Chapter 33
What we have learned about extreme metal production from HSDP
Krzysztof Izdebski, Enrico Di Lorenzo, Yuling Yan & Matthew Blanco

Abstract

Here we describe laryngeal kinematics present during extreme heavy metal singing (EHMS). The aim was to explain how EHMS can be practiced without causing true vocal fold (TVF) trauma. The following EHMS vocal qualities known as growl (Gr), scream (Sc), and high & low pitch false vocal folds scratched voice (HPFSV & LPFSV) were investigated in a single subject. Glottis area was observed with distal chip scope (DCS) trans-nasal flexible fiberscopy, trans-oral HSDP, and acoustics.

Gr and Sc showed that TVF edges touched only occasionally and glottis stayed open for most of the production time while the false vocal folds (FVF) were active. HPFSV and LPFSV were produced with simultaneous steady and symmetric vibration of the TVF and FVF with the TVF adducted and the FVF partially adducted. The sound spectrographs of Gr and Sc showed a spectrum dominated by noise. It was not possible to distinguish the harmonic structure of the TVF-generated sound, proving the virtual absence of the glottic activity in the production of these aggressive sound qualities.

We observed different sets of movements (vibrations) from the supraglottic structures (SGS), generated by FVF, arytenoid caps (AC), and aryepiglottic folds (ARF). These combined vibrations were responsible for creating multi-periodic oscillations that characterize Gr and Sc. We noticed that the activity of SGS during EHMS was different from the sphincter-like activity that we find in dysphonic pathologic phonations that engage SGS.

It is our opinion that the brief duration of the glottic closure during Gr and Sc is to produce the safe range of the subglottic pressure (Ps) and to allow for the generation of high airflow required to set the SGS structures in motion. These factors, in our opinion, allow the EHM singers to produce aggressive voice qualities while avoiding phonotrauma to the TVF.

Keywords: heavy metal, extreme metal, growl, scream, scratched voice, HSDP, DCS, sonogram, supraglottic activity, supraglottic phonation, VF vibration, phonotrauma

Introduction and background

Since the first releases in the 70’s, the musical phenomenon generally known as “Heavy Metal” has produced multiple diversified subgenres, which share elements of vocal aggression and power. Similarly to what can be observed for the instrumental components of EHMS, the use of the voice also differs according to a given EHMS subgenre. In EHMS (e.g., Death Metal, Black Metal), sound production may share elements with Pop and Rock. In general, EHMS vocals entail elements of aggression and such extreme loudness that these sounds lose the ordinary features of the “human voice” and hence purpose-
fully take on the vocal image of a monster or an animal. Sounds known as Gr and Sc have been considered to typify the singing of Death Metal and Black Metal. The HPFSV and the LPFSV also form EHMS styles.

Our previous studies showed that these sounds represent predominantly supraglottic phonations with only partial participation of the TVF [1]. We demonstrated that in Gr and Sc, the TVF and SGS are involved with the simultaneous vibration of the FVF, AC, and AEF. The TVF, when participating, seem to vibrate in a chaotic manner keeping an open position without showing true abduction or adduction. The maintenance of such glottic gap allows for the airflow to activate the SGS while disengaging approximation of the TVF, keeping the subglottic pressure in a safe range. From these same studies, we learned that Gr and Sc can be of low and high pitch with Gr being in the lower part and Sc being in the higher part of the pitch range.

Studies conducted by the Danish group at the Complete Vocal Institute in Copenhagen (completevocalinstitute.com) demonstrated, with the use of conventional laryngovideostroboscopy (LVS), simultaneous presence of multiple sound sources. However, only when HSDP was used did a more specific explanation on sound source become apparent [2-3].

The aim of this study was to investigate EHMS using HDSP to obtain a clearer definition of how the glottic and supraglottic vibrating structures contribute to the creation of these sounds. We also wanted to explain how these potentially abusive and aggressive sounds can be produced without resulting in phonotrauma to the TVF.

**Methods**

We engaged a single male subject, a well-trained expert in EHMS. As expected, he presented with completely normal voice. The subject underwent trans-nasal flexible fiberscopy with distal chip scope (DCS), trans-oral LVS, HSDP, and acoustics. HSDP results were analyzed using Vocalizer® and acoustics were analyzed using both spectrography and Vocalizer®.

The distal chip transnasal fiberscopic exam was carried out with an Olympus ENF-VH HD DCS (3.9 mm in diameter) and a LED CLL-S1 light. During the exam, the subject performed HPFSV, LPFSV, Gr, Sc, and a glissando from Gr to Sc.

HSDP was carried out in color mode with a Kaypentax 90° rigid laryngoscope connected to a Kaypentax color high speed video system (CHSV), model 9710. Recording rate was set at 2000 f/s. During this exam, the subject performed HPFSV, LPFSV, Gr, and a vocal quality that was developed for the present study, which we called soft supraglottic phonation (SSP). It was not possible to analyze Sc using HSDP due to the peculiar resonance needs in producing this quality, which makes it impossible to perform this sound with the rigid laryngoscope inserted in the oral cavity.

