Vocal distortion and real-time processing of roughness

Marta Gentilucci, Luc Ardaillon, Marco Liuni

Vocal distortion and real-time processing of roughness

Conference Paper · August 2018

CITATION
1

READS
190

3 authors:

Marta Gentilucci
Harvard University
2 PUBLICATIONS **2** CITATIONS
SEE PROFILE

Luc Ardaillon
Institut de Recherche et Coordination Acoustique/Musique
9 PUBLICATIONS **21** CITATIONS
SEE PROFILE

Marco Liuni
Institut de Recherche et Coordination Acoustique/Musique
32 PUBLICATIONS **91** CITATIONS
SEE PROFILE

VOCAL DISTORTION AND REAL-TIME PROCESSING OF ROUGHNESS

Marta Gentilucci
Research Creation Interfaces department
IRCAM, Paris, France
marta@martagentilucci.com

Luc Ardaillon
Analysis/Synthesis team
Science and Technology of
Music and Sound (UMR9912,
IRCAM/CNRS/Sorbonne
Université), Paris, France
luc.ardaillon@ircam.fr

Marco Liuni
ERC CREAM project,
PDS team
Science and Technology of
Music and Sound (UMR9912,
IRCAM/CNRS/Sorbonne
Université), Paris, France
marco.liuni@ircam.fr

ABSTRACT

This paper shows the results of our research on parametric control of the singing voices distortion. Specifically, we present a real-time software device that allows the manipulation and control of vocal roughness. The compositional interest to work with classically trained opera singers and with vocal distortion led us to start a research in the signal processing domain. The need has been to develop a tool that could facilitate the production of distorted sounds without the direct effort of the singer. In that way, the singer can perform a non distorted or lightly distorted sound, and the software tool will recreate or magnify in real-time the distorted part.

1. INTRODUCTION

The vocal extended techniques and distortion have been widely used in the musical and theatrical contexts for many decades. There are many works that thematize or use different types of distortions, or extended techniques connected to the production of noise, and the list of composers that worked on the voice's transformation in an unconventional creative ways is very long. Karlheinz Stockhausen (*Spirale*), Peter Maxwell Davies (*Eight Songs for a Mad King*), Helmut Lachenmann (*TemA*), and in recent years Georgia Spiropoulos (*Les Bacchantes*) are just few examples. There are other contexts in which the distortion is part of vocal sound palette, for example it is known as 'growl' in some pop music milieu. In this paper, the distortion has been part of the authors' wider research on vocal extended techniques connected to the vibrato. The specific study of the distortion started with the encounter and the close work with the mezzo-soprano Marie-Paule Bonnemason, who had both a classical opera vocal training, and a deep connection with Roy Hart [1] vocal techniques and with the Panthéâtre [2] as well. Since then, if the main goal has been to enlarge the classical vocal possibilities, the distortion is not intended as a color or gesture, or as a theatrical effect, but as a sound quality that can be composed as much as the harmony, rhythm, pitch, etc. By working with other classical singers, it has become evident that the fragility and the difficulties of stable completely controllable distorted sounds, make their use a challenge for any singer who did or did not have a specific training. Furthermore, there are physical limitations of the vocal folds, for example the amount of time in which a classical singer can hold or reproduce repeatedly distorted sounds, or the rapid alternation between distorted and not distorted sound in different vocal ranges. The first explicit use of the distortion in Marta Gentilucci's work, appears in *Auf die Lider* (2016), for soprano, percussion and electronic.

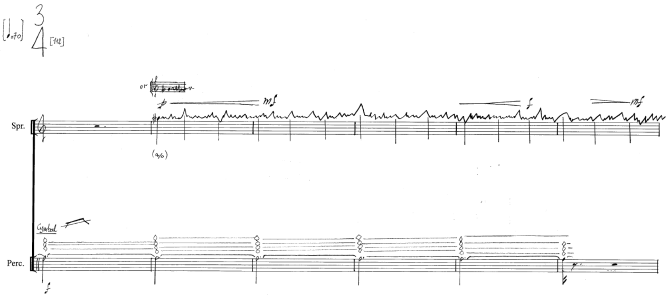

Figure 1: Score excerpt from *Auf die Lider* (2016), for soprano, percussion and electronic.

