Computer-aided musical analysis of extended vocal techniques for compositional applications

Leonardo Fuks
leofuks@serv.com.ufrj.br
UFRJ-Escola de Música, Rua do Passeio 98, 200021-290,Rio de Janeiro, BRASIL
KTH-Speech, Music, Hearing, Royal Institute of Technology, SWEDEN

Abstract

This paper describes an ongoing investigation on various extended vocal effects, assisted by personal computer, for the application in contemporary music. Phenomena of co-oscillation between the vocal folds and other structures may cause amplitude modulation on a quasi-periodic glottal signal, thus establishing subharmonics and their integer partials. By means of computer-aided analysis, relevant aspects of several types of phonation may be revealed, complemented by physiological and aerodynamical measurement methods. Such phenomena may include the participation of the false vocal folds (vocal-ventricular mode, VVM, found in Tibetan and Mongolian chants), epiglottis (vocal growl, used in Jazz and Pop music), and segregated mass-spring subsystems within the vocal folds (hypothesised in the case of vocal fry or creaky voice). In addition, cases of biphonation (diplophonia), where the pair of vocal folds oscillates at two or more different frequencies are considered. Computer-aided analysis is shown to be an efficient and inexpensive platform for such study, providing interesting data for voice modelling, music performance and compositional applications.

INTRODUCTION

Human voice is produced by a flow of pulmonary air through the resistive larynx, which sustains oscillations of its elastic components. The glottal source, i.e., the sound obtained by the modulation of the airflow, is then filtered by the vocal tract and modified by several structures that serve as articulators (Fant, 1960). The glottal oscillator, however, is to some extent aperiodic and chaotic (see Davis & Fletcher, 1996).

Since our trained ears are quite sensitive to how the glottal oscillator behaves, those characteristics impose great challenges in the modelling and synthesis of naturalistic vocal sounds, which cannot be effectively accomplished by simplified methods (Chowning, 1980). Several methods of voice synthesis are available, and the most promising ones seem to be based on physical modelling of the whole phonatory apparatus, which still require more knowledge on the glottal, filtering and articulatory mechanisms.

Chaos theory has been found a suitable tool in describing some phenomena associated with complex vocal fold vibration (e.g., Berry et al., 1994; Davis & Fletcher, 1996). Bifurcations and chaos have been identified in the cries of newborns (Herzel & Reuter, 1996) and in asymmetric vocal fold oscillations (Steinecke & Herzel, 1995). Besides what could be called "normal phonation", human voice is highly flexible, varying in form according to culture and context (see Zemp, 1996). This indicates that different vocal practices may reflect different production mechanisms.

MATERIALS AND METHODS

One musically trained subject, the author, produced all vocal examples in this study. For each phonation type the subject practiced for a period long enough to ensure maximal control and repeatability in the tasks. The recordings were done directly on a personal computer provided with a Turtle Beach Pinnacle Plus soundboard. An Ono-Sokki LA-215 sound level meter was used as an input device, with a MI-421 microphone.

All recordings were performed in an acoustically treated room, and the microphone positioned at a distance of 0.3 m from the subject. SoundSwell (Nyvalla DSP AB, Stockholm) and Cool Edit Pro

(Syntrillium Co., Arizona) software packages were used for recording, editing and analysis of the sound signals. Additional measurements in this project included the techniques of high-speed filming (Speedcam+, Weinberger AG, Germany), inverse-filtering (Rothenberg mask, USA), and electro-glottography (Glottal Enterprises type SC-1B). These latter data are not documented in the present paper (see Fuks et al., 1998).

EXTENDED VOCAL TECHNIQUES AND VOCAL REGISTERS

Some traditional and contemporary music make use of special vocal techniques adding new sonorities to the available resources known and employed in western classical music. These may be labelled extended vocal techniques (Barnett, 1977). Since it is a definiton by exclusion, a more precise classification system is required in order to describe such techniques.

On the other hand, conventional classification systems for voice production include a redundant and abusive amount of terms, when it refers to registers (Hollien, 1974). The most widely accepted subdivision of registers considers three categories: pulse register (or vocal fry, or creaky voice), modal register (or chest voice, or heavy register), and loft register (or head voice, falsetto, light register). These three registers differ perceptually from each other, or when examined through a spectral analysis.

We collected effects that derive from different forms of phonation therefore they could be described as belonging to different vocal registers. For instance, a vocal tradition in different regions in Central Asia such as Tibet, Tuva, Mongolia and Ladakh have been reported to include very low pitched drones, often complemented by shifting, salient high pitched partials (Smith et al., 1967; Zemp, 1996; Bretèque, 1988; Bloothhooft, 1992). These effects are covered by the examples employed in this study.

Vocal-ventricular Mode

In a previous study, a particular phonation mode, labelled vocal-ventricular mode (VVM), produced by a healthy, musically-trained subject was judged as perceptually identical to that used in the Tibetan chant tradition (Fuks et al., 1998; Fuks et al., 1999). VVM covered a range close to an octave, starting at about 50 Hz. High-speed glottography revealed that the ventricular folds oscillated at half the frequency of the vocal folds thus yielding a frequency of f0/2. Phonation at f0/3 was also possible. Presumably, aerodynamic forces produced by the glottal flow pulses sustained the vibrations of the ventricular folds.

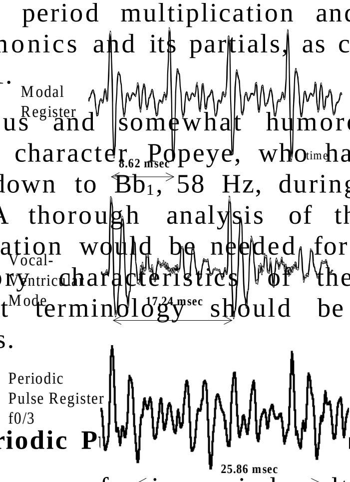


Figure 1. Amplitude modulation in voice. Audio signals of three types of vocal registers are shown: modal (top), vocal-ventricular (VVM, middle) and periodic pulse (vocal fry, bottom). For VVM, in this case, the wave is modulated causing a period doubling. For pulse register, the wave may have its period multiplied by 2,3,... n, in this case by 3

By use of a personal computer, the audio signals of VVM were analysed and revealed a case of amplitude modulation (AM), creating period multiplication and establishing denser spectra of subharmonics and its partials, as compared to modal phonation, see Figure 1.

A curious and somewhat humorous variant is the voice of the cartoon character Popeye, who had a very low-pitched and harsh voice (down to Bb_1, 58 Hz, during singing), possibly produced by VVM. A thorough analysis of the kind applied in the present investigation would be needed for an exhaustive description of the phonatory characteristics of these tone production modes. A coherent terminology should be based on the results of such analyses.

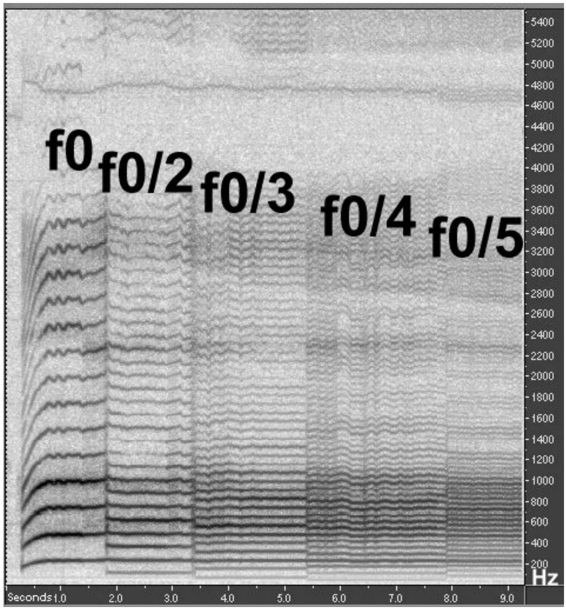
1.1 Periodic P… l Fry

Another case of voice period multiplication, periodic pulse register (PPR) or vocal fry, produced motion patterns differing from those of VVM, as observed by electroglottography. Ocurrence of f0/2 or f0/3 modes has been already documented (Barnett, 1977; Titze, 1994). This register has been studied and described by several authors (e.g. Hollien, 1974; Titze, 1994; Blomgren & Chen, 1998). In our study, PPR presented a wider range than VVM, tipically more than an octave and a fifth, and the possibility of subharmonics varying from modes 1 up to 5, i.e., generating a subharmonic series with five clearly distinctive sounds departing from a stable fundamental frequency. Also, the sound intensity achieved by PPR was considerably lower than in VVM, usually 12dB less in extreme cases. This is probably due to the fact that VVM admits considerably higher subglottal pressures than PPR.

However, accurate measurements of the degree of vocal fold adduction, stiffness of the vocal fold margins and their precise movements are still required for a complete understanding of this mechanism.

Figure 2 shows a case of vocalization in which the sound evolves from a starting pitch of G_3 (f_0, ca. 250 Hz) to G_2 ($f_0/2$, 125 Hz), to C_2 ($f_0/3$, ca. 83 Hz), then to G_1 ($f_0/4$, 63 Hz) and finally to Eb_1 (Fo/5, 50 Hz). It can be noted that the frequency 250 Hz, here called the *prime frequency*, keeps stable throughout the whole sound production. This is clearly a subharmonic series, in which integer numbers sequentially multiply the period. This phenomenon has a parallel in chaos theory, where period doubling is a central occurrence. However, it is not quite clear why the multiples 3 and 5 occur. Our current hypothesis for this behavior is that some portions of the vocal folds segregate from the main oscillating parts, defining independent mass-spring subsystems, which are able to co-oscillate at subharmonic modes, see item co-oscillation below. This requires future studies, with combined use of computer-based methods and high-speed imaging tools.

Figure 2. A spectrogram of a periodic pulse register (vocal fry) vocalization in which the sound evolves from a starting pitch of G3 (f0, ca. 250 Hz) to G2 (f0/2, 125 Hz), to C2 (f0/3, ca. 83 Hz), then to G1 (f0/4, 63 Hz) and finally to Eb1 (f0/5, 50 Hz).Note that the first partial of f_0 keeps constant throughout the whole process.



1.2 Growl

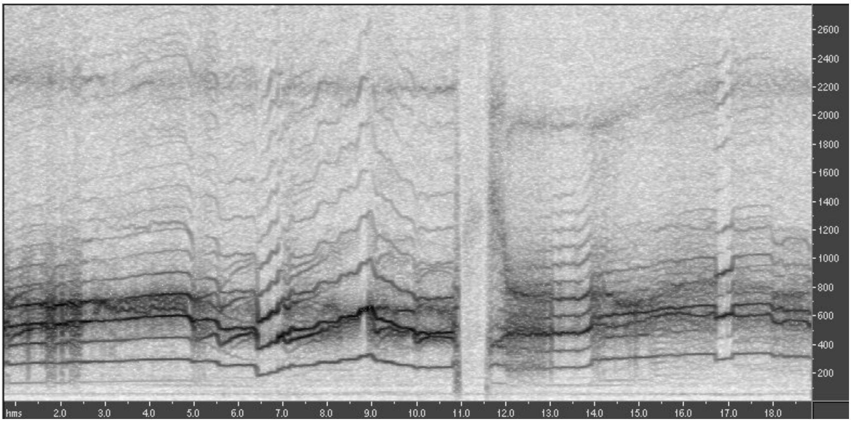
Vocal "growl" is sometimes used by jazz and pop musical singers, e.g., Louis Armstrong. Phonation is usually breathy, complemented by a constriction of pharyngeal/laryngeal structures, as revealed by video-stroboscopy (Thalén M, personal communication). This supraglottal valve mechanism produces a modulation of the airflow, and hence of the sound quality, adding irregular or regular low rate oscillations to the sound. In some cases, the "growl" vibrations may present a phase lock to the vocal folds in the f0/2 or f0/3 modes. Spectral and perceptual analysis of some of Armstrong's recordings did not show similarity with our VVM examples, although some subharmonics were observed.

1.3 Biphonia/Diplophonia

The pair of vocal folds may be set into oscillation at different frequencies one from the other, under certain conditions. This is called biphonia or diplophonia, and may occur in pathological cases (Tigges et al. 1997) or by the use of extended techniques. The performer establishes assymetrical degrees of contraction in the muscles that control vocal tension, helped by some abduction of the glottis, thus enabling non-coupled vibrations in the larynx.

Therefore, a multiphonic glottal source is filtered by the same vocal tract, with some breathy noise due to the partial adduction of the glottis. This can be observed at the spectrum, see Figure 3.