All acquired signals were analyzed visually and acoustically by processing with Vocalizer®. Automatic tracing of glottal area waveform (GAW), kymographic segmentation, and Nyquist plots were derived for HPFSV, LPFSV, and Gr.

The sonograms were realized with 600 points (KayPENTAX CLS program) from sounds captured with a microphone standing two meters from the mouth to avoid the distortion of the audio signal, which could easily be induced because of the intrinsically loud nature of these sounds. HPFSV, LPFSV, Gr, Sc, glissando from Gr to Sc, and “A-E-I-O-U” test in both Gr and Sc were analyzed. The obtained sonograms were visually analyzed for presence of harmonic or noisy sound structure, stability and uniformity of the sound itself, and structure and distribution of formants (F0, F1, F2, and F3). In addition, the vocal signals acquired simultaneously during HSDP were processed by Vocalizer®.
Results

Trans-nasal flexible fiberoscopy with DCS

DCS allowed for a clearer description than standard fiberoptic scope of the potential role of the various laryngeal structures during the various vocalizations and the role of the associated structures of the vocal tract.

LPFSV & HPFSV

During LPFSV and HPFSV, the TVF were “tension” adducted with simultaneous closure of the FVF. This was completely different from what can be analyzed in the pathological phonations involving SGS. Complete coverage of the glottis by the FVF was absent with only the anterior third and part of the medial third of the glottis being obstructed by the FVF.

Gr & Sc

During Gr and Sc, we observed activation of the SGS comprising vibrations of the FVF, AC, and ARF. In both of these voice qualities, the TVF appeared to stay practically open during the whole phonatory event. During Gr, the larynx was in a low position with an increase in the latero-lateral and anteroposterior diameters of the vocal tract. During Sc, the larynx appeared in a neutral position, the latero-lateral diameter looked reduced along with a passive distortion of the epiglottis while the anteroposterior diameter appeared to increase. During glissando from Gr to Sc, the phonatory mechanism did not undergo modifications. When the perceptive pitch moved upwards to Sc, the progressive rise of the larynx and the reduction of the latero-lateral diameter of the vocal tract was noted.

HSDP & Vocalizer® Analysis

HSDP offered the detailed vision of the vibrating patterns of the TVF and the SGS, neither of which can be otherwise evaluated by a steady light or by LVS.

LPFSV

During LPFSV, we observed the simultaneous steady and symmetric vibration of the TVF and the FVF. The VF appeared “tension” adducted while the FVF do not completely cover the VF as it is mentioned above (Figure 1). The rate between the vibrating cycles of the FVF and the TVF appears to be 1:2 (i.e., a FVF cycle every two glottis cycles).

Figure 1. HSDP frames of LPFSV (closed and open FVF).
Figure 2 (A-C) shows Vocalizer® analysis of LPFSV taken from HSDP and simultaneous acoustic signal recordings. The analysis from the visual data included glottal area waveform (GAW), kymographic segmentation, and Nyquist plot. The analysis from the acoustic data included signal envelope, spectrogram, FFT, frequency distribution, and Nyquist plot.

**Figure 2A.** Generation of GAW, kymographic segmentation, and Nyquist plot from HSDP of LPFSV using Vocalizer®.

**Figure 2B.** Acoustical analysis of LPFSV using Vocalizer®.
What we have learned about extreme metal production from HSDP

Figure 2C. Comparison of visual and acoustical analysis of LPFSV using Vocalizer®.

HPFSV

During HPFSV, we observed the same vibrating features described for LPFSV (Figure 3), but the rate between the vibrating cycles of the FVF and the TVF appeared to be smaller than 1:2. With the increase of the vibrating cycles of the TVF, in accordance with the highest pitch, no increase in frequency FVF vibration was noted, which appeared to vibrate at the same frequency analyzed for LPFSV.

Figure 3. HSDP frames of HPFSV (closed and open FVF).

Figure 4 (A-C) shows Vocalizer® analysis of HPFSV taken from HSDP and simultaneous acoustic signal recordings acoustics as previously described.
Figure 4A. Generation of GAW, kymographic segmentation, and Nyquist plot from HSDP of HPFSV using Vocalizer®.

Figure 4B. Acoustical analysis of HPFSV using Vocalizer®.
What we have learned about extreme metal production from HSDP

Gr

HSDP clearly shows that during Gr, the TVF touch only occasionally keeping the glottis open almost for the whole duration of the phonation (Figure 5). The vibration of the TVF appears highly asymmetric and irregular in time both in terms of frequency and width of the vibrating cycles.