Here the natural distorted voice is introduced and accompanied by the percussionist, who plays a cymbal with a bow, producing a distorted sound rich in high harmonics. If these sounds are not available and feasible for every classical singer, the necessity has been to find alternative ways to produce the distortion digitally, and to help the singer during the performance. Then, two main questions arouse: how can a composer write a score for a classical singer, that is not limited to the traditional vocal sounds, but that integrates the use of the distortion in a feasible way? How can distorted sounds, by nature fragile and unstable, be manipulated and steadily controlled? The results of this interdisciplinary collaboration across composition, signal processing and computer music design, allows to answer these questions with a research that open a new perspective on sound distortion through live signal processing of roughness.

The paper has the following structure: section 2 is a state-of-the-art about the roughness and its synthesis/transformation,

[1] http://roy-hart-theatre.com/legacy/
[2] https://www.pantheatre.com/2-voix-fr.html

while section 3 details the methods and implementation of the real-time software tool, with some examples of the obtained results (section 3.3), followed by some conclusions in section 4.

2. SYNTHESIS AND TRANSFORMATION OF VOCAL ROUGHNESS

In the spectral domain, a rough voice is characterized by a low Harmonic-to-Noise Ratio (HNR) [1, 2], with the presence of noise and subharmonics (sinusoids present between the harmonics of the voice) [3, 4, 5, 6]. In some more extreme examples of vocal roughness, the sound becomes completely noisy and it is not even possible to identify a pitch in the signal [5].
In the time domain, rough voices are mainly characterized by the presence of important degrees of jitter and shimmer [6, 7, 8]. Jitter can be defined as a period-to-period irregularity of pitch (each pitch cycle may have a different duration), while shimmer is defined as period-to-period amplitude variations (each glottal pulse may have a different amplitude). In some cases, one can observe some macro-pulses, where a macro-pulse is a group of pulses (with varying shapes and amplitudes) that exhibit a certain periodicity at a lower rate than the real f_0 [5]. Rough voices may also be more or less stable, and in some cases (e.g. in screamed voices) present bifurcations, defined as sudden and uncontrolled transitions between different vibratory behaviors (e.g. different number of sub-harmonics) and possibly to a chaotic regime [9, 4, 10].

From the perceptual point of view, the perception of roughness is related to the ability of the auditory system to perceive and resolve close individual sinusoids presented together, as explained in [11]*"The condition for our ability to hear one of 2 equally strong spectrum partials as an individual tone is that they are separated by at least one critical band. [...] All pairs of partials that are similar in amplitude and separated by less than a critical band contribute to the roughness of the timbre. If the pair of partials is high in amplitude, the contribution is substantial."* This explains for instance why the presence of sub-harmonics in the voice spectrum is perceived as roughness. Similar conclusions were drawn in [12] from the study of amplitude and frequency-modulated tones, which adds that *"the entire roughness is composed of the partial roughnesses which are contributed by adjacent critical bands"*.

