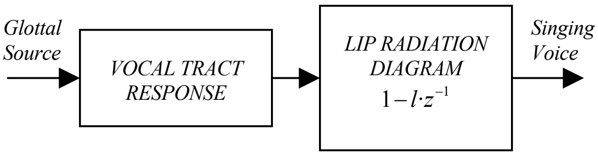


Fig. 1. Non interactive source filter model

In case of vibrato production, a question arises of how VTR and GSD parameters evolve during vibrato. From the analysis point of view, Inverse Filtering techniques have been proposed in order to split both elements [9, 10]; however these techniques are transparent to vibrato because they are based on short time window analysis, compared to the slow fundamental frequency variation typical of vibrato. In Fig. 2 an example is shown of such technique results, corresponding to the approach proposed in [10]. The VTR curve shows in the 2500–3000 Hz frequency region a set of close peaks typical of the singers formant in male voices [1].

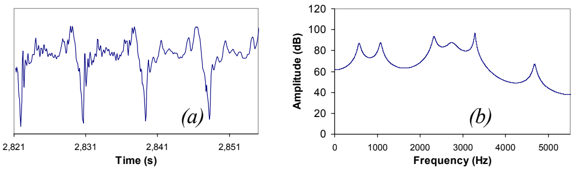


Fig. 2. Inverse filtering results. *(a)* GSD, *(b)* VTR. *Baritone recording, Vowel "a", F_o= 106 Hz*

2.2. Sinusoidal model

As it has been mentioned, vibrato can be also defined as a set of sinusoidal components (partials), whose frequencies and amplitudes change along time, and as a result, it has traditionally been described by the Sinusoidal Model. This signal model is naturally characterized by the Instantaneous Amplitude $a_i(t)$ (IA), and Frequency $f_i(t)$ (IF), of the harmonics, plus a stochastic residual $r(t)$, which is modelled by a time-varying spectral density,

$$s(t) = \sum_{i=1}^{M} a_i(t)\cos\theta_i(t) + r(t) \qquad (1)$$

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

In [11], detailed information is given about the measurement of these instantaneous magnitudes. Here, we will limit ourselves to remark what can be inferred about vibrato production from Sinusoidal Modelling results. It is qualitatively known that the IA of the harmonics is a consequence of their IF variation and the effect of the VTR. Several research efforts have been driven to establish the perceptual relevance of this feature [4, 5], and to find a more quantitative relationship between IF and IA [12]. All those works have in common the way in which they use the IA. In [12], the AM-FM representation is defined as the instantaneous amplitude versus instantaneous frequency representation, with time being an implicit parameter. This representation is compared to the magnitude response of an all pole filter, which is typically used for VTR modelling, and it is concluded that only when anechoic recordings are used, these two representations can be fairly compared. Otherwise, reverberation will corrupt the instantaneous magnitudes estimation. Fig. 3 constitutes a good example of the kind of AM-FM representations we are talking about. On it, each harmonic's instantaneous amplitude is represented versus its instantaneous frequency. In the figure only two vibrato cycles, where the vocal intensity does not change significantly, have been depicted.

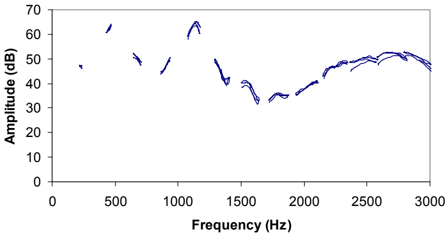


Fig.3. AMFM Representation for the first 20 Harmonics. Anechoic Tenor recording. *F_o= 220 Hz, Vowel "a"*

As the number of harmonic increases, the frequency range swept by each harmonic widens. Comparing Fig. 3 to Fig. 2.b, the AM-FM representation is very similar to the VTR of Fig. 2.b. However, in the case of the AM-FM representation, no Source-Filter separation has been made, and thus both elements are melted in the representation.

3. NON INTERACTIVE SOURCE-FILTER MODEL WITH VIBRATO

In this context, a novel Non Interactive Source-Filter Model with vibrato is proposed as a signal model that links the acoustical features of vibrato (frequency and amplitude variations) with the GSD and VTR typical of voice production. We will first make some basic assumptions regarding what is happening with GSD and VTR during vibrato, which are based on perceptual aspects of vibrato, and on the AM-FM representation for natural singing voice. Assuming that the sound intensity does not change:

1. The GSD characteristics remain constant during vibrato, and only the fundamental frequency of the voice changes. This assumption is justified by the fact that perceptually there is no phonation change during a single note
2. The VTR almost does not change along with vibrato. This assumption supports in the fact that vocalization does not change along the note

