Acoustic Features and Auditory Impressions of Death Growl and Screaming Voice

Conference Paper · October 2013

DOI: 10.1109/IIH-MSP.2013.120

Akinori Ito
Tohoku University

Acoustic Features and Auditory Impressions of Death Growl and Screaming Voice

Keizo Kato
Graduate School of Engineering
Tohoku University
Sendai, Japan
k-kato@spcom.ecei.tohoku.ac.jp

Akinori Ito
Graduate School of Engineering
Tohoku University
Sendai, Japan
aito@spcom.ecei.tohoku.ac.jp

***Abstract*—In the contemporary music scene, the death growl and screaming voice are often used in the extreme metal, and have been one of the indispensable singing styles. In this study, we made an attempt to clarify the acoustic feature of the death growl and screaming voice. We chose jitter, shimmer and HNR as the acoustic features, and found that the death growl and screaming voice have much larger jitter and shimmer, lower HNR compared with the normal voice. Next, we investigated the relationship between subjective impression and acoustic feature, and found that the screaming voice has an optimum jitter.**

***Keywords*-Singing voice; Death growl; Scream; Pathological voice; Jitter; Shimmer; Harmonic-to-Noise ratio**

I. Introduction

Death growl (death grunt, deadly hawl) and screaming are vocal techniques often used in the metal music such as the extreme metal or metal core. Voices generated by both techniques sounds very bold and animal-like. The death growl is often used by singers such as Chris Barnes, vocalist of Cannibal Corpse; screaming is used by those such as Philip Labonte, vocalist of All That Remains. These vocal techniques have been evolved as a singing style in the heavily distorted guiter and thick drum and bass sound. Originally, these vocal techniques were used in songs with violent and antisocial lyrics, but nowadays they are more commonly used.

Note that the word “growl” includes varieties of vocal techniques, which are wider than the “death growl,” target of this paper. Sakakibara et al. investigated generation of “growl” singing, which denotes singing style used in ethnic singing (such as Japanese *Noh*) as well as jazz singing such as that by Louis Armstrong [1]. Tsai et al. investigated growl-like voice used in Beijing opera, and found that the harmonics-to-noise ratio (HNR) as well as the impression of “aggressiveness” is related to the thickness of abdominal muscles [2]. Their work suggests the general generation mechanism of growl-like voice; however, the voices they analyzed seems to be different from the death growl, our target of analysis.

The death growl and screaming needs training, or the singing style might be harmful for the singer’s vocal fold. To use the death growl or scraming voice in a song when the singer is not familiar with these singing technique, it is desired to synthesize these voices or convert normal singing voice into death growl or screaming voice. There were attempts to implement the conversion from normal singing voice to those extreme voices. [3], [4]. However, it is still unexplored what kinds of physical features are related to auditory impression of “growl-ness” or “screaming-ness.”

In this paper, we investigate physical properties that features the death growl and screaming voice, and then conduct subjective evaluation experiments to analyze the relationship between physical features and subjective impressions.

II. Spectral observations of the death growl and screaming

In this section, we observe the basic properties of death growl and screaming voices, and introduce acoustic features for further analysis.

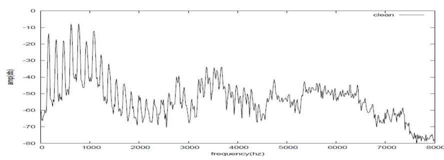
A. Vocal samples

We collected samples of the death growl and screaming voices for observation. We employed two singers, who were amateur vocalists familiar with the extreme metal. One of the two singers uttered five vowels, /a/, /i/, /u/, /e/ and /o/ in an ordinary utterances and the death growl style, and the others did in an ordinary style and screaming style. The utterances were recorded in a soundproof chamber and sampled at 16 kHz, 16 bit/sample.

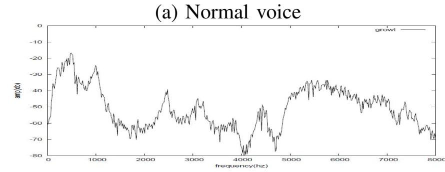
B. Spectral features

Figure 1 and 2 show the spectra of normal and death growl/screaming utterance of vowel /a/. 1024-point FFT with Hamming window was used for the analysis, and we took an average of 10 frames of the stable part. Normal utterances of both singers have obvious harmonic structure. On the contrary, we cannot see any harmonic structure in the death growl and screaming utterances. Spectrum of the death growl utterance does not seem to have any obvious structure other than formant of /a/, while that of screaming utterance has several obvious peaks. The most prominent peak is around 860 Hz, which seems to be the second harmonics of the peak around 440 Hz. There is another peak around 150 Hz, the one-third of the frequency of the second peak. Another

