Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Perceptual and Acoustic Study of Voice Quality in High-Pitched Heavy Metal Singing

Alexsandro R. Meireles[1]

[1] Phonetics Laboratory, Federal University of Espírito Santo, Brazil
meirelesalex@gmail.com

Abstract

This paper studies high registers of heavy metal singing based on the voice profile analysis scheme (VPAS) and acoustic correlates. The f0 range varied from 366 to 666 Hz. The application of VPAS for singing is unique in the phonetic literature. Two professional and two amateur singers sang *Iron Maiden*'s *Aces High* with instrumental playback through headphones. Two very high-register excerpts were selected from this song to verify the vocal strategies used by experienced singers while singing at extreme registers of vocal extension. Experienced judges (vocal coaches, speech therapists, phoneticians) analyzed their vocal productions by perceptual analysis of voice quality and voice dynamics with VPAS. The acoustic analyses were run with the software VoiceSauce that automatically extracted thirteen parameters of long-term measures (H1, H1H2, H1A3, CPP, Energy, HNR05, HNR15, HNR25, HNR35, F1, F2, B1, B2). Results indicate that the two groups of singers use distinctive articulatory strategies in singing at high registers and that settings strategies are factors that influence these measures. Although both groups of singers used tense vocal tract and larynx settings, open jaw and raised larynx settings were only found in the professional voices. These different articulatory settings were statistically corroborated by the acoustic analysis.

Index Terms: singing; voice quality; heavy metal; auditory perception; acoustic analysis

1. Introduction

Heavy metal is characterized by a combination of vocal settings such as pharyngeal constriction, raised larynx, tense vocal tract and larynx, and complex modes of phonation (falsetto, creaky voice, harsh voice, whispery voice). Some varieties also use wide vocal range and timbre variation. Examples of the combination of long-term complex vocal settings can be found in songs written by classic heavy metal bands such as Iron Maiden and Judas Priest, and hard rock bands such as Guns N' Roses and Aerosmith. Furthermore, the adoption of these settings requires special attention in clinical voice analysis as it may result in future speech pathologies.

Metal singers use many types of voice adjustments technically known as vocal drives, which can be physiologically produced with different vocal tract configurations [15]. Among these techniques, belting is one of the most used in the high notes of hard rock and heavy metal (eg. Bruce Dickinson, *Iron Maiden*'s lead singer).

Belting is originated in musical theater and then has expanded to popular genres such as pop music, rhythm and blues, country, rock, heavy metal and jazz. Popeil [16, p.79] states that "belt production is characterized by: 1) thicker edge of vocal fold; 2) tenser TA (vocalis) muscle; 3) lack of zippering action in vocal opening/closing—more of a clapping action; 4) high speed quotient—folds snap shut quickly; 5) high closed quotient—longer closed phase (over 50%); 6) increased sensation of breath-holding; 7) heightened activity of jaw and extrinsic laryngeal muscles; 8) possible pulling forward of hyoid bone; 9) higher larynx position, but singer has some lifting and lowering ability; 10) epiglottis more horizontal, lessening space in vallecula; 11) increased "support"/subglottal pressure; 12) even distribution of amplified harmonics up to 4 kHz and spectral energy above 15 kHz.

Gonsalves and colleagues [7, p. 196] discuss numerous studies on singing voice. However, the majority of these works have studied classically trained singers who are trained in traditional singing methods, and constitute a minority among singers. Other singing styles, such as pop, rock, samba, country, and others, require further study, as each musical style possesses its own characteristics and specific vocal techniques. Yet, their perceptual results are contradictory, since, despite showing rock presented the greatest vocal tension among the styles they have studied, this tension was analyzed as light and moderate. Besides, they say that muscular tensions do no seem to be affected by the vocal technique and also that tense singing may result in normal laryngoscopy [7, p.199].

Even though rare, some studies have focused on rock singing, such as the study by Oliveira and Behlau [14], Thalen and Sundberg [17], and Gonsalves and colleagues [7]. However, the style of singing used there is not vocally related to the vocal style of this paper. Despite drawing inspiration from 1960s' rock [2], as declared by members of bands that gave rise to heavy metal such as Black Sabbath and Motorhead, heavy metal is a much more aggressive variation of classical rock and was explored to an even lesser degree in academia.