As vocal roughness covers multiple different voice qualities, several approaches may be used to model each voice quality. In [5], the author classified roughness-related vocal effects found in different singing styles, denoted as "Extreme Vocal Effects", into 5 categories (rattle, distortion, growl, grunt, and scream, from the softest to the more extreme), giving a description and musical references for each. Based on the Wide-band Harmonic Sinusoidal Modeling algorithm [13], the author tried to reproduce those effects, from the analysis of recordings, by a combination of various treatments: global stretching of the spectral envelope; spectral filtering to introduce macro-pulses (using a different filter for each glottal pulse); addition of noise based on phase randomization; addition of pitch variations (jitter); and a negative gain on the fundamental. Each one on those parameters has a different setting, depending on the effect to be produced.
In [14], the author proposed 2 approaches to create such effects. The first approach consists in transposing down the original signal by a certain number of octaves, then shifting and scaling several copies of this transposed signal with various delays and gains values with a certain amount of randomness (to introduce jitter and shimmer), and finally summing them together.
The second approach presented in [14] consists in adding sub-harmonics to the signal directly in the frequency domain, based on the phase vocoder. In this approach, sub-harmonics are added only in the range [f_0-8kHz]. The phase and amplitude patterns of these sub-harmonics are imposed based on the analysis of real growl sounds.
Another frequency-domain approach is presented in [6], where authors proposed to use spectral morphing to mix an original "clean" voice to be transformed with a sample of rough voice with the desired voice quality. This is achieved by first inverse filtering the rough sound by its spectral envelope to get a residual signal, apply the target f_0 curve by time-domain resampling, filtering it back with the spectral envelope of the original clean sound, and finally transforming the harmonics in order to match the phases and amplitudes of the original sound. The rough source can also be looped if necessary to match the target duration.
Other approaches are based exclusively on jitter and shimmer modeling, for transforming speaking or singing voices. In [7], jitter is defined as the average intensity in a band around the fundamental in the spectrum of a normalized pitch contour. The jitter can thus be introduced or modified by changing the mean and variance of the energy in this band.
In [15], a generative model based on statistical analysis of natural hoarse voices is used to modify the jitter and shimmer properties of a modal (or "clean") voice. The jitter is first obtained by high-pass filtering the f_0 contour. Then some statistics are extracted on the degree of jitter and the numbers of consecutive pitch cycles without alternations of the jitter derivative. "Jitter banks" are built to store the original pitch variations due to the jitter. Then, based on these statistics and jitter banks, a new pitch curve including jitter can be generated and applied by time-domain resampling of each individual pitch cycle obtained by a pitch-synchronous analysis (using envelope preservation).

3. A REAL-TIME PARAMETRIC TOOL FOR VOCAL DISTORTION

Analysis and synthesis of singing voice techniques linked to rough vocalizations, like growl or distortion, are extremely challenging, as roughness is generated by highly unstable modes in the vocal fold and tract. For a composer, having a mean to parametrically control a distortion effect implies the possibility to deeply explore such techniques with a significant reduction of the singer's effort, as well as a precise tool to represent and characterize such vocals, that can help the definition of new strategies for roughness and distortion notation in a score.

Moreover, having a real-time software allows to apply the

effect during rehearsals or live performances. Being an additive technique (see section 3.1), this effect allows also to create acoustical illusions on a natural singing voice, by simply placing a loudspeaker close to the interpreter.

3.1 F0-driven amplitude modulations

Our approach to model such effects parametrically is based on a simple amplitude modulation and time-domain filtering to efficiently add sub-harmonics in the original signal. Amplitude modulation simply consists in multiplying in the time-domain a carrier signal with a modulating signal that has a lower frequency and an amplitude varying in the range [0-1], centered around 1.
Let $x_c(t) = A_c \cos(\omega_c t)$ be the carrier signal, with an angular frequency ω_c and an amplitude A_c, and $x_m(t) = 1 + h\cos(\omega_m t)$ be the modulating signal with an angular frequency ω_m and a modulation depth $h \in [0-1]$ (also called the modulation index). We then have, as a result of the modulation:

$$\begin{aligned} y(t) &= x_m(t)x_c(t) \\ &= (1 + h\cos(\omega_m t))A_c\cos(\omega_c t) \\ &= A_c\cos(\omega_c t) + A_c h\cos(\omega_m t)\cos(\omega_c t) \\ &= A_c\cos(\omega_c t) + \tfrac{A_c h}{2}\cos((\omega_c+\omega_m)t) \\ &\quad + \tfrac{A_c h}{2}\cos((\omega_c-\omega_m)t) \\ &= x_c(t) + y_+(t) + y_-(t) \end{aligned} \quad (1)$$