978-0-7695-5120-3/13 $26.00 © 2013 IEEE
DOI 10.1109/IIH-MSP.2013.120

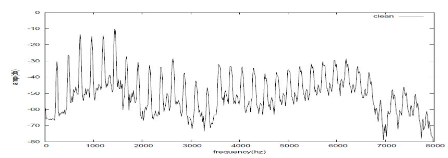


(a) Normal voice

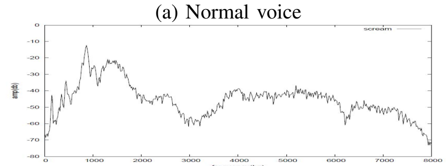


(b) Growl voice

Figure 1. Spectra of normal and death growl voice



(a) Normal voice



(b) Screaming voice

Figure 2. Spectra of normal and screaming voice

peak around 340 Hz is not a harmonic component of other peaks, which seems to be a subharmonics generated from the false cord vibration.

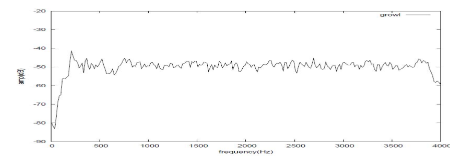
C. Spectrum of source signals

The spectra shown in Fig. 1 and 2 were affected by the vocal tract shape. Then we observed spectra of source signal by removing transfer function of the vocal tract using LPC analysis. We conducted 25th-order LPC analysis on the vocal signals, and then observed spectra of the residue signals. Figure 3 shows the spectra of the death growl and screaming voice. We can see that the residue of the death growl is almost white, and a small peak is observed around 210 Hz. In the spectrum of the screaming voice, the three peaks can be observed more clearly, while the second harmonics of the 440 Hz peak is diminished.

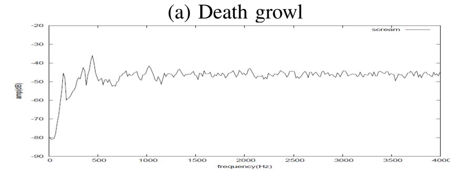
III. Analysis of Acoustic Features

A. Acoustic features

Next, we calculated acoustic features to conduct a quantitative analysis. As a metric of roughness of voice, harmonic-to-noise ratio (HNR) is often used [5]. In addition, jitter and shimmer are used for analysing pathological voices [6]. Therefore, we employ jitter, shimmer and HNR as indice of extremeness of the voice.



(a) Death growl



(b) Screaming voice

Figure 3. Spectra of LPC residue

Jitter expresses fluctution of glottal pulse interval. Consider that we have a source signal that has N glottal pulses. Let the time and amplitude of i-th pulse be t_i and A_i, respectively. The length of the i-th interval is

$$T_i = t_{i+1} - t_i \quad (1 \le i \le N-1). \tag{1}$$

Then the (relative) jitter is expressed as follows.

$$R_{\text{jitter}} = \frac{\frac{1}{N-2}\sum_{i=1}^{N-2} |T_{i+1} - T_i|}{\frac{1}{N-1}\sum_{i=1}^{N-1} T_i} \tag{2}$$

Shimmer is a fluctuation of amplitude of glottal pulses, calculated as follows.

$$R_{\text{shimmer}} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |A_{i+1} - A_i|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \tag{3}$$

The harmonic-to-noise ratio (HNR) is the ratio of energies of harmonic and non-harmonic components of the signal. While there are a couple of methods for calculating HNR [7], we exploited Boesma's algorithm [8]. Let the autocorrelation function of input signal $x(t)$ be

$$r_x(\tau) = \sum_{t=0}^{N} x(t)x(t+\tau). \tag{4}$$

Here, the lag that gives the maximum autocorrelation be

$$\tau_{\max} = \arg\max_{0<\tau<L} r_x(\tau) \tag{5}$$

where L is the length of the analysis window, and the maximum normalized autocorrelation function be

$$r'_x(\tau_{\max}) = \frac{r_x(\tau_{\max})}{r_x(0)}. \tag{6}$$

Then the HNR (in dB) is given as follows.

$$HNR = 10\log_{10}\frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})} \tag{7}$$

When calculating these values, we first removed tranfer function of the vocal tract by taking residue of the LPC analysis. Then we chose 256 period where the signal seems

Table I
ACOUSTICAL FEATURES OF THE DEATH GROWL VOICES

Singer	Jitter(%)		Shimmer(%)		HNR (dB)	
	Normal	Growl	Normal	Growl	Normal	Growl
AY	0.438	27.82	1.51	41.84	9.11	-6.81
OY	0.512	26.16	2.09	48.57	8.52	-5.79
MT	0.281	24.10	3.44	45.48	7.25	-5.90
NA	0.523	23.40	2.56	47.72	8.46	-6.68
YT	0.440	9.23	3.51	16.14	7.46	-1.96
KK	0.274	11.57	1.44	13.14	9.11	-1.99