Similarly to Gonsalves et al. [7], Butte and colleagues [3] conducted an analysis of different musical styles but did not examine the rock genre. They justify the absence of rock because they did not record audio that could be acceptable for acoustic analysis, since the instruments were contained in the voice material. On the other hand, differently from the other study, the authors have used acoustic analysis to interpret the data.

Kent and colleagues [6] affirm that acoustic analysis, though objective and reliable, should be combined with perceptual analysis of voice, which provide an overview of vocal attributes by analyzing the effect that the sound has on the listener. To that end, the research group coordinated by

 doi: 10.21437/SpeechProsody.2016-256

John Laver in the 1980s developed a profile for describing voice quality based on the perception of auditory, acoustic, articulatory and physiological data (Voice Profile Analysis Scheme, VPAS, cf. Laver et al., 1981). This study provided a foundation for further auditory-perceptual studies on voice.

Camargo and Madureira [4] translated and adapted the VPAS profile to Brazilian Portuguese for voice quality and voice dynamics examination of phonetic and speech therapy analysis in Brazil. However, this profile is not designed specifically for the singing voice and may need to be adapted for this purpose. Among the vocal settings used by singers that needs to be interpreted in VPAS, we can cite, for example, epiglottal and false vocal folds adjustments, which are used in Mongolian singing throat and death metal style (eg. Max Cavalera, former lead singer of *Sepultura*).

In addition to the above, this study specifically aims at describing the complex interactions of voice quality settings in singing with perceptual and acoustic analysis, so as to contribute for a scientific investigation of the voice in heavy metal. Moreover, this study contributes for the correlation between acoustic and perceptual data on singing, since there are very few studies on the field, and adds the heavy metal style to the possibilities of research.

2. The Voice Profile Analysis Scheme

Laver [8], based on Abercrombie's study [1], defines voice quality as all the vocal characteristics that are related to speech, including laryngeal, supralaryngeal, and muscular tension and voice dynamical features. Thus, both linguistic and extralinguistic, as well as paralinguistic aspects of speech are examined in this study. It is important to remind that only vocal aspects that frequently occur in speech should be considered in the analysis. According to Laver [9, p. 152], these long-term voice characteristics correspond to "all factors which can potentially be prolonged beyond the domain of the segment".

Based on studies by Laver, a group of researchers from Queen Margaret University College in Scotland developed a voice quality analysis profile called VPAS (Vocal Profile Analysis Scheme, see [10, 11, 12, 13]. The protocol is based on the neutral configuration of the vocal tract, which is similar to the schwa vowel. Based on this neutral setting, settings that deviate from this voice configuration are described. Laver [8] discusses 53 possible types of settings.

As Camargo and Madureira [4] pointed out, "the adoption of the neutral setting as reference can be considered a landmark in the investigation of voice qualities since it does not introduce a rupture between normality and voice disorder". Furthermore, this methodology i) allows for comparisons of voice quality among dialects and different languages, as the neutral standard of vocal tract configuration is universal, and ii) eliminates the classic dichotomy between normal and pathological speech. Any voice quality setting can be replicated with this profile, and it prevents us from determining a priori whether a particular type of voice quality is attributable to speech pathology.

The 2007 VPAS version [12] (adapted to Brazilian Portuguese by [4]) is divided into three large sections: 1) vocal tract features (supralaryngeal settings); 2) overall muscular tension (laryngeal and supralaryngeal settings); and 3) phonation features (laryngeal adjustments). In addition to these settings, four parameters of voice quality and voice dynamics are described: 1) pitch (f0 variation); 2) loudness (variation in dB); 3) time (including speech continuity and rate); and 4) respiratory support features.

