Listener Expertise Enhances Intelligibility of Vocalizations in Death Metal Music

Kirk N. Olsen, William Forde Thompson, &
Iain Giblin
Macquarie University, Sydney, Australia

Death Metal music with violent themes is characterized by vocalizations with unnaturally low fundamental frequencies and high levels of distortion and roughness. These attributes decrease the signal to noise ratio, rendering linguistic content difficult to understand and leaving the impression of growling, screaming, or other non-linguistic vocalizations associated with aggression and fear. Here, we compared the ability of fans and non-fans of Death Metal to accurately perceive sung words extracted from Death Metal music. We also examined whether music training confers an additional benefit to intelligibility. In a 2 $\times$ 2 between-subjects factorial design (fans/non-fans, musicians/nonmusicians), four groups of participants ($n = 16$ per group) were presented with 24 sung words (one per trial), extracted from the popular American Death Metal band *Cannibal Corpse*. On each trial, participants completed a four-alternative forced-choice word recognition task. Intelligibility (word recognition accuracy) was above chance for all groups and was significantly enhanced for fans (65.88%) relative to non-fans (51.04%). In the fan group, intelligibility between musicians and nonmusicians was statistically similar. In the non-fan group, intelligibility was significantly greater for musicians relative to nonmusicians. Results are discussed in the context of perceptual learning and the benefits of expertise for decoding linguistic information in sub-optimum acoustic conditions.

Received: June 11, 2017, accepted March 3, 2018.

Key words: auditory perception, emotion, expertise, lyrics, perceptual learning

It has been proposed that expert musicians exhibit enhanced auditory skills that are underpinned by specific changes in the brain (Kraus & Chandrasekaran, 2010; Musacchia, Sams, Skoe, & Kraus, 2007). These enhancements include a heightened ability to comprehend speech in the presence of noise (for a review, see Coffey, Mogilever, & Zatorre, 2017). One explanation for this enhancement is that expert musicians are highly sensitive to fine-grained acoustic information that is important not only for music perception, but also for speech perception, thereby permitting enhanced speech intelligibility under noisy conditions. However, it is not clear whether this benefit of expertise extends to the perception of sung lyrics in the presence of music, a complex auditory environment in which a singer's vocal line is embedded within multiple instrumental textures.

For nonmusicians, evidence suggests that recognition of sung words is enhanced when that music is familiar (Russo & Pichora-Fuller, 2008). However, the nature of this benefit is not well understood. One possibility is that the benefit is specific to particular songs and does not extend to unfamiliar songs from a familiar genre of music. A second possibility is that familiarity with a genre of music permits enhanced intelligibility of the lyrical content for all exemplars of that genre. In that case, recognition of sung words should be better for fans than for non-fans of a particular genre of music, even if the specific songs in question are unfamiliar. The present study focused on a genre of music with an extremely "noisy" combination of vocal signal and background instrumentation: Heavy Metal music and its sub-genre of Death Metal. We predicted that intelligibility of lyrical content in Death Metal should be better for fans than for non-fans of the genre, even when the specific musical pieces are unfamiliar to all listeners.

An important motivation for conducting this research is the violent nature of the lyrical content in a significant proportion of Death Metal music, which is considered to be an example of violent media. Death Metal music is characterized by instrumental and vocal timbres with unnaturally low fundamental frequencies and high levels of distortion, roughness, and intensity that are experienced by listeners as loud, aggressive, and highly arousing (Arnal, Flinker, Kleinschmidt, Giraud, & Poeppel, 2015; Berger & Fales, 2005; Tsai et al., 2010; Walser, 2014). Animalistic sounds normally associated with aggression and fear, such as growls and screams, are also commonplace in Death Metal vocalizations (Arnal et al., 2015; Tsai et al., 2010). Both vocal and instrumental

Music Perception, volume 35, issue 5, pp. 527–539, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2018.35.5.527

sounds in Death Metal music have aggressive connotations, but the most explicit representation of violence is contained in lyrical content that often describe acts of rape, murder, decapitation, and suicide. The themes embedded within the lyrics are well known to fans, freely available to view in online community sites, and contribute to the aesthetic appreciation by fans of the genre.