Table II
ACOUSTICAL FEATURES OF THE SCREAMING VOICES

Singer	Jitter(%)		Shimmer(%)		HNR (dB)	
	Normal	Scream	Normal	Scream	Normal	Scream
TK	0.204	13.919	3.61	46.77	8.11	-5.12
SM	0.532	17.856	4.22	44.872	8.52	-4.12
OY	0.112	19.599	1.6	42.595	8.32	-5.5
MT	0.463	22.182	3.44	27.725	7.25	-6.15
MY	0.781	24.716	2.51	25.781	9.11	-6.3
NA	0.523	22.563	2.56	31.845	8.46	-6.68
YT	0.44	9.939	3.51	15.139	7.46	-3.12
KK	0.274	5.184	1.44	12.628	9.11	1.77

to be stable, and measured the time and amplitude of pulses manually. The jitter and shimmer were calculated using these manually-chosen pulses. When calculating the HNR, we used an analysis window with 8 times wider than the fundamental period of the pulses, and frame shift was the half of the window length. The average of HNR values of all frames was used as the HNR value of the utterance.

B. Voice samples

We recorded samples of death growl and screaming utterances uttered by experienced and inexperienced singers. We employed six singers for death growl utterances, three of them had experience as vocalists in an amateur extreme metal band, and the others had no experiences. Similarly, we employed eight singers for screaming utterances, five of them had experiences and the other three did not. We asked the singers to utter /a/ vowel five times, in the specified singing style. We instructed the singers to utter in their easiest way, and the singing voice to be as stable as possible. We recorded the voices in a soundproof chamber, 44.1 kHz sampling rate and 16 bit quantization.

C. Acoustic features of experienced and inexperienced singers

Table I and II shows the results of death growl and screaming voice, respectively. In these tables, the last three singers are the inexperienced singers. We can see that death growl and screaming voice have much larger jitter and shimmer as well as much smaller HNR. At the same time, the death growl and screaming voice generated by the inexperienced singers are similar to normal voices, having smaller jitter and shimmer and larger HNR. Among the three inexperienced singers, singer NA could generate death growl and screaming voice similar to that of the experienced singers. These results suggest that some people have talent to sing in the extreme voice and the others don't, and it needs training to generate "proper" extreme voices.

D. Subjective evaluation

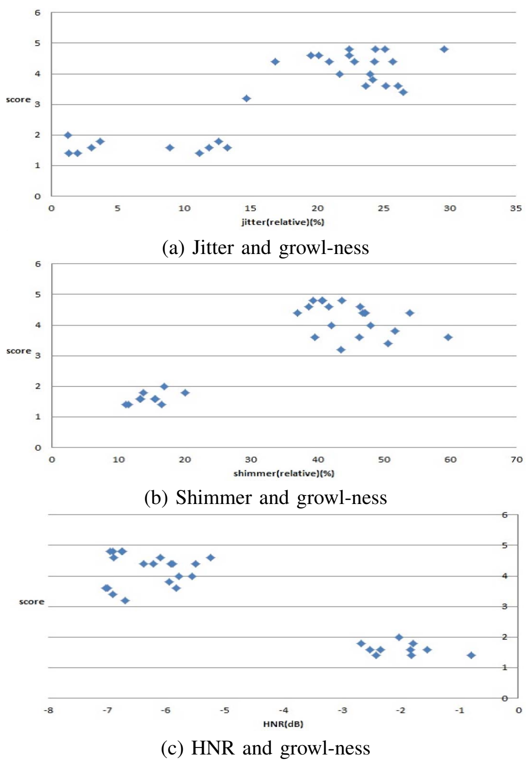
Next, we conducted subjective evaluation to investigate relationship between the acoustic features and listener's impression. We employed five evaluators who have listened to extreme metal songs for more than three years and have experiences of playing extreme metal pieces in amateur bands. One of the five evaluators were who participated in the recording of the samples. They evaluated how the utterance sounds like a death growl or screaming voice using five-point Likert scale. Finally, an average of the scores by the five evaluators was used as the score of the utterance. The presented sound signals were normalized at -3dB.

The experimental results are shown in Fig. 4 and 5. Figure 4 shows the relationship between the three acoustic features and the "growl-ness," and Fig. 5 shows that of the "scream-ness." It is obvious that higher jitter, higher shimmer and lower HNR is related to more "growl-ness" and "scream-ness."

We can also observe the difference between the death growl and screaming voice. As for the jitter, the death growl voices with higher jitter sounded more "growl-like," while the screaming voice had the optimum jitter around 15%, and higher jitter makes the voices less "screaming-voice-like." The similar tendency was observed in HNR, too. Conversely, effect of shimmer was similar for both of death growl and screaming voice, where voice with higher shimmer sounded more like growl or screaming voice. From the evaluators' insight, screaming voices with high jitter were evaluated as "thin voices," which seems to mean voices with less periodicity. Investigation of the relationship between those impressions and the acoustic features is the future issue.

IV. CONCLUSION

In this paper, we made observation of the death growl and screaming voices, kinds of singing voices used in extreme metal songs. From the observation of the spectra, both the death growl and screaming voice were more noise-like compared with the normal voices, and the screaming voice seemed to have a subharmonic peak in the spectrum. Next, we examined three kinds of acoustic parameters: jitter, shimmer and HNR. As a result, we could confirm that the death growl and screaming voice had much higher jitter and shimmer as well as lower HNR values. At the same time, it was found that the voices by an inexperienced singer were similar to the normal voices. Finally, we investigated relationship between the acoustic features and subjective

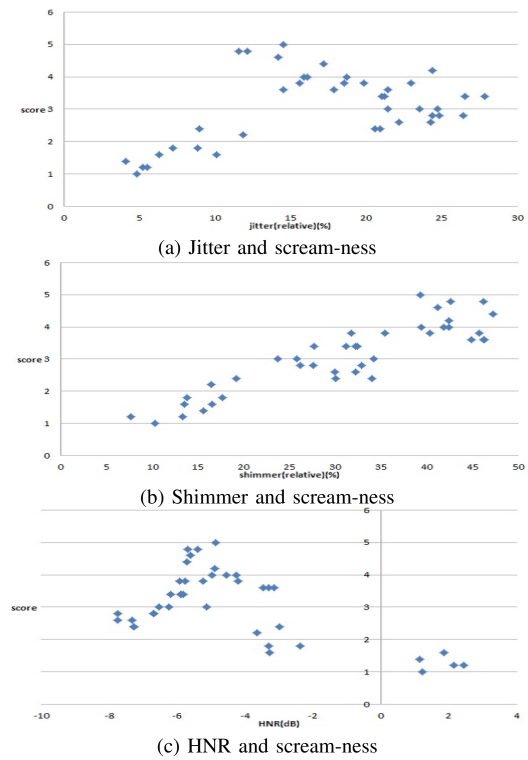


(a) Jitter and growl-ness

(b) Shimmer and growl-ness

(c) HNR and growl-ness

Figure 4. Relationship between the acoustic features and “growl-ness”



(a) Jitter and scream-ness

(b) Shimmer and scream-ness

(c) HNR and scream-ness

Figure 5. Relationship between the acoustic features and “scream-ness”

impressions. Basically, higher jitter, higher shimmer and lower HNR makes the voice sound more extreme-voice-like, but we found that the screaming voice had an optimum jitter.

As a future work, we need to make further investigation on the relationship between the acoustical features and subjective impression such as “deep voice/thin voice,” which seems to be deeply related to the death growl and screaming voice. At the same time, we need to find a method to convert or synthesize the extreme voices for helping creation of songs of various genres.

ACKNOWLEDGMENT

Part of this work was supported by JSPS A3 foresight program “Ultra-realistic acoustic interactive communication on next-generation Internet.”

REFERENCES

[1] K. Sakakibara, K. Fuks, H. Imagawa and N. Tayama, “Growl Voice in Ethnic and Pop Styles,” Proc. Int. Symp. on Musical Acoustics, 2004.

[2] C.-G. Tsai, L.-C. Wang, S.-F. Wang, Y.-W. Shau, T.-Y. Hsiao and W. Auhagen, “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.

[3] A. Loscos and J. Bonada, “Emulating Rough and Growl Voice in Spectral Domain,” Proc. Int. Conf. on Digital Audio Effects (DAFx’04), 2004.

[4] O. Nieto, “Voice Transformations for Extream Voice Effects,” Master’s thesis, Universidad Pompeu Fabra, 2008.

[5] G. de Krom,“Spectral correlates of breathiness and roughness for dipcrent types of vowel fragments,” Proc. Int. Conf. on Spoken Language Proc., pp. 1471–1474, 1994.

[6] J. Kreiman and B. R. Gerratt, “Perception of aperiodicity in pathlogical voice,” J. Acoust. Soc. Am., vol. 117, no. 4, pp. 2201–2211, 2005.

[7] F. Severin, B. Bozkurt and T. Dutiot, “HNR Extraction in Voiced Speech, Oriented Towards Voice Quality Analysis,” in Proc. 13th European Signal Proc. Conf. (EUSIPCO ’05), 2005.

[8] P. Boesma, “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” Inst. of Phonetic Science Proceedings, Univ. of Amsterdam, vol. 17, pp. 97–110, 1993.