Taking into account these assumptions, the simplified Non Interactive Source filter Model with Vibrato could be represented by the block diagram in Fig. 4:

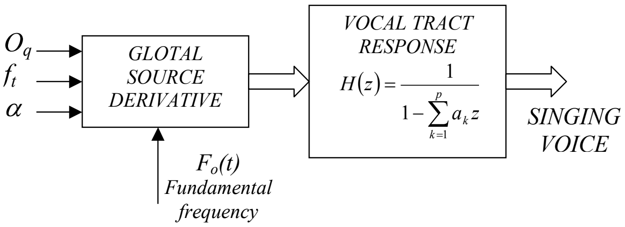


Fig. 4. Non interactive Source Filter Model with vibrato

This model anticipates that a long term relationship can be established between Inverse Filtering and the AM-FM representation: Taking into account that GSD features remain constant during vibrato, the AM-FM representation of each harmonic will represent a local section of the VTR, and each representation will be shifted in amplitude depending on the GSD spectral shape. Therefore, by removing the GSD effect from the AM-FM representation only the VTR will be represented. This relationship is shown in Fig. 5 for a synthetic signal described by the proposed model. The GSD has been modelled by the LF model [13]. The vocal tract response filter corresponds to vowel "a" of a baritone, and F_o= *100 Hz*:

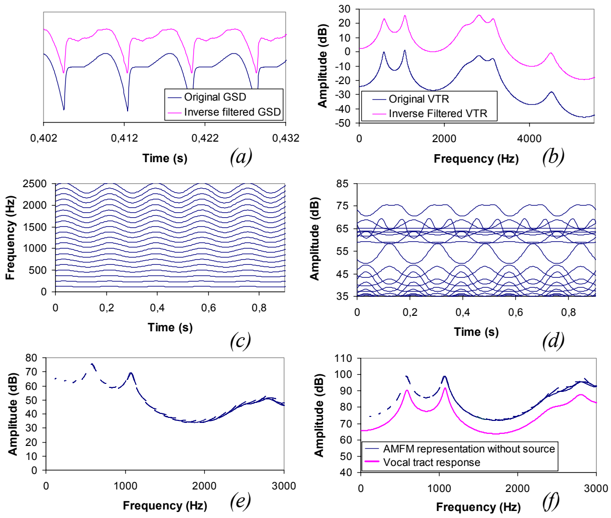


Fig.5. Synthetic signal. I. Filt. Results. *(a) GSD (b) VTR*. S. Modelling results *(c) IF (d) IA. (e) AMFM representation (f) AMFM representation without source effect*

In Fig. 5 the analysis procedure is shown: on the one hand, the Inverse Filtering approach proposed in [10] has been used in order to separate GSD and VTR, (*(a)* and *(b)*). On the other hand, Sinusoidal modelling has been used and the IF and IA of the harmonics have been estimated. In *(c)* the frequency variation of the harmonics is evident, and in *(d)* the resulting amplitude variation is shown. By representing IA versus IF and comparing *(e)* and *(b)*, the differences are evident because in the AM-FM representation no Source Filter deconvolution is applied. However, by comparing *(f)*, where the source effect has been removed using the spectral features of *(a)*, to *(b)*, it is clear that both analysis provide very similar vocal tract information.

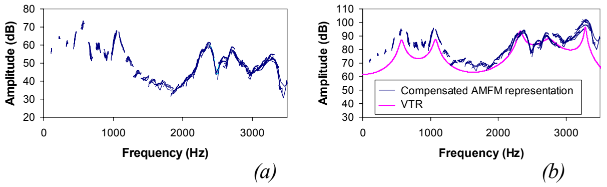


Fig. 6. Baritone recording. *Vowel "a", F_o= 106 Hz. (a)* AMFM representation *(b)* AMFM representation without source effect

In order to check how close the new vibrato model characterises natural vibrato production, the above mentioned procedure has been applied in natural singing voice recordings. Therefore, the same analysis explained for the synthesized signal has been applied to the baritone recording shown in subsection 2.1. Comparing Fig. 6 to Fig.5, which is representative of natural singing voice results, it is possible to say that both signals' analysis provide very similar results, what can validate the model's assumptions about natural singing voice. In other words, it might be concluded that the GSD and VTR do not change so much along with the fundamental frequency variation typical of vibrato.