For example, the American Death Metal band *Cannibal Corpse* released multiple albums in the top 40 American Billboard charts (Billboard, 2017) with titles such as "Kill" and "Torture," and song titles such as "Relentless Beating," "Hacksaw Decapitation," "Necropedophile," and "She Was Asking for It." Although Death Metal is a sub-genre of the more mainstream Heavy Metal genre, it still receives a relatively large worldwide following with bands such as *Cannibal Corpse* selling over two million albums (Blabbermouth.net, 2017).

It is surprising, therefore, that violent Death Metal music has received relatively little scientific investigation, especially when considering the possible social and psychological implications of exposure to such extreme violent and anti-social themes. Indeed, only a small body of research has addressed how exposure to violent music might influence short- and long-term psychological and behavioral outcomes, or the mechanisms underlying such outcomes (e.g., Anderson, Carnagey, & Eubanks, 2003; Arnett, 1991; Gowensmith & Bloom, 1997; Mast & McAndrew, 2011; Sharman & Dingle, 2015; Warburton, Roberts, & Christenson, 2014). The present study was designed to shed light on the early stages of this process; that is, perception and intelligibility of violent antisocial vocalizations in Death Metal music. Indeed, it is the words carried within Death Metal vocalizations—the lyrics—that are the most direct, explicit channel with which violent themes are conveyed to listeners.

Any potential consequences of violent music in society are likely to apply to those who listen to that music on a regular basis, such as those who identify as fans. To date, research on violent music has examined its impact on individuals without consideration of whether they are fans of that music. Recent evidence suggests that fans and non-fans have very different experiences of Death Metal music (Thompson, Geeves, & Olsen, 2018). In particular, differences in sensitivity to the lyrical content is likely to be one of the most significant ways in which fans and non-fans of Death Metal differ. However, it is not yet clear whether fans of Death Metal music accurately perceive words and messages conveyed in the characteristically noisy and violent Death Metal vocal timbre. Nevertheless, it is likely that fans of Death Metal music have become attuned to the way in which words and messages are conveyed through the characteristic vocalization style. As a result, "genre-specific expert listeners" (i.e., fans) of Death Metal music may show an intelligibility advantage similar to what would be expected from listeners who normally exhibit enhanced auditory skills, such as those with many years of formal music training (Kraus & Chandrasekaran, 2010). Expert musicians show a range of enhanced auditory skills in domains outside of music, such as greater intelligibility of speech in noise (Coffey et al., 2017), and it's likely that these skills will also transfer to perception and intelligibility of noisy Death Metal vocalizations.

Three empirical questions were posed in this investigation: First, to what extent can listeners understand the lyrical content of Death Metal music? Second, is intelligibility of violent Death Metal vocalizations greater for fans of Death Metal music, relative to non-fans who have little-to-no exposure to the distorted vocalizations common in Death Metal music? Third, is intelligibility of violent Death Metal vocalizations greater for musicians with many years of music training than for individuals with little-to-no training in music? Thus, the present study considers two forms of expertise: informal genre-specific listening expertise and formal music training. As reviewed below, existing theory and evidence are consistent with the hypothesis that both forms of expertise may be associated with listeners' sensitivity to the linguistic content of Death Metal lyrics.

MECHANISMS UNDERLYING SPEECH INTELLIGIBILITY

Intelligibility is defined as the ability to accurately perceive and understand speech. For intelligibility to occur, the perceptual system must transform incoming acoustic signals into a form that allows information in the signal to interface with one's mental lexicon. That is, the listener must succeed in mapping sound to meaning. Linguistic research has demonstrated that there are multiple levels of representation in mapping sound to meaning (Hickok & Poeppel, 2007). For example, distinctive features are the basic inventory of sounds that characterize human languages and are thought to be the smallest constituents of speech that have an acoustic interpretation. Collections of distinctive features are bundled together to create segments/phonemes, and the segments/phonemes are then organized into language-specific syllable structure. The featural, segmental/phonemic, and syllabic levels constitute the pre-lexical phonological representation, and these levels of representation provide the initial analysis of the acoustic signal for subsequent intelligibility (Liberman & Mattingly, 1989; Price, Thierry, & Griffiths, 2005).