The VPAS method involves perceptual analysis by experienced judges in two steps. The first step involves defining whether a non-neutral setting is present. If a non-neutral setting is present, we go to the second step, in which settings are classified as either moderate (1 to 3) or extreme (4 to 6). Camargo and Madureira [5, p. 80] summarize the settings as follows: 1 – small difference from the neutral setting; 2 – slight difference from the neutral setting; 3 – moderate difference from the neutral setting; 4 – noticeable difference from the neutral setting; 5 – significant difference from the neutral setting; and 6 – extreme difference from the neutral setting. Camargo and Madureira [5, p. 95] attest to the validity of VPAS, which provides "the rupture of the line that distinguishes normality from voice alteration, allowing one to analyze various situations that involve the voice, such as disorder rehabilitation and voice professional assistance." Extending the use of VPAS to the professional voice, we propose a unique application of the method to the professional voice: the analysis of singing. As mentioned before, due to singing characteristics, we will need to adapt the protocol in the future for an appropriate description of voice quality in singing.

3. Methodology

Four singers between 20 and 40 years of age (2 professionals: J and A1; 2 amateurs: A2 and Informant I) sang *Iron Maiden*'s *Aces High* with instrumental playback through headphones. The playback served to assist the singers with tone and recording time uniformity. Only the audio from the singers' voices was recorded so that the harmonies did not interfere with the sample evaluation. Materials for the above analysis were produced in a recording studio at the Federal University of Espirito Santo, Brazil, for informant A1 and in a private recording facility (*Ocean Studios*, Brazil) for informants J, A2 and I. Both recordings were produced at a sampling rate of 44.1 kHz stereo and stored as .wav files. In the first studio, a condenser microphone was used, and in the second facility, a unidirectional cardioid microphone was used. In addition to collecting audio recordings, video recordings of the vocal performances were also captured. For the acoustic analysis we used the software VoiceSauce [19] that automatically extracted thirteen parameters of long-term measures (H1*[1], H1*H2*, H1*A3*, CPP, Energy, HNR05, HNR15, HNR25, HNR35, F1, F2, B1, B2. For a detailed description of these values, see [18, 19].

The experienced singers matched the professional profile, as informant A has been a vocalist for more than 15 years and sings in an *Iron Maiden* cover band, and Informant J is graduated in vocal performance at the School of Music of Espírito Santo and is the lead singer of a famous heavy metal band in Espírito Santo, Brazil. The amateur singers A2 and I do not possess any singing experience in singing at high registers and do not sing in heavy metal bands. Besides, singers I and A2 have studied singing for only four months.

Two very high-register excerpts were selected from the song *Aces High* to verify the vocal strategies used by

[1] The asterisk (*) means that spectral magnitudes H1, H2 and A3 have been corrected for the effects of formants [18, 19].

experienced singers while singing at extreme registers of vocal extension. The excerpts are listed as part 1 and part 2 below (figure 1).

After preparing the sample, the data were subjected to a VPAS analysis. The judges, who have large experience in analyzing data with VPAS, included three speech-language pathologists, one phonetician, one lyrical singer, and professors and/or graduate students of PUC-SP. We recall that to be a VPAS evaluator, one must take a minimum of 20 hours of practice with the method at the University of Glasgow or at qualified centers around the world. One of these centers is situated at the Integrated Acoustic Analysis and Cognition Laboratory (LIAAC) at PUC-SP, Brazil.

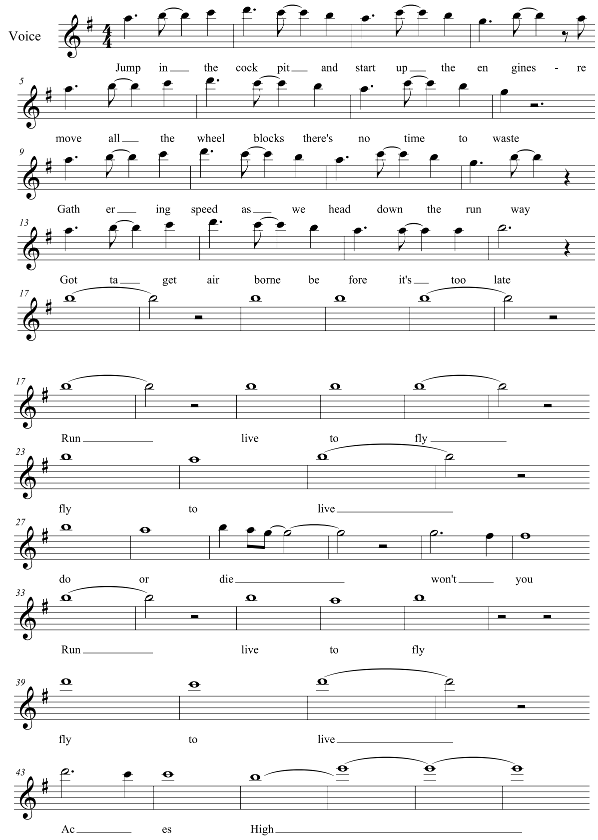

Figure 1: *Parts 1 (measures 1-17) and 2 (measures 17-45) of Iron Maiden's Aces High.*

4. Results

4.1. Perceptual analysis

The amateur singers (I, A2) varied less their voice qualities while singing at extremely high notes (above tenor C note). The judges observed only overall vocal tract and larynx tension settings for these singers. Both exhibited grade 3 for vocal tract and larynx tension. However, as these singers used different strategies while performing musical notes, different phonation settings were perceived.

A2 used modal voice in the beginning of Part 1 [00:00 to 00:14] then changed to falsetto [00:15 to 00:30]. In Part 2, the singer used modal voice in "Run, live to fly, fly to live, do or die. Run" [00:45 to 01:00] and falsetto in "Live to fly, fly to live. Aces high" [01:01 to 01:13]. The passages in which falsetto was used are those that correspond with the highest notes in the song. On the other hand, singer I predominantly used a harsh, whispery voice throughout the song.

The professional singers, while also using muscular tension settings, demonstrated higher levels of these settings. Regarding vocal tract and larynx tension, A1 received a score of 5 in both settings, and singer J, a score of 6. Furthermore, we detected the following settings: open jaw (A1: grade 2; J, grade 4), raised larynx (A1, grade 4; J, grade 5), minimized tongue body range (J, grade 2), and backed tongue body (J, grade 2). As for phonation settings, modal voice with creak (creaky voice) predominated in both singers. This type of setting corresponds to what is expected in heavy metal style.

4.2. Acoustic analysis

A t-test of subjects, gathered together as professionals (A1, J) and amateurs (A2, I), as a function of the 13 voice quality measures (see tables 1 and 2) revealed that these categories of singers were statistically different from each other. Except for H1A3 and NR35 for part 1, all other comparisons were highly significant ($p < 4e^{-16}$). We highlight that the big amount of data contributed for this significance (Part 1: N=57390; Part 2: N=94124). Also, it supported a long-term acoustic analysis of voice quality in the same manner of the perceptual analysis.

We also tried to correlate the main differences between the groups of singers. As can be noticed in section 4.1, the evaluators perceived tense vocal tract and tense larynx for all speakers. Yet, open jaw and raised larynx were only found in the professional voices. The F1 data on Tables 1 and 2 indirectly corroborate that the professional speakers used a more open jaw and higher larynx positions than the amateur singers.

Table 1. *Voice quality measures (mean) for Part 1. S* stands for Subject; *N*, HNR.

S	H1	H1-H2	H1-A3	CPP	Energy
A1	-4.1	-0.04	-3.3	20.7	4.34
J	-11.9	-0.9	-6.9	20.1	0.8
A2	-11.6	-1.9	-5.4	20.4	0.3
I	-2	-6.9	-4.8	20.0	2.41

S	F1 (B1)	F2 (B2)	N5	N15	N25	N35
A1	685 (236)	1574 (200)	37	31	30	33
J	698 (225)	1692 (230)	34	31	31	32
A2	566 (184)	1572 (263)	34	36	36	37
I	467 (119)	1577 (280)	24	26	26	28

Table 2. *Voice quality measures (mean) for Part 2. S* stands for Subject; *N*, HNR.

S	H1	H1-H2	H1-A3	CPP	Energy
A1	4.9	-0.2	-5.2	24.3	8.5
J	-5.6	1.8	-6.1	24.1	1.37
A2	-7.8	8.6	10.7	6.2	0.73
I	-5.1	-3.6	0.2	22.9	0.77

S	F1 (B1)	F2 (B2)	N5	N15	N25	N35
A1	732 (272)	1573 (199)	48	36	36	38
J	725 (175)	1642 (204)	43	38	38	38
A2	638 (232)	1558 (243)	50	48	50	50
I	521 (221)	1456 (296)	29	30	34	36

Among the acoustic parameters that seem to be correlated with the perceptual parameters (tense vocal tract and tense larynx), H1H2, H1A3, Energy and HNR were the most robust (see tables 1 and 2). The statistical analysis has shown that H1H2 is higher for the professional speakers in part 1, suggesting a more breathy voice for the professionals ([21] have showed that higher levels of H1H2 are usually correlated with breathy voice). Nevertheless a stronger Energy level is added to the quality of this professional breathy voice. Also, we found, in part 1, higher values of HNR for the professional speakers, suggesting a more modal phonation than the amateur speakers ([20] suggest higher values for modal phonation). Nevertheless, in part 2, an inverted pattern occurred, the professional singers showed lower HNR levels suggesting higher degree of breathiness than the other singer group. In addition, as in part 1, the professional singers showed in part 2 higher Energy levels. Yet, the spectral slope (H1A3) and H1 were smaller for the professional speakers, what suggests an addition of air escape in the vocal folds but with the addition of high acoustic energy.

Finally, we would like to comment that all these measures were made based on the singing of very high notes, what generates very high f0s compared to the normal male voice, which is known to be around 100 Hz (see table 3).

Table 3. *f0 mean (max) values for Parts 1 and 2*

Subject	Part 1	Part 2
A1	459 (600)	366 (612)
J	460 (667)	398 (630)
A2	447 (589)	400 (598)
I	463 (666)	437 (572)

5. Discussion

Taking into consideration that the VPAS was developed based on the perceptual ability of evaluators to hear voices as complementary combinations of settings, whose method can be learned [5, p. 89], we verified the validity of applying the protocol to studies on singing given that these combinations of complementary settings are long-term muscular settings of vocal tract structures.

Regarding the analyzed singing material, despite the clear distinction between articulatory strategies applied by the amateur (A2, I) and experienced (A1, J) singers, individual strategies were also found for extreme register singing. The professional singers were most significantly distinct in their continuous maintenance of high tension in the vocal tract and the vocal folds, which was found intermittently and to a lesser degree among the amateur singers. Additionally, the experienced singers held an open jaw position and raised larynx vocal posture to reach the high notes. These settings were not observed in the amateur singers.

Furthermore, it is important to emphasize that singer J experienced greater difficulty in reaching the high notes than singer A1. According to singer J, this difficulty was not directly related to the presence of high notes but to the aggressive heavy metal vocal style employed. The subject reported being able to sing even higher notes in his metal band but using a different vocal style than the style studied. We thus plan in the future to examine settings that singer J uses to reach even higher notes

6. Conclusions

This paper has presented a new methodology for studying voice quality on singing: the use of VPAS allied to acoustic analysis. As we have presented here, this combination of procedures was able to differentiate vocal strategies for singing at high registers. We acknowledge, nevertheless, that other analysis such as EGG, ultrasound and MRI may refine the understanding of the strategies used by professionals to sing extreme high notes in heavy metal. Therefore, in the future, we intend to complement the results found here with articulatory and/or physiological analysis.

Finally, an adaptation of VPAS is essential for a better perceptual analysis of singing, as some available settings are irrelevant to singing while other settings may need to be reinterpreted in the protocol (such as vibrato rate and laryngeal settings related to vocal drives). To do so, we will build a database to define optimal parameters for studying various singing styles and relate them to voice quality settings.

7. Acknowledgements

We thank Dr. Zuleica Camargo, Dr. Sandra Madureira and the students of the Graduate Program in Applied Linguistics and Language Studies at PUC-SP for their perceptual evaluations of this study. We also thank Fapes and CNPq for financing the study, and the vocal coach Ariel Coelho for detailing some vocal aspects of Bruce Dickinson's style (*Iron Maiden*'s lead singer). Finally, a special thanks to Dr. Aline Pessoa-Almeida for discussing the paper.

8. References

[1] D. Abercrombie. *Elements of general phonetics*. Edinburgh: Edinburgh University Press, 1967.

[2] G. Bayer. *Heavy metal music in Britain*. London: Ashgate popular and folk music series, 2009.