The resulting signal thus contains the original sinusoidal carrier signal $x_c(t)$, and 2 new sinusoids:

$$y_+(t) = \frac{A_c h}{2}\cos((\omega_c+\omega_m)t)\ ,$$

$$y_-(t) = \frac{A_c h}{2}\cos((\omega_c-\omega_m)t)\ ,$$

with amplitudes $\frac{A_c h}{2}$ at frequencies $\omega_c+\omega_m$ and $\omega_c-\omega_m$.

Now, let's consider $x_c(t)$ being a voice signal, approximated by a simple sum of N harmonic sinusoids: $x_c(t) = \sum_{i=1}^{N} A_i\cos(i\omega_0 t)$ (where $\omega_0 = 2\pi f_0$). The result of the modulation of this signal by $x_m(t)$ would simply be the sum of each harmonic modulated individually:

$$\begin{aligned} y(t) &= x_m(t)x_c(t) \\ &= x_m(t)\textstyle\sum_{i=1}^{N} A_i\cos(i\omega_0 t) \\ &= \textstyle\sum_{i=1}^{N} x_m(t)A_i\cos(i\omega_0 t) \\ &= x_c(t) + \textstyle\sum_{i=1}^{N}(y_{+_i}(t) + y_{-_i}(t)) \end{aligned} \quad (2)$$

with

$$y_{+_i}(t) = \frac{A_i h}{2}\cos((i\omega_0+\omega_m)t)\ ,$$

$$y_{-_i}(t) = \frac{A_i h}{2}\cos((i\omega_0-\omega_m)t)\ .$$

By choosing an appropriate value for ω_m, it is thus possible to generate sub-harmonics between each harmonics at frequencies $i\omega_0 \pm \omega_m$, the distance of each sub-harmonic to its related harmonic i being thus equal to ω_m. Thus, setting $\omega_m = \frac{\omega_0}{k}$, this would generate a pair of sub-harmonics around each harmonic at $i\omega_0 \pm \frac{\omega_0}{k}$.
A particular case is $\omega_m = \frac{\omega_0}{2}$ where the upper sub-harmonic generated by the i[th] harmonic and the lower sub-harmonic generated by the (i+1)[th] harmonic have the same frequency ($i\omega_0+\omega_m = (i+1)\omega_0-\omega_m$). This results in a single sub-harmonic being generated between each pair of harmonics (as can be often observed on real signals).
It is also possible to use a sum of sinusoids for the modulating signal, in order to generate more sub-harmonics:

$$x_m(\omega_0, t) = 1 + \sum_{k=1}^{K} h_k\cos(\frac{\omega_0}{k}t) \quad (3)$$

For instance, in order to generate 3 equally-spaced sub-harmonics, one may use the sum of 2 sinusoids at $\omega_0/2$ and $\omega_0/4$.

Note that in terms of signal's characteristics, temporal amplitude modulation can be related to some kind of shimmer (in this case with a regular periodic pattern and not random variations, as the modulation frequency is directly related to the f_0). Sub-harmonics may also be obtained using frequency modulation (which would then be rather related to jitter). In [16], the author states that such non-linear combination of 2 signals with amplitude and phase modulations produce lateral waves and relates this phenomena as an evidence of coupling between the 2 vocal folds. However, frequency modulation generates an infinite series of sub-harmonics (lateral waves) with more complex amplitude relations, which are thus more complex to control for our purpose.