The smallest constituents of language that mediate meaning are morphemes, and these morphemes have an active role in word recognition (Whalen et al., 2006). To accurately perceive speech in ideal listening conditions (e.g., a spoken word in quiet surrounds), listeners must analyze the acoustic signal, construct the pre-lexical representation, and then map the resulting representation to the correct item in the lexicon (Poeppel & Hackl, 2008). Thus, speech intelligibility involves the ability to extract the acoustic cues in a speech signal that encode the identity of the phonological segments in the target language (consonants, vowels, stress, etc.). This can be conceptualized as a bottom-up/pre-lexical process. It is important to note, however, that interpretation of a speech signal can be influenced by higher level cognitive processes. That is, we need not rely solely on the pre-lexical representation to identify words, but can also use cues that are not present in the immediate acoustic signal. These cues might be linguistic (syntactic or semantic cues) or extra-linguistic (such as facial expression, gesture, word frequency, or broader discourse context) (Yorkston, Strand, & Kennedy, 1996). The pre-lexical representations and knowledge-driven inferences that listeners adopt are not mutually exclusive. In speech perception, listeners can use both strategies and maintain intelligibility (Mattys, Davis, Bradlow, & Scott, 2012; Smiljanic & Chandrasekaran, 2013). Therefore, when perceiving speech, listeners will integrate both lexical (top-down) and pre-lexical (bottom-up) information to maximize intelligibility.

SPEECH INTELLIGIBILITY IN THE CONTEXT OF MUSIC

Speech perception develops quickly and can be achieved with very high accuracy in ideal listening conditions. However, intelligibility becomes difficult when words are sung rather than spoken (Collister & Huron, 2008; Condit-Schultz & Huron, 2017). For example, spoken words are recognized up to seven times more accurately than their sung versions when produced by unaccompanied soloists in controlled acoustic conditions (Collister & Huron, 2008). The decline of intelligibility for sung words is mostly due to errors of vowel centralization, where vowels are perceived as more centralised in a sung word relative to a spoken word (e.g., the target word "steel" is heard as "still"). In addition, errors in perception of stopped consonants (e.g., /b/, /t/, /p/, /k/) strongly contribute to low intelligibility of sung words. This is most likely because stopped consonants require a stoppage of air flow in the mouth that is very difficult to achieve while sustaining a sung word.

The genre or style of music in which words are sung is also an important factor for intelligibility (Condit-Schultz & Huron, 2015; Fine & Ginsborg, 2014). For example, sung words in Jazz and Country music are significantly more intelligible than Pop, Rock, and Classical music (e.g., Opera). In the case of Classical music, difficulty in perceiving sung words can arise because classical vocalists are typically trained to concentrate energy in the "singer's formant" (between $\sim$3–4 kHz for female singers and $\sim$2–3 kHz for male singers) (Huron & Condit-Schultz, 2015). This vocal technique functions to maximize the intensity of vocal production, which was especially desirable in historical periods when no amplification was available. However, increased vocal intensity comes at a cost: as the intensity of sung words increases, intelligibility decreases (Huron & Condit-Schultz, 2015; Scotto Di Carlo, 2007). This trade-off is relevant to music characterized by high-intensity vocal styles, such as Rock music and Heavy Metal/Death Metal music in particular. Indeed, the findings of Condit-Schultz and Huron (2015) show that only $\sim$66% of words sung in Rock music were accurately identified. One would expect, therefore, that lyrics produced in more extreme versions of Rock music—such as Heavy Metal and Death Metal—might be more difficult to identify because of the extreme levels of intensity at which vocalizations are delivered. When high intensity is combined with "noisy" acoustic characteristics, intelligibility should decrease further.

Consider acoustic *roughness*—a primary determinant of vocal quality and a key attribute in the so-called "hoarse" or "croaky" voice (Eddins, Kopf, & Shrivastav, 2015; Fastl & Zwicker, 2007). Typically, roughness results from relatively quick changes in an acoustic signal produced by amplitude modulated frequencies in the range of 30–150 Hz, with a peak at 60 Hz (Arnal et al., 2015; Eddins et al., 2015; Fastl & Zwicker, 2007). Roughness in speech is commonly measured by calculating the harmonics-to-noise ratio of an acoustic signal (i.e., the dominance of periodic over aperiodic levels), as well as the magnitude of acoustic jitter and shimmer (defined as cycle-to-cycle variations and perturbations of fundamental frequency and amplitude, respectively) (Fastl & Zwicker, 2007). High levels of jitter and shimmer are common acoustic features in pathological speech (e.g., a defective laryngeal mechanism), where intelligible vocalizations are difficult for speakers to produce (Kreiman & Gerratt, 2005; Wendahl, 1966). When auditory roughