A SPECTROGRAPHIC ANALYSIS OF VOCAL TECHNIQUES IN EXTREME METAL FOR MUSICOLOGICAL ANALYSIS

Eric Smialek (1, 2) Philippe Depalle (2, 3) David Brackett (1)
CIRMMT, McGill University, Montréal, QC, Canada

ABSTRACT

Extreme metal genres such as death metal and black metal force music analysts to seek alternative methods to Western notation-based analysis, especially when one asks what means of expression their vocalists may draw from in order to seem convincingly powerful to fans. Using spectrograms generated by AudioSculpt, a powerful sound analysis, processing, and re-synthesis program, this paper demonstrates a mixed application of spectrograms and conventional music analysis to vocals in two separate contexts: an a cappella recording in a soundproof laboratory and a commercial recording with a full band. The results support an argument for the utility of spectrograms in revealing articulations and expressive nuances within extreme metal vocals that have thus far passed unnoticed in popular music scholarship.

1. INTRODUCTION

Popular music scholars have long insisted that details of musical sound such as rhythmic and melodic inflections or timbral characteristics are central in importance to popular musicians and their audiences [1, 2, 3, 4]. Accordingly, some of the leading popular music scholars such as Richard Middleton [1] and Philip Tagg [2] have expressed criticism of analyses of popular music that overlook these features, often due to a reliance on Western music notation. One alternative is to visually represent sound through spectrograms, and it has now been nearly thirty years since Robert Cogan made his strong case for their utility to model musical sound in a way that can account for melodic, rhythmic, or timbral nuances [5]. Since then, despite the importance popular music scholars have placed in these features, they have rarely used spectrograms to study them. In the few instances where they have appeared, such as [6], reviewers have argued that spectrograms are superfluous or have voiced a general distrust of analytical technology such as spectrum photography or digital signal processing of spectral imagery [7,8,9].

Such a distrust of musical spectrograms has not, of course, extended to fields such as electroacoustic composition and analysis where there exists an epistemological tradition that has long supported the use of music technology and where the limitations of Western notation are obvious [10]. With this in mind, the study of extreme metal presents something of a disciplinary middleground in its foundations in popular music studies and its incompatibility with Western notation. Because extreme metal vocalists place greater importance on the timbral variations of vocals than on pitch and harmony, methods of analysis that rely on conventional notation do not reveal much information about their expressive screams. Extreme metal vocalists thus provide an opportunity to analyze the role and possibilities of spectrographic technology to reveal information about musical expression in a genre of music for which little to no analytical methods have been established. Using real-time spectrographic displays, this paper will demonstrate how spectrograms can be useful research tools for scholars who require or seek alternatives to notation-based analysis.

2. THE EXTREME METAL VOICE

2.1. Basic Aspects of Vocal Production

To produce the vocal sounds characteristic of death metal and black metal (as well as related sub-genres), extreme metal vocalists pass air through the *vocal cords* (or “false vocal chords”) located a few millimetres above the vocal folds (see [12] for anatomical details). This allows extreme metal vocalists to achieve the large spectral spread of energy visible in spectrograms.

Extreme metal screams can be performed by either inhaling or exhaling, resulting in two very distinct styles of screaming. The different directions of air flow can be thought of as akin to the linguistic distinction made between voiced and unvoiced methods of articulating consonants: when performing exhaled vocals, one’s larynx vibrates, indicating that the vocal cords are vibrating—rather forcefully—whereas this vibration does not occur with inhaled vocals. This basic difference has a profound effect on the overall sound quality produced, the ease with which different articulations can be made, the ability for a vocalist to sustain a long scream, and the degree of strain put on the voice.

Figure 1 demonstrates some of the acoustic differences between inhaled and exhaled vocals. Here, a volunteer extreme metal vocalist was asked to freely inhale or exhale, resulting in two very distinct styles of screaming. The different directions of air flow can be thought of as akin to the linguistic distinction made between voiced and unvoiced methods of articulating consonants: when performing exhaled vocals, one’s larynx vibrates, indicating that the vocal cords are vibrating—rather forcefully—whereas this vibration does not occur with inhaled vocals. This basic difference has a profound effect on the overall sound quality produced, the ease with which different articulations can be made, the ability for a vocalist to sustain a long scream, and the degree of strain put on the voice.

Figure 1 demonstrates some of the acoustic differences between inhaled and exhaled vocals. Here, a volunteer extreme metal vocalist was asked to freely

---

Footnotes:

1. A similar process is used in Mongolian throat singing where singing voice formants are used to convey melodic information [13].

2. This technical information offered here is based on [11].
A SPECTROGRAPHIC ANALYSIS OF VOCAL TECHNIQUES IN EXTREME METAL FOR MUSICOLOGICAL ANALYSIS

Eric Smialek (1, 2) Philippe Depalle (2, 3) David Brackett (1)
CIRMMT, McGill University, Montréal, QC, Canada

ABSTRACT

Extreme metal genres such as death metal and black metal have long been overlooked by music scholars. The focus of popular music scholarship has often been on rhythm and timbral characteristics, which are central to the experience of extreme metal. As a result, much of the technical information about extreme metal vocal techniques is known only to those who are well versed in the technical aspects of these genres. In this paper, we present some basic distinctions between extreme metal vocal techniques as demonstrated by a volunteer extreme metal vocalist.

1. INTRODUCTION

Popular music scholars have long insisted that details of musical sound such as rhythmic and melodic inflections or timbral characteristics are central in importance to popular musicians and their audiences [1, 2, 3, 4]. Accordingly, some of the leading popular music scholars such as Richard Middleton [1] and Philip Tagg [2] have expressed criticisms of the analyses of popular music that overlook these features, often due to a reliance on Western music notation. One alternative is to visualize the vocal sound through spectrograms, and it has now been nearly thirty years since Robert Cogan made his case for the utility of spectrographic technology to reveal articulations and expressive nuances within extreme metal vocals that have thus far passed unnoticed in popular music scholarship.

2. THE EXTREME METAL VOICE

2.1. Basic Aspects of Vocal Production

To produce the vocal sounds characteristic of death metal and black metal (as well as related sub-genres), extreme metal vocalists pass air through the ventricular folds (or “false vocal chords”) located a few millimetres below the vocal folds (see [12] for anatomical details). This allows extreme metal vocalists to achieve the large spectral spread of energy visible in spectrograms [5]. Extreme metal vocalists can be classified using spectrograms. Vocalists will tend to keep the vocal cords close together to shorten the vocal tract length for a higher scream. They will work the vocal cords and do a raspy voice, or spread the cords to produce a lower, more raspy voice.

3. SIGNAL PROCESSING AND RE-SYNTHESIS USING AUDIOSCULPT

In order to investigate the aforementioned features of the extreme metal voice and their roles in music pieces, we decided not only to perform spectrographic analyses, but also to extract parameters that allow for a partial resynthesis of key sound components in order to validate our assumptions. To explain, the following is a brief description of the basic principles underlying AudioSculpt functionalities: AudioSculpt generates a short-term Fourier transform (STFT) representation of the sound. In order to optimally display the sonogram image, STFT parameters must be chosen including analysis window size and shape, step size (that determines how the STFT fits over the signal characteristics), and fast Fourier transform (FFT) sampling. For each sound example in this study, we used a Blackman window of size M=2400 with a step size of M/8 and a number of channels equal to 4096. Once a satisfactory image is in view, it is ready to be analyzed.

Qualitative judgments can be made based on the obtained spectrographic image but more precise measurements of vowel formant quality are also taken by first synthesizing a new sound file based on the original sound’s vowel formants using one of AudioSculpt’s several spectral gain filters (in each case during this study, the pencil filter tool). These filters change the gain of certain frequency regions defined by the user within the resynthesized signal. For this study, a similar process is used in Mongolian throat singing where singing voice formants are used to convey melodic information [13].

The technical information offered here is based on [11].
the regions to be filtered were defined according to a computer-based analysis of the original sound file’s partitions. Specifically, AudioSculpt’s “Partial Tracking Analysis” feature was used to track partitions utilizing multiple breakpoint functions for each sinusoidal component (a procedure that works with inharmonic signals as well as harmonic ones). A new sound was then generated by amplifying the formant regions of the original signal.

AudioSculpt’s diapason tool was then used to take frequency readings from the synthesized partitions. When the tool is pointed at a particular place on the sonogram, its frequency and amplitude are displayed, a corresponding sine tone sounds, and a two-dimensional spectrogram of the synthesized signal appears. This procedure allows for an analysis of the original sound by hypothesizing that certain components of it appear to be important, re-synthesizing those components, and, consequently analyzing them.

4. A RECORDED IMPROVISATION

In order to investigate the features of extreme metal voices presented in section 2 in a more musical context, we asked a vocalist to perform a vocal improvisation using whatever extreme metal vocal techniques he wished. Given the somewhat artificial performance environment—the room was virtually devoid of reverberation and there were no instruments to accompany the vocalist—the volunteer vocalist’s improvisation shown in Figure A1 (see URL1), if not an exact replication of the vocalist’s intended performance, might make in a concert setting or full-band recording, from exhales to those familiar with extreme metal music occur in the recording, they appear in the chart beneath one another whenever a portion of their rhythms match other segments (e.g. measures 2–6 and 11–12 have common rhythms for beats 3 and 4). Thus, a comparative slice of musical syntax would be drawn.4

The vocalist also appears to have reserved at least one of the vowels with especially high formant region reaching upwards of 4000Hz, a circumstance inhaled portions of the improvisation come close to this. The vocalist’s screams steadily increase in volume and dynamic markings at measure 7); each half contains a difference in the highest regions of spectral energy frequently invoked in discussions of basic musical composition: the improvisation shows a clear binary division into two roughly equal sections of music, separated by a pause that can be executed quickly; similarly, the quick eighth-note alternations between /i/ (the r-colored vowel in “sure”) and /u/ (as in “heud”) in measure 12 are easily performed because they do not require elaborate jaw motions, only quick shifts between lip and tongue positions.

4.2. The Importance of /i/ Sounds

This last point raises one of the most conspicuous consistencies observable throughout the recording sample. The vocalist very frequently alternates between /i/ and /u/ sounds with both inhaled and exhaled vocals. These phonemes occur for nearly all the articulations contained within the solid box given in the left-centre column and, in the case of the one exception, the /i/ combination is merely displaced by a beat, and this is marked by the arrow in the left-centre column. The versatile /i/ of Standard North American English is especially worthy of emphasis here for both the acoustical and physiological advantages it offers the vocalist. This sound does not require a stoppage in airflow so it can be used as a vowel (e.g. /b/ in “fit”) or as a consonant (e.g. /l/ in “rapt”). As a result, it is especially suited to the physiological difficulties of articulating consonants with vocals (note again that nearly all of the other consonants are exhaled). Because it can be articulated with a continuous air flow, the /i/ can be used to slightly alter vowels. Specifically, when a vowel is rhoticized, i.e. coloured by an /r/, the third formant becomes lowered.

Even if this third-formant lowering is not as directly tied to an impression of heaviness as the lowering of the first two formants, it nevertheless provides a way for the vocalist to create variety and, on a social-perceptual note with regards to paralanguage, it seems more than a coincidence that the /i/ sounds so often emulate the snarls of wild beasts. Having drawn a number of inferences as to why certain patterns appeared in the improvisation, the strongest and most basic point here is that there exists a consistency to the vocalists’ use of particular phonemes in such a way that they seem fundamental to the most salient musical features of the improvisation.

4 An online appendix that includes larger images such as Figure A1 can be accessed at http://www.music.mcgill.ca/~depalle/ICMC2012/ICMC2 012.instr.htm.

5 One of the clearest introductions to paradigmatic analysis available can be found in [14].
the regions to be filtered were defined according to a computer-based analysis of the original sound file’s specifics. Specifically, AudioSculpt’s “Partial Tracking” feature was used to track partials using multiple breakpoint functions for each sinusoidal component (a procedure that works with inharmonic signals as well as harmonic ones). A new sound was then synthesized by amplifying the formant regions of the original signal.

AudioSculpt’s diapason tool was then used to take frequency readings from the synthesized partials. When the tool is pointed at a particular place on the sonogram, its frequency and amplitude are displayed, a corresponding sine tone appears, and a two-dimensional slice of the synthesized signal appears. This procedure allows for an analysis of the original sound by hypothesizing that certain components of it appear to be important, resynthesizing those components, and closely analyzing them.

4. A RECORDED IMPROVISATION

In order to investigate the features of extreme metal voices presented in section 2 in a more musical context, we asked a vocalist to perform a vocal improvisation using whatever extreme metal vocal techniques he wished. Given the somewhat artificial performance room and the vocal’s obvious lack of reverberation and there were no instruments to accompany the vocalist—the volunteer vocalist’s improvisation shown in Figure A1 (see URL), if not an exact indicator of the performance decisions a vocalist might make in a concert setting or full-band recording, can be considered an accurate reflection of the kinds of performance choices that are possible using inhaled and exhaled voices and the variations in formant frequency available with different vowel combinations.

4.1. A Spectrographically-Informed Music Analysis

In addition to the spectrogram, Figure A1 also contains phonemic and rhythmic transcriptions as well as analytical annotations provided at the bottom of the figure. As indicated by the rhythms given below the spectrogram, the vocals fit neatly into a regular 4/4 meter, indicating that the vocalist kept a regular pulse in the recording, the vocals fit neatly into a regular 4/4 above. Partly because of this rhythmic regularity, the strengths and weaknesses of his earlier performances discussed in Section 3 are marked contrast to his earlier performances shown in Figure A2 (see URL 1), if not an exact indicator of the performance decisions a vocalist might make in a concert setting or full-band recording, can be considered an accurate reflection of the kinds of performance choices that are possible using inhaled and exhaled voices and the variations in formant frequency available with different vowel combinations.

4.1.1. An Impression of Tight Control

This impression likely results from several musical features exhibited by the improvisation that are frequently invoked in discussions of basic musical composition: the improvisation shows a clear binary division into two roughly equal sections of music, separated by a clear change in register (see the dynamic markings at measure 7); each half contains a recurring, slightly varying rhythm (labelled x and y) that generally does not occur in the other half; and, following the contours visible in the spectrogram, both divisions exhibit an arch-like quality of intensity whereby the vocalist’s screams steadily increase in volume and high spectral frequencies (measures 2–4, 9–11) before rapidly calming with quieter, lower vowels (measures 6–8, 12–15). The total result is a sample of improvised extreme metal vocals which demonstrates controlled musical techniques in a regular manner; one that could be heard as loosely narrative or rhetorical in the sense that it creates regularly spaced climaxes, moments of repose, and gradual variations. If there is a clear sense of musicality to be found here, what can be inferred about the vocal techniques used to achieve it?

4.1.2. Inhaled vs. Exhaled Vocals

Inhaled vocals in Figure A1 are shown by boxes around the notation below the spectrogram, indicating that exhaled and inhaled vocals are employed about equally during the improvisation. Though aurally distinguishable from each other, exhaled vocals are harsher in their timbre than inhaled vocals, the distinguishing acoustic characteristics of the inhaled vocals are not always immediately apparent from the spectrogram (at least at the resolution given in the example). An extra tool can be used to make a vocal distinction; namely, a clear difference in the highest regions of spectral energy reached in measures 4–5 and 10–11. None of the exhaled portions of the improvisation come close to this region; inhaled vocals at measures 5, 9, 11, 12 reach both of these segments in the right. Inhaled and exhaled vocals are indicated by different shades of grey horizontal boxes but not in the case of reappearing segments. Consequently, the chart can be followed in real time along with the recording by reading only the shaded segments, proceeding downwards from the top-left to the bottom-right corners. To draw attention to areas of greatest interest, solid boxes outline regions that show both similar rhythms and noticeably repeated phonemes while dashed boxes indicate rhythmic identities with little phonemic similarity.

4.2. Some Rhythmic and Phonetic Motives

The left-column brings into relief how the vocalist created a series of variations during the first half of the improvisation (measures 1–6). Here, he has clearly focused on a particular rhythm that he’s using to imitate the snarls of wild beasts. Having a number of inferences as to why certain patterns appeared in the improvisation, the strength here is that there exists a consistency to the vocalist’s use of particular phonemes in such a way that they seem fundamental to the most salient musical features of the improvisation.

1 An online appendix that includes larger images such as Figure A1 can be accessed at: http://www.musuc.mcgill.ca/~deppel/ICMC2012/ICMC2 012majalek.htm.

2 The alveolar ridge is the sloping region located by the upper jaw’s front teeth. One’s tongue quickly touches this plain to stop the flow of air through the vocal tract when producing the consonant /d/ [15].

3 Paralanguage can roughly be understood to include all forms of non-verbal communication. These can include facial expressions, vocal intonation, body movements, gesticulations, voice tone, and sighs, as well as prosaic and timbral modifications to ordinary speech that shade meaning [16].
5. AN EXCERPT FROM “THE VOWEL SONG”

Framed as a public service promoting literacy, “The Vowel Song” by death metal band Zimmer’s Hole begins with vocalist Chris Valagao reciting the vowels of the alphabet (henceforth “letters”) to distinguish from phonetic vowels) in long sustained screams, harmonized by three guitars in homorhythm (see Figure A3). The slow punctuations of each homorhythmic attack, combining voice, low power chords (shown only as roots in the example), and two harmonized lead guitars, not only lend a certain satirical grandiosity to song, they also help to create the sensation of Valagao’s unchilled screams possessing a kind of melody, drawn by a precise control of formant frequency locations.

5.1.1. Formants in Flux

Although it may not be immediately evident in the spectrograms given here, there is quite a great deal of variation to the formant frequencies used in the example. In order to illustrate this variation, Figure 3 plots the position of each letter that Valagao screams within vowel space, i.e. a graph which plots vowels according to the frequency of the first formant on the y-axis and second formant on the x-axis. Of course, some of the letters Valagao screams are actually diphthongs or triphthongs. Accordingly the most steady-state vowel within each letter is identified with a dot on the graph and arrows leading up to or away from it depending on how the phonetic transitions occur. To illustrate an example, Valagao’s screamed letter “u,” represented by plot #5, is performed in such a way that it traverses vowel space beginning near /i/ (as in “hit”), reaching its most steady-state point near /u/ (as in “harm”), and finally moving towards the lower formant frequencies in between /i/ and /u/ (as in “hook” and “caught” respectively). It becomes clear how much formants in flux taking into consideration that there is usually a frequency range of around 1000Hz over which each formant’s energy is significant (the values in Table 1 sample the average frequency for the first formant) and that the voice is in rhythmic unison with the guitars, it does not seem far-fetched for a listener to perceptually connect the guitar melody and the formant movements, thereby imagining a kind of melodic motion assigned to the last letter by screaming the words “and” before a vowel is stabilized and sustained.

5.1.2. Interactions between the Voice and Guitar

If these changes in formant frequency are compared with the pitch contour of the guitar parts (which move in parallel motion), a surprising correspondence appears between the guitar’s changes of pitch direction and the movements of the first formant. As the first letter changes to the next, the lower formant decreases in frequency paralleling the descent of the guitar (see the “changes of direction” arrows in Figure A3). With the next letter, the lower formant reverses direction just as the guitar does. This pattern of alternating upwards and downwards directions, shared between the guitars and the voice’s first formant, continues until the guitars and voice break the homorhythm. What’s more, there is an even stronger correspondence between the highest lead guitar part and the voice’s first formant movements. This is shown in Table 1 (see next page) which compares the high lead guitar melody with the first formant of each vowel at its point of greatest stability and sustain (the dotted points in Figure 3). Both are shown as frequency values in Hz and in terms of pitches that correspond to those frequencies. A comparison of these pitch values (including their octave position) reveals a striking relationship. With a margin of about one semitone above and below the upper guitar part, the voice’s first formant parallels the exact contour of the guitar part.

5.2. Formants in Flux

In order to illustrate this variation, Figure 3 plots the position of each letter that Valagao screams within vowel space, i.e. a graph which plots vowels according to the frequency of the first formant on the y-axis and second formant on the x-axis. Of course, some of the letters Valagao screams are actually diphthongs or triphthongs. Accordingly the most steady-state vowel within each letter is identified with a dot on the graph and arrows leading up to or away from it depending on how the phonetic transitions occur. To illustrate an example, Valagao’s screamed letter “u,” represented by plot #5, is performed in such a way that it traverses vowel space beginning near /i/ (as in “hit”), reaching its most steady-state point near /u/ (as in “harm”), and finally moving towards the lower formant frequencies in between /i/ and /u/ (as in “hook” and “caught” respectively). It becomes clear how much formants in flux taking into consideration that there is usually a frequency range of around 1000Hz over which each formant’s energy is significant (the values in Table 1 sample the average frequency for the first formant) and that the voice is in rhythmic unison with the guitars, it does not seem far-fetched for a listener to perceptually connect the guitar melody and the formant movements, thereby imagining a kind of melodic motion assigned to the last letter by screaming the words “and” before a vowel is stabilized and sustained.

5.2.1. Interactions between the Voice and Guitar

If these changes in formant frequency are compared with the pitch contour of the guitar parts (which move in parallel motion), a surprising correspondence appears between the guitar’s changes of pitch direction and the movements of the first formant. As the first letter changes to the next, the lower formant decreases in frequency paralleling the descent of the guitar (see the “changes of direction” arrows in Figure A3). With the next letter, the lower formant reverses direction just as the guitar does. This pattern of alternating upwards and downwards directions, shared between the guitars and the voice’s first formant, continues until the guitars and voice break the homorhythm. What’s more, there is an even stronger correspondence between the highest lead guitar part and the voice’s first formant movements. This is shown in Table 1 (see next page) which compares the high lead guitar melody with the first formant of each vowel at its point of greatest stability and sustain (the dotted points in Figure 3). Both are shown as frequency values in Hz and in terms of pitches that correspond to those frequencies. A comparison of these pitch values (including their octave position) reveals a striking relationship. With a margin of about one semitone above and below the upper guitar part, the voice’s first formant parallels the exact contour of the guitar part.

5.2.2. Interactions between the Voice and Guitar

If these changes in formant frequency are compared with the pitch contour of the guitar parts (which move in parallel motion), a surprising correspondence appears between the guitar’s changes of pitch direction and the movements of the first formant. As the first letter changes to the next, the lower formant decreases in frequency paralleling the descent of the guitar (see the “changes of direction” arrows in Figure A3). With the next letter, the lower formant reverses direction just as the guitar does. This pattern of alternating upwards and downwards directions, shared between the guitars and the voice’s first formant, continues until the guitars and voice break the homorhythm. What’s more, there is an even stronger correspondence between the highest lead guitar part and the voice’s first formant movements. This is shown in Table 1 (see next page) which compares the high lead guitar melody with the first formant of each vowel at its point of greatest stability and sustain (the dotted points in Figure 3). Both are shown as frequency values in Hz and in terms of pitches that correspond to those frequencies. A comparison of these pitch values (including their octave position) reveals a striking relationship. With a margin of about one semitone above and below the upper guitar part, the voice’s first formant parallels the exact contour of the guitar part.

6. CONCLUSION

Having observed extreme metal vocal techniques in both a relatively controlled recording session and at work in a commercial studio recording, it should now be clear that the extreme metal voice is far from the simplistic percussive device that it is often assumed to be. If such

<table>
<thead>
<tr>
<th>No. 1</th>
<th>No. 2</th>
<th>No. 3</th>
<th>No. 4</th>
<th>No. 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>C6/1050 Hz</td>
<td>G5/780 Hz</td>
<td>Eb 6/1240 Hz</td>
<td>C6/1050 Hz</td>
<td>Eb 6/1240 Hz</td>
</tr>
<tr>
<td>C5/540 Hz</td>
<td>G4/380 Hz</td>
<td>F5/700 Hz</td>
<td>B4/510 Hz</td>
<td>E5/650 Hz</td>
</tr>
</tbody>
</table>

Table 1. A comparison of frequency values between the highest lead guitar’s melody and the first vocal formant in its point of greatest stability.

REFERENCES


interdisciplinary approach to the extreme metal voice, the results of this paper support ongoing arguments for the musicological utility of spectrograms in drawing attention to subtle means of musical expression that can easily be overlooked.
5. AN EXCERPT FROM “THE VOWEL SONG”

Framed as a public service promoting literacy, “The Vowel Song” by death metal band Zimmer’s Hole begins with vocalist Chris Valagao reciting the vowels (henceforth “letters”) to distinguish from phonetic vowels in long sustained screams, harmonized by three guitars in homorhythm (see Figure A3). The slow punctuations of each homorhythmic attack, combining voice, low power chords (shown only as roots in the example), and two harmonized lead guitars, not only lend a certain satirical grandiosity to song, they also help to create the sensation of Valagao’s unpitched screams possessing a kind of melody, drawn by a precise control of formant frequency locations.

5.1.1. Formants in Flux

Although it may not be immediately evident in the spectrograms given here, there is quite a great deal of variation to the formant frequencies used in the example. In order to illustrate this variation, Figure 3 plots the position of each letter that Valagao screams within vowel space, i.e. a graph which plots vocals according to the frequency of the first formant on the y-axis and second formant on the x-axis. Of course, of the letters Valagao screams are actually diphthongs or triphthongs. Accordingly the most steady-state vowel within each letter is identified with a dot on the graph and arrows leading up to or away from it depending on how the phonetic transitions occur. To illustrate an example, Valagao’s screamed letter “u,” represented by plot #5, is performed in such a way that it traverses vowel space beginning near /i/ (as in "hi"), reaching its most steady-state point near /u/ (as in "harm"), and finally moving towards the lower formant frequencies in between /u/ and /o/ (as in "hook" and "caught" respectively). It becomes clear how much formant variation is involved in this song introduction from only the steady-state plot points, let alone taking into account all the quick diphthong transitions indicated by the arrows.

5.1.2. Interactions between the Voice and Guitar

If these changes in formant frequency are compared with the pitch contour of the guitar parts (which move in parallel motion), a surprising correspondence appears between the guitars’ changes of pitch direction and the movements of the first formant. As the first letter changes to the next, the lower formant decreases in frequency parallelly the descent of the guitars (see the “changes of direction” arrows in Figure A3). With the next letter, the lower formant reverses direction just as the guitar does. This pattern of alternating upwards and downwards directions, shared between the guitars and the voice’s first formant, continues until the guitars and voice break the homorhythm. What’s more, there is an even stronger correspondence between the highest lead guitar part and the voice’s first formant movements. This is shown in Table 1 (see next page) which compares the high lead guitar melody with the first formant of each vowel at its point of greatest stability and sustain (the dotted points in Figure 3). Both are shown as frequency values in Hz and in terms of pitches that correspond to those frequencies. A comparison of these pitch values (including their octave position) reveals a striking relationship. With a margin of about one semitone above and below the upper guitar part, the voice’s first formant parallels the exact contour of the guitar part.

6. CONCLUSION

Having observed extreme metal vocal techniques in both a relatively controlled recording session and at work in a commercial studio recording, it should now be clear that the extreme metal voice is far from the simplistic percussive device that it is often assumed to be. If such assumptions are in no small part the result of deeply entrenched habits of describing vocal music primarily in terms of pitched melodies, the extreme metal voice can serve as an invitation to approach the study of musical expression in new ways. Having taken an interdisciplinary approach to the extreme metal voice, the results of this paper support ongoing arguments for the musicological utility of spectrograms in drawing attention to subtle means of musical expression that can easily be overlooked.