However, using only this modulation doesn't result in a natural-sounding voice signal. The reason for this is that the amplitudes of the lowest sub-harmonics (and especially that of the first one, below the fundamental) are too high. Observing real signals, one can see that the amplitudes are usually much lower for the lowest sub-harmonics.
It is thus necessary to high-pass filter the sub-harmonics. As the original signal $x_c(t)$ is fully preserved in the modulated signal, the generated sub-harmonics can easily be isolated by simply subtracting this original signal from the modulated one: $y_{sub}(t) = y(t) - x_c(t)$. Once the sub-harmonics have been isolated, they can be high-pass filtered before being added back to the original signal by a simple summation. We thus obtain the final rough voice signal as:

$$y_{rough}(t) = x_c(t) + \alpha y_{sub}^{HP}(t) \quad (4)$$

where $y_{sub}^{HP}(t)$ denotes the high-pass filtered sub-harmonics, and $\alpha > 0$ is a mixing factor.
The chosen parameters should vary to give various degrees and qualities of roughness, depending on the desired effect. Especially, mixing factor α (or equivalently the modulation index) could be adjusted to obtain a more or less intense effect. According to our observations on various recordings, it appears that real voices usually contain from 1 to 5 equally-spaced sub-harmonics.
From the physiological point of view, the amplitude modulation may possibly be related to the interaction between the vocal folds and other vibrating supra-glottal structures such as the ventricular folds. For instance in [10], the author observed vibrations of the ventricular folds at frequencies $f_0/2$ or $f_0/3$, which may then modulate the original

sound wave generated by the vocal folds. But it remains unclear how the high-pass filtering of the sub-harmonics relates to the voice physiology.
Using appropriate settings, this simple approach has proved to give very natural results on several examples. Some subjective listening tests are planned to evaluate the naturalness and perception of the sounds produced. As rough voices are usually related to a rather high vocal effort, those effects should however better be applied on voices that are already tense or loud to obtain natural results.

The presented approach is suitable for generating sounds with stable sub-harmonics. However, for more unstable types of rough voices, the number of sub-harmonics and the modulation parameters can be changed along time to create bifurcations between different regimes. It is also possible to introduce some degree of noise in the modulating signal, in order to obtain more chaotic signals.

3.2 Real-time implementation

Although other approaches had been proposed in the litterature to generate sub-harmonics to create roughness in voice (e.g. [5, 14, 6]), the main advantage of this new approach, beyond its simplicity and the naturalness of the results obtained, is its efficiency. The only operations required to apply this effect are one multiplication for the amplitude modulation, a subtraction to isolate sub-harmonics, a few multiplications and additions for the filtering (depending on the order of the filter), and an addition and a multiplication for the final mixing step. This thus makes this approach especially suitable for real-time, to be used as an audio effect on a real voice.

A real-time implementation of this algorithm, called *angus* [3] , has thus been implemented as an open source patch based on the closed source software Max [4] . The most computationally-heavy step for applying this effect in real-time is the f_0 estimation, that is necessary for setting appropriate frequencies for the modulating signal. We used for this a real-time implementation [5] of the yin algorithm [17], which allowed us to implement the effect with no audible latency. The main parameters of this software tool are :

- the number of amplitude modulators, each one with independent depth and high-pass filtering of the generated sidebands; the modulators' frequencies vary as sub-multiples of the estimated f_0 (i.e., given N modulators, frequencies are $f_0/2, (...), f_0/(N+1)$).
- A temporal envelope to dynamically control the effect's level.
- The possibility to add noise on the modulators' frequencies, by specifying a parameter *noise_amp* that is multiplied by the estimated f_0 and then by the noise value itself (varying between 0 and 1 at the audio sample rate). This noise component can be then low-pass filtered, with the parameter *noise_smooth* specifying the period of the filter in milliseconds.

All of these parameters can be organized in presets and subsequently recalled and interpolated in real-time. To avoid discontinuities in the generated spectra when applying presets with different numbers of modulators, *angus* automatically generates smooth transitions between close sub-harmonics, controlled by a further time parameter specifying the transition's duration.

3.3 Results and examples

The proposed examples have all been realized with *angus*, and are available at the following url:
`http://www.martagentilucci.com/ICMC_2018/`

To clarify the testing process, we chose a soprano who sings a F#4 (figure 2) with little or no vibrato, and no distortion. The first step has been to apply three presets progressively to the original sound: the first one has three modulators, that result in adding to the fundamental three pairs of sidebands (sub-harmonics); the second has only one modulator, the third again three. Here, all the modulators' frequencies change with no interpolation (figure 3). The obtained sound results artificial and static, with an abrupt transition from a state to another.

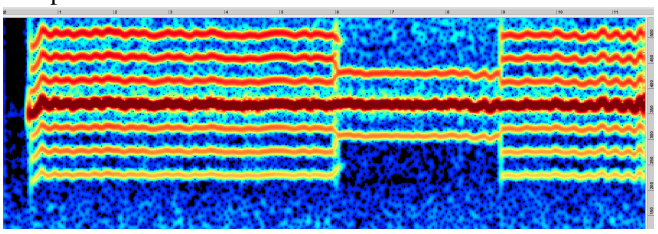

Figure 3: Spectrogram centered around the fundamental frequency of the F#4 note, applying three presets progressively, with no interpolation.

Smooth transitions between the sidebands are obtained with a transition duration of 100 ms (figure 4). This kind of preset, when associated to an adapted time envelope of the overall level of the effect, can generate natural growl-like transformations of the natural voice. However, a sense of steadiness and unnatural unfold of the sound could appear.

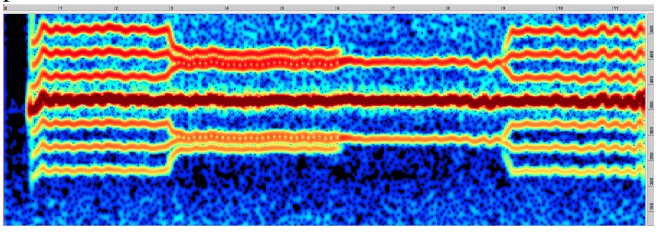

Figure 4: Spectrogram the F#4 note, applying three presets progressively, with a smooth transition in 100 ms.

If we look at the original distorted vocal sound of a soprano who sings a G#5 note (figure 5), we can observe that the sidebands behavior is less ordered and much more chaotic. We also observe that there is more energy around the fundamental frequency, especially at the beginning where it is lightly unstable.

[3] `http://forumnet.ircam.fr/product/angus/`
[4] `https://cycling74.com/`
[5] `http://forumnet.ircam.fr/product/max-sound-box-en/`

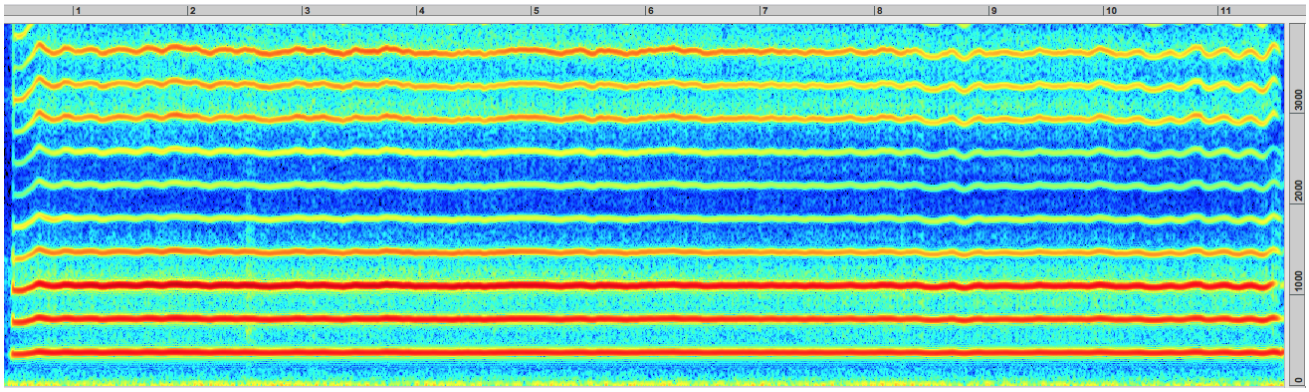

Figure 2: Spectrogram of a F#4 note sung by a soprano, no vibrato and no distortion.

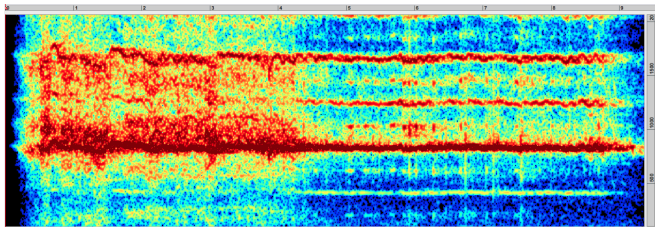

Figure 5: Spectrogram of a distorted vocal sound produced by a soprano who sings a G#5 note.

To reproduce a similar chaotic behavior of the sidebands, we tested several *noise_amp* factors between 1 and 5 for the noise component of the modulator's frequencies, and *noise_smooth* between 0 and 200 ms for low-pass filtering (figure 6).

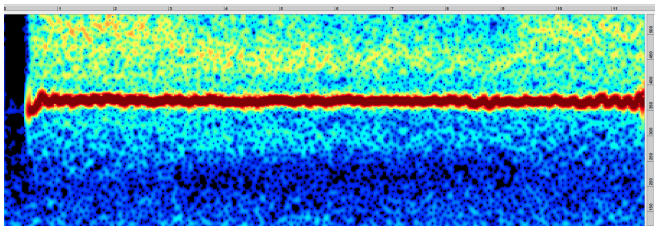

Figure 6: Spectrogram of the F#4 note, with a multiplication factor of 5, and 200 ms low-pass filtering.

The turning point for our research has been when we added a higher number of sub-harmonics (between 5 and 25), and we dynamically changed the modulators number as well as the transition time between 100 and 500 ms (figure 7).

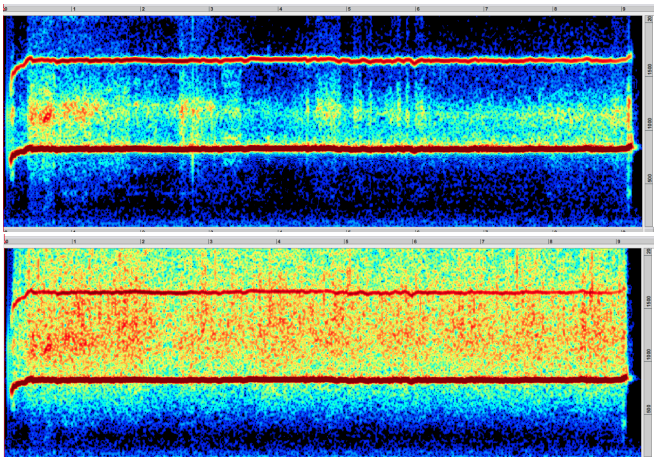

Figure 7: Spectrograms of the G#5 note. Top: natural voice with no distortion. Bottom: transformation with a noise multiplication factor of 5, a number of modulators dynamically changing between 5 and 25, and transition time between 100 and 500 ms .

The natural development of our research will be to adapt the parameters to a specific musical situation and to a specific voice. Only by mixing and calibrating the number of subharmonics, the noise ratio and the transitions between different subharmonics settings to a particular voice, we can obtain a sound that is similar to an actual distorted singing voice .

4. CONCLUSIONS

The goal of our research has been to investigate the distortion of the signing voice. Specifically, we wanted to facilitate the singer and allow him or her to sing without any or little vocal effort to produce distorted sounds. The real-time software tool we built allows the production and control of vocal distortion without having the singer to produce it directly. To improve the sound quality of the tool, a effective strategy consists in integrating compression and reverb to the processing chain. Because our tool does not depend on the spectral characteristics of the input sound but only on its fundamental frequency, its use can easily be extended to non vocal monodic sounds. The current model operates as synthesis tool with manual adjustment of all parameters. As future developments, we aim to extend it to an analysis-synthesis model that allows the automatic

analysis of a target sound (the model) and the synthesis applied to a chosen sound source.