[3] C. Butte; Y. Zhang, H. Song and J. Jiang. "Perturbation and Nonlinear Dynamic Analysis of Different Singing Styles". *Journal of Voice*. v.23, p.647-652, 2009.

[4] Z. Camargo and S. Madureira. "Voice quality analysis from a phonetic perspective: Voice Profile Analysis SchemeProfile for Brazilian Portuguese (BP-VPAS)". In: *Fourth Conference on Speech Prosody - Abstract book and Proceedings*. Campinas: Capes, Fapesp, CNPq. v.1, p.57-60, 2008a.

[5] Z. Camargo and S. Madureira. "Avaliação vocal sob a perspectiva fonética: análise preliminar". *Distúrbios da Comunicação*. v.20(1), p.77-96, 2008b.

[6] R. D. Kent, G. Weismer, J. F. Kent, H. K. Vorperian and J. R. Duffy. "Acoustic studies of dysarthric speech: methods, progress, and potential". *Journal of Communication Disorders*, v.32, p.141-186, 1999.

[7] A. Gonsalves, E. Amin and M. Behlau. "Análise do grau global e tensão da voz em cantores de roque". *Pró-Fono R. Atual. Cient*. [online]. vol.22, n.3, p.195-200, 2010.

[8] J. Laver. *The phonetic description of voice quality*. Cambridge: Cambridge: Cambridge University Press, 1980.

[9] J. Laver. Principles of phonetics. Cambridge: Cambridge University Press, 1994.

[10] J. Laver. "Phonetic evaluation of voice quality". In: *Voice quality measurement*. R.D Kent, Ball M.J. (ed). San Diego: Singular Publishing, p.37-48, 2000.

[11] J. Laver, S. Wirs, J. Mackenzie and S. M. Hiller. *A perceptual protocol for the analysis of vocal profiles. Edinburgh; Edinburg University*. Department of Linguistics; p.139-55. [Workin Progress, 14], 1981.

[12] J. Laver and J. Mackenzie-Beck. *Vocal Profile Analysis Scheme – VPAS*. Queen Margareth University College-QMUC, Speech Science Research Centre, Edinburgh, 2007.

[13] J. Mackenzie-Beck. "Perceptual analysis of voice quality: the place of vocal profile analysis". In: *A figure of speech: a festschrift for John Laver*. W.J. Hardcastle, J. Mackenzie-Beck (ed). Mahwah: Lawrence Erlbrum, p.285-322, 2005.

[14] L. Oliveira and M. Behlau. *Perfil vocal de cantores amadores de banda de roque*. [monografia]. São Paulo(SP): Centro de Estudos da Voz, 2004.

[15] G. Pecoraro, A. Duprat, S. Bannwarth and M. Andrada e Silva. "Cantores de rock: ajustes dinâmicos de trato vocal, análise perceptivoauditiva e acústica das vozes ao longo de cinco décadas". In: *Anais do 18o Congresso Brasileiro de Fonoaudiologia*. Curitiba, 2010.

[16] L. S. Popeil. "The multiplicity of belting". *Journal of Singing*, v.64, n.1, p.77-80, 2007.

[17] M. Thalen and J. Sundberg. "Describing different styles of singing- a comparison of a female singer's voice source in ''Classical'', ''Pop'', ''Jazz", and ''Blues". *Log Phon Vocol*. v.26, p.82-93, 2001.

[18] Y.-L. Shue. *The Voice Source in Speech Production: Data, Analysis and Models*. PhD Thesis, UCLA, 2010.

[19] Y.-L. Shue, P. Keating, C. Vicenik and K. Yu (2011) VoiceSauce: A program for voice analysis, *Proceedings of the ICPhS XVII*, 1846-1849, 2011.

[20] E. Yumoto, W. Gould and T. Baer, T. ''Harmonics-to-noise ratio as an index of the degree of hoarseness''. *J. Acoust. Soc. Am.* 71, 1544– 1550, 1982.

[21] P. Keating, C. M. Esposito, M. Garellek, S. u. D. Khan, J. Kuang. "Phonation Contrast across languages". UCLA Working Papers in Phonetics, No. 108, pp. 188-202.