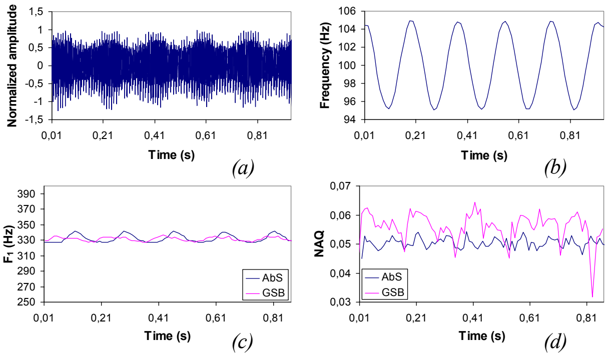
4. EVALUATION OF THE MODEL

In the last section a non interactive model with vibrato has been proposed in order to relate the most relevant acoustical features of vibrato to the voice production elements. As a result, it has been shown that the amplitude variations of the harmonic are a result of the frequency modulation of the GSD and the filtering effect of the VTR. The key point of the model, as it is clear from the analysis, is the fact that the GSD and VTR features, (O_q, α, f_c) and central frequencies of the formants respectively, remain constant during vibrato.

In order to confirm these assumptions the proposed model will be evaluated and further analysis will be developed, where synthesized signals along with natural singing voice recordings are compared. As it has been mentioned, Inverse Filtering analysis is a short time analysis, and does not take into account the slow fundamental frequency variations typical of vibrato. Thus, in this section, these techniques are going to be applied in such a way that the analysis window will be moved along time. In this way, and using the appropriate parameterisation for the VTR and GSD, we will see how these elements behave during vibrato. It must be stressed that Inverse filtering techniques have a fundamental frequency dependence [10], and in order to avoid measurement errors synthetic and natural singing voice signals will be analysed.

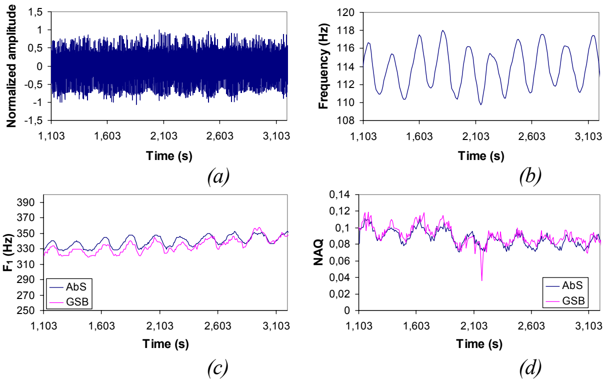
In this situation an apposite parameterisation must be selected for both elements. For the VTR the central frequency of its two first formants is selected, because it is straight forwardly obtained from the Inverse Filtering results. However, this analysis provides a waveform representing the GSD, so it must be parameterised in a few numerical values. This problem has been the topic of interest of many researchers in the voice quality context, and there are several possible choices, [14]: on the one hand, there are direct estimation methods, where the source parameters are obtained by measuring time domain landmarks. However, they are not very robust because those landmarks are not easy to determine in natural inverse filtered signals. On the other hand, there are fitting estimation methods, where a mathematical model is fitted to the inverse filtered GSD, and that model's parameters parameterised the waveform, however, these approaches have a high computational load. Therefore, we have decided to adopt an alternative parameterisation method proposed in [15], which is amplitude and fundamental frequency normalised, and avoids time domain landmark measurement. Moreover, in [16] it is shown that it is also a global parameter as it depends on the three above mentioned glottal source parameters. We are talking about the Normalised Amplitude Quotient (NAQ) which is defined as the quotient of the maximum of the Glottal Source and the minimum of the Glottal Source Derivative.

Two distinct inverse filtering approaches have been applied in natural and synthetic signal analysis, proposed in [9, 10], for not limiting ourselves to a unique analysis. Later, the central frequencies of the first two formants of the VTR have been measured and the NAQ has been calculated for each period of the GSD. The results of these analysis are shown in Fig. 7 and 8, respectively:



(a) *(b)* *(c)* *(d)*

Fig. 7. Synthetic signal. Vowel "i". *(a)* Acous. Signal, *(b)* Fundamental frequency variation, *(c)* First formant central frequency, *(d)* NAQ



(a) *(b)* *(c)* *(d)*

Fig. 8. Baritone recording. Vowel "i". *(a)* Acous. Signal, *(b)* Fundamental frequency variation, *(c)* First formant central frequency, *(d)* NAQ

Comparing Fig. 8 and Fig. 7 it can be said that in both cases the GSD and VTR parameters (*NAQ* and F_1) show slight variations around their average values. Taking into account that in the proposed model these parameters are kept constant, this variation might be a result of the fundamental frequency dependence of Inverse Filtering techniques, otherwise in the synthetic signal they should be constant during vibrato. However, it has to be mentioned that in natural singing voice, the relative variation of these parameters is a little bit higher than in synthetic signal, what can be explained considering that the Source-Filter is a simplified model, and even, oscillatory movements of many surrounding physical structures (pharyngeal walls, the velum, the tongue, the epiglottis, and the jaw), which have been reported in [17], could also be the cause of the variations. In spite of this, we conclude that the proposed model for vibrato production is very close to natural vibrato production.

5. CONCLUSIONS

In this paper, a novel Non-Interactive Source-Filter model for singing voice has been proposed, that includes vibrato as a possible feature. Thanks to this model, the most relevant acoustic features of vibrato have been related to voice production mechanisms, at least from the signal processing point of view. This model assumes that the GSD and VTR do not suffer major changes during vibrato, and only the fundamental frequency of the voice varies. Natural singing voice analysis has shown that this assumption is supported in vocal vibrato, thus the amplitude variation of each harmonic can be related to a local section of the vocal tract response. Finally, the proposed model has been evaluated through the Inverse Filtering techniques, and it has been shown that the GSD and VTR parameters do not change appreciably during vibrato, as it was assumed in the model.

6. ACKNOWLEDGEMENT

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

REFERENCES

[1] J. Sundberg, ''Acoustic and psychoacoustic aspects of vocal vibrato,'' *in Vibrato (Singular, London)*, 1995

[2] E. Prame, "Vibrato extent and intonation in professional Western lyric singing", *J. of the Acoust. Soc. of Amer.*, Vol. 96, nº 4, pp. 1979-1984, October 1994

[3] I. Arroabarren, M. Zivanovic, J. Bretos, A. Ezcurra, A. Carlosena, "Measurement of Vibrato in Lyric Singers", *IEEE Trans. on Instrumentation and Measurement*, Vol. 51, nº 4, pp. 660-665, August 2002

[4] R. Maher, J. Beauchamp, "An investigation of vocal vibrato for synthesis", *Applied Acoust.*, nº 30, pp. 219 - 245, 1990

[5] M. Mellody, G. H. Wakefield, "Modal Distribution Analysis and Synthesis of a Soprano's Sung Vowels", *Voice Foundation Symposium*, Philadelphia, June 2000

[6] I. Arroabarren, M. Zivanovic, A. Carlosena, "Analysis and Synthesis of vibrato in Lyric singers", *in Proc. of the EUSIPCO*, September 3-6, 2002, Toulose, France

[7] G. Fant, "Acoustic theory of speech production", *Mounton, The Hague*, 1960

[8] N. Henrich, B. Doval, C. d'Alessandro, "Glottal open quotient estimation using linear prediction", *in Proc. of the Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*, Firenze, September 1999

[9] H. Fujisaki, M. Ljungqvist, "Proposal and evaluation of models for the glottal source waveform", *in Proc. of the IEEE ICASSP*, 1986

[10] I. Arroabarren, A. Carlosena, "Glottal Spectrum Based Inverse Filtering", *in Proc. of the EUROSPEECH*, September 1-4, 2003, Geneva, Switzerland

[11] X. Serra, J. Smith III, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System based on Deterministic plus Stochastic Decomposition", *Computer Music Journal*, Vol. 14, nº 4, pp. 12- 24, Winter 1990

[12] I. Arroabarren, M. Zivanovic, X. Rodet, A. Carlosena, "Instantaneous Frequency and Amplitude of Vibrato in Singing Voice", *in Proc. of the IEEE ICASSP*, April 6-10, 2003, Hong Kong, China

[13] G. Fant, J. Liljencrants, Q. Lin, "A four-parameter model of glottal flow", *STL-QPSR*, Vol. 85, nº 2, pp. 1-13, 1985

[14] H. Strik, "Automatic parametrization of differentiated glottal flow: Comparing methods by means of synthetic flow pulses", *J. of the Acoust. Soc. of Amer.*, Vol. 103, nº 5, pp. 2659-2669, May 1998

[15] P. Alku, T. Bäckström, "Normalized amplitude quotient for parametrization of the glottal flow", *J. of the Acoust. Soc. of Amer.*, Vol. 112, nº 2, pp. 701-710, August 2002

[16] I. Arroabarren, A. Carlosena, "Glottal source parameterization: A comparative study", *in Proc. of the VOQUAL*, August 27-29, 2003, Geneva, Switzerland

[17] M. Hirano, S. Hibi, S. Hagino, ''Physiological aspects of vibrato,'' *in Vibrato (Singular, London)*, 1995