Biphonic-like sounds can be convincingly synthesised by the adding of two vocal sounds and some white noise, passing the resulting signal through formant filters simulating the vocal tract. In this case, Figure 3, there seem to be two main formants at 650 Hz and



2200 Hz.

Figure 3. Spectrogram of biphonic sounds. Note that there are two families of harmonically related sounds, presumably produced by assymetrical and independent oscillations of the vocal folds.

1.4 Overtone singing

All cases of phonation above, VVM, PPR, growl and even biphonia may be skillfuly controlled by a singer, by means of changes in the filtering properties of the vocal tract. If the singer manages to establish resonating cavities with a high Q-value (resonance peak), it may originate the effect of segregation of vocal harmonics and formants, and being perceived as overtone singing, a well-described technique employed by several eastern cultures. The same or similar effects have been referred to by other terms, including diphonic singing (Smith et al., 1967), "diplophony" (Dmitriev, 1983), "throat-singing", harmonic singing, etc. Figure 4 shows a spectrogram of overtone singing of the melody "Oh Susannah" performed on a fixed VVM drone at approximately 53 Hz.

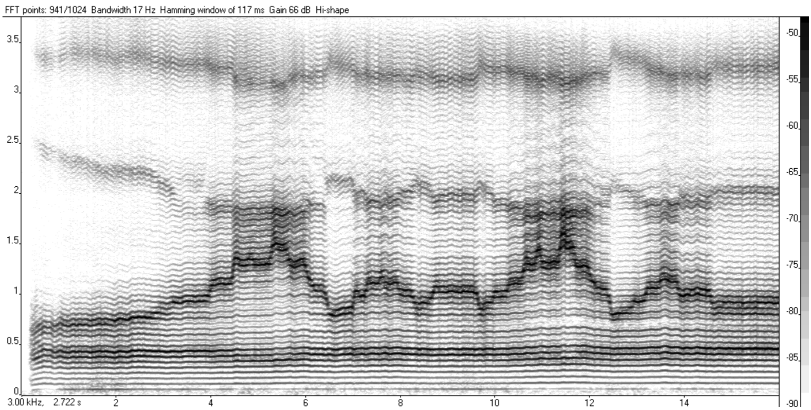


Figure 4. Overtone singing ("Oh, Susanah" melody) spectrogram from a **VVM** utterance, at a fixed fundamental frequency, approx. 53 Hz (From Fuks, 1998).

2 DISCUSSION

2.1 Subharmonics, Co-Oscillation and Period Multiplication

The existence of subharmonic partials, a rather enigmatic concept in music perception and theory, can be shown and explained in a straightforward way in the examples here presented.

Together with the normal vibrations of the vocal folds, other structures may initiate and maintain coupled oscillations with them. In order to keep steady self-sustained vibrations, it is highly preferable that the co-oscillating system has harmonically-related natural frequencies. For instance, if the vocal folds vibrate at f_0, they will be able to excite coupled systems at natural frequencies of

f_0/n, $f_0,/n-1$,…, f_0, $f_0.2$, $f_0.3$, $f_0.(n-1)$, $f_0.(n)$

The coupling between the vocal folds and the additional systems may be mechanical (in PPR) or aerodynamic (in VVM and growl), due to the fact that in voice production there are important driving forces due to tissue vibration and airflow modulation. Figure 1 showed the audio signals of three different types of phonation modes: modal, vocal-ventricular and periodic pulse register (or vocal fry).

The presence of harmonically-related co-oscillation in the airways produces a modulation inside the modulation. Since the amplitude of the pressure pulse of the airflow is affected, this is a case of amplitude modulation. In amplitude modulation (AM), there is a carrier frequency and a modulating frequency. In our case, the carriers are all the partials of the vocal sounds and the modulating frequency is the frequency of the co-oscillating structure, always in a harmonic relationship with f0 . This creates a rich spectrum with every partial **plus and minus** the modulating frequency. Our spectral representation corroborates the model of AM and also matches well with the observations by other methods (electroglottography, inverse filtering and high-speed imaging, in Fuks, 1998).

Not only AM has been observed, but also minute variations in the duration of the cycles, called *jitter*, that if regularly present is a form of frequency modulation (FM). We plan to study these phenomena in the future.

3 Conclusions

Computer-aided tools for sound analysis were shown to be efficient and inexpensive for our purposes, providing insights on the nature of the vocal vibrations during extended techniques.

For synthesis applications, this study may provide directions for the building of computer algorithms.