Both TVF and SGS showed simultaneous different sets of “mucosal waves” (Figure 6), some of which are characterized by higher frequency, smaller width, and shorter propagation, and others characterized by a lower frequency, larger width, and longer propagation.
Figure 7 (A-C) shows Vocalizer® analysis of Gr taken from HSDP and simultaneous acoustic signal recordings as previously described. As remarked previously, it was not possible to evaluate Sc in HSDP because the shapes of the vocal tract, the mouth, and especially the tongue in this voice quality did not allow for the study with rigid laryngoscopy. However, from the study of Sc with DCS, we noticed how the phonatory mechanisms of Gr and Sc can overlap. Therefore, we feel justified to extrapolate the information obtained from Gr in HSDP to Sc as well.

Figure 6. HSDP frame of extreme supraglottic activity during abducted glottis in Gr.

Figure 7A. Generation of GAW, kymographic segmentation, and Nyquist plot from HSDP of Gr using Vocalizer®.
Figure 7B. Acoustical analysis of Gr using Vocalizer®.

Figure 7C. Comparison of visual and acoustical analysis of Gr using Vocalizer®.
SSP

To question if it was possible to activate the SGS without obtaining a contact of the free edges of the TVF, we studied a vocal style we termed “Soft Supraglottic Phonation” (SSP). This type of voice is a quiet supraglottic phonation, with a low level of loudness but with the perceptive traits of Gr. This quality (it must be kept in mind) does not have any artistic relevance to EHMS, as it was developed only for the sake of this investigation.

The HSDP results showed the same elements which were found in Gr, but with a clearly smaller width of the mucosal waves. The TVF showed the same phonatory behavior found in Gr but without the edges ever contacting during the whole phonation (Figure 8).

![Figure 8. HSDP frames of SSP (TVF never touch each other).](image)

_Acoustics by Sonography_

Gr and Sc showed a spectrum dominated by noise. It was not possible to distinguish the harmonics, proving irrelevance of the glottic activity in the production of these sound qualities.

Figure 9 shows that most of the energy during Gr is concentrated below 1000 Hz with the position of the third formant at about 2500 Hz, which appears functional in the artistic performances where EHMS is employed (i.e., to make the voice cut through the massive sound wall, which is typical of this genre).

![Figure 9. Sonogram of Gr.](image)
In Sc, on the contrary, a wide energy distribution above 1000 Hz can be observed with no structuration of the third formant. Below 1000 Hz, the concentration of energy appears to be slightly reduced (Figure 10).

![Figure 10. Sonogram of Sc.](image)

During the “A-E-I-O-U” test in quality of Gr and Sc, it was possible to observe the position of F1 and F2. During Gr, F1 appears attracted towards the low frequencies with a strongly limited movement between vowels (Figure 11). During Sc, the excursion of F1 and F2 does not show substantial differences if compared to the same test carried out in conversational voice (Figure 12).

![Figure 11. Sonogram of “A-E-I-O-U” in Gr voice.](image)

![Figure 12. Sonogram of “A-E-I-O-U” in Sc voice.](image)
The analysis of the glissando from Gr to Sc shows how the energy distribution of the sound changes gradually from the architecture typical of Gr to that typical of Sc through various intermediate degrees (Figure 13). HPFSV and LPFSV show diplophonia. In this case it is important to notice stability in time and in the energy distribution, which differs from the pathological diplophonia.

![Sonogram of glissando from Gr to Sc.](image)

**Figure 13.** Sonogram of glissando from Gr to Sc.

**Discussion**

EHMS is characterized by an aggressive acoustic sound that evokes perception of severe vocal abuse. Nonetheless, during the daily use of the voice, well-trained performers do not show any elements of dysphonia. Our study explains this dichotomy and offers, thanks to HSDP examination, a clear illustration of the phonatory physiology present in these voice qualities.

All the vocal types studied thus far showed traits of multiphonia, making it possible to observe the simultaneous activation of different structures in the generation of sound. In particular Gr and Sc proved to display a multiphonic and multiperiodic phenomena with different active structures simultaneously during phonation, and different sets of mucosal waves at different frequencies and width being active simultaneously with the same structure. All of this occurred without hard glottic attacks and only with occasional contact of the VF edges, clearly in opposition to what happens in the pathological phonations that involve the same structures.

It is our opinion that the highly brief duration of the glottic closure during Gr and Sc is behind the persistence of Ps at a safe range and allows for high airflow to occur. Such high airflow, in our opinion, is required to activate oscillations of the SGS structures. With the HSDP analysis of SSP, we were able to clearly show how the vibration of the SGS can also be obtained without abusive influence on the VF.

**Conclusions**

Combining the data obtained from the DCS, HSDP, Vocalizer®, and sonograms, we have shown how the EHMS is produced without causing TVF phonotrauma. In particular, for the most extreme qualities of EHMS (i.e., Gr and Sc), the HSDP showed how the free edges of the TVF touch only occasionally during the phonatory event, keeping the subglottic pressure within a safe range and guaranteeing an adequate airflow, which is necessary for the activation of the SGS. The study of the SSP also showed that the artistic supraglottic activation present in EHMS differs from supraglottic phonations present in pathological conditions.
References