5. REFERENCES

[1] A. C.-g. Tsai, L.-c. Wang, S.-f. Wang, Y.-w. Shau, and T.-y. Hsiao, "Aggressiveness of the Growl-Like Timbre: Acoustic Characteristics, Musical Implications, and Biomechanical Mechanisms," *Music Perception: An Interdisciplinary Journal*, vol. 27, no. 3, pp. 209–222, 2010.

[2] Y. Sasaki and H. Okamura, "Harmonics-to-noise ratio and psychophysical measurement of the degree of hoarseness," *Journal of Speech and Hearing Research*, vol. 27, no. 2-6, 1984.

[3] K.-i. Sakakibara, L. Fuks, H. Imagawa, and N. Tayama, "Growl Voice in Ethnic and Pop Styles," in *Proceedings of the International Symposium on Musical Acoustics*, Nara, Japan, 2004.

[4] R. Neubauer and H. Herzel, "Calls out of chaos: the adaptive significance of nonlinear phenomena in mammalian vocal production," *Animal Behaviour*, vol. 63, no. 3, pp. 407–418, 2002.

[5] O. Nieto, "Voice Transformations for Extreme Vocal Effects," Master thesis, Pompeu Fabra University, Barelona, Spain, 2008.

[6] J. Bonada and M. Blaauw, "Generation of growl-type voice qualities by spectral morphing," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6910–6914.

[7] A. Verma and A. Kumar, "Introducing roughness in individuality transformation through jitter modeling and modification," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, 2005, pp. 5–8.

[8] T. M. Jones, Ã. M. Trabold, Ã. F. Plante, B. M. G. Cheetham, and J. E. Ã. Earis, "Objective assessment of hoarseness by measuring jitter," *Clinical Otolaryngology*, vol. 26, no. 1, pp. 29–32, 2001.

[9] A. Lagier, T. Legou, C. Galant, B. Amy, D. L. Bret, Y. Meynadier, and A. Giovanni, "The shouted voice: A pilot study of laryngeal physiology under extreme aerodynamic pressure," *Logopedics Phoniatrics Vocology*, 2016.

[10] L. Bailly, "Interaction entre cordes vocales et bandes ventriculaires en phonation : exploration in-vivo , modélisation physique , validation," PhD thesis, Université du Maine, Le Mans, France, 2009.

[11] J. Sundberg, *The science of singing voice*, 1990.

[12] E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," *Acta Acustica united with Acustica*, vol. 30, no. 4, pp. 201–213, 1974.

[13] J. Bonada, "Wide-band harmonic sinusoidal modeling," in *Proc of the 11th Int Conference on Digital Audio Effects (DAFx08)*, 2008. [Online]. Available: http://mtg.upf.es/files/publications/WBHSM.pdf

[14] A. Loscos and J. Bonada, "Emulating rough and growl voice in spectral domain," in *Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx'04)*, Naples, Italy, 2004.

[15] D. Ruinskiy and Y. Lavner, "Stochastic models of pitch jitter and amplitude shimmer for voice modification," in *IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, 2008, pp. 489—-493.

[16] A. Giovanni, M. Ouaknine, R. Guelfucci, T. Yu, M. Zanaret, and J. M. Triglia, "Nonlinear behavior of vocal fold vibration: the role of coupling between the vocal folds." *Journal of voice : official journal of the Voice Foundation*, vol. 13, no. 4, pp. 465–476, 1999.

[17] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music." *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002. [Online]. Available: http://audition.ens.fr/adc/pdf/2002{_}JASA{_}YIN.pdf