The occurrence of period multiplication, amplitude modulation and subharmonics has been demonstrated and clarified through the computer-aided methods.

A physical explanation could be offered for the period multiplication phenomena, rather than by arguments derived from chaos theory.

Music performance may benefit from computer-aided methods for training and identification of singing techniques.

4 ACKNOWLEDGEMENTS

This study was done under the supervision of Johan Sundberg, from the Royal institute of Technology, KTH, Sweden. The kind and expert assistance of Stellan Hertegård, Per-Åke Lindestad and Britta Hammarberg from Karolinska Hospital and Karolinska Institute, is gratefully acknowledged. Lobsang Tenzin Kyaga, from Tibet/Sweden, kindly judged live demonstrations of various types of vocal sounds. The author profited from valuable discussions with Svante Granqvist, Johan Sundberg, Gunnar Fant, Johan Liljencrants, and Anders Askenfelt. This research was supported by a scholarship from the Brazilian Ministry of Education, through Capes Foundation, and by the Rio de Janeiro Federal University, UFRJ, School of Music. The Division of Graduate Studies, UFRJ School of Music, supports the participation in this symposium.

5 REFERENCES

Barnett BM (1977). Aspects of vocal multiphonics, *Interface* 6: 117-149.

Berry DA, Herzel H, Titze I & Krischer K (1994). Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions, *J Acoust Soc Am* 95: 3595-3604.

Blomgren M & Chen Y (1998). Acoustic, aerodynamic, physiologic and perceptual properties of modal and vocal fry registers, *J Acoust Soc Am* 103/5: 2649-2658.

Bloothooft G, Bringmann E, van Cappellen M, van Luipen JB & Thomassen KP (1992). Acoustics and perception of overtone singing, *J Acoust Soc Am* 92/4: 1827-1836.

Bretèque BA de la (1988). Le chant diphonique, *Revue de Laryngologie* 109/4: 387-388.

Chowning J M (1980). Computer synthesis of the singing voice, in *Sound Generation in Winds, Strings, Computers*, ed. Johan Sundberg, Royal Swedish academy of Music, no. 29, Stockholm, 1980.

Davis P & Fletcher N (editors) (1996). *Vocal fold physiology-controlling complexity and chaos*, a collection of invited papers presented at the 9th Vocal Fold Physiology Symposium, Australia-1995, San Diego: Singular Publishing Group.

Dmitriev LB, Chernov BP & Maslov VT (1983). Functioning of the voice mechanism in double-voice Touvinian singing, *Folia Phoniat* 35: 193-197.

Fant G (1960). *Theory of Speech Production*, The Hague: Mouton.

Fuks L, Hammarberg B & Sundberg J (1998). A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences, *KTH TMH-QPSR* 3/1998, 49-59, Stockholm.

Fuks L, Lindestad P A, Hertegard S (1999). A vocal-ventricular mode in a Tibetan-like phonation: acoustical, aerodynamic, electroglottographic and high-speed visual evidences, *Proceedings of the 2nd World Voice Congress*, p 62, São Paulo,

Fuks L (1998). *FROM AIR TO MUSIC - Acoustical, physiological and perceptual aspects of reed wind instrument playing and vocal-ventricular fold phonation*, PhD Thesis in music acoustics, Royal Institute of Technology-KTH, Department of Speech, Music and Hearing, Stockholm.

Herzel H & Reuter R (1996). *Biphonation in voice signals, non-linear chaotic and advanced signals processing methods for engineers and scientists*. Woodbury: AIP Press.

Hollien H (1974). On vocal registers, *Journal of Phonetics* 2: 125-143.

Smith H, Stevens KN & Tomlinson RS (1967). On an unusual mode of chanting by certain Tibetan Lamas, *J Acoust Soc Am* 41/5: 1262-1264.

Steinecke I & Herzel H (1995). Bifurcations in an asymmetrical vocal-fold model, J Acoust Soc Am 97/3: 1874-1884.

Tigges M, Mergell P, Herzel H, Wittenberg T & Eysholdt U (1997). Observation and modelling of glottal biphonation, *Acustica/Acta acustica*, 83: 707-714.

Titze IR (1994). Principles of voice production, N.J: Prentice Hall.

Zemp H (editor) (1996). *Voices of the world - an anthology of vocal expression, Paris*: UMR 9957-Musée de l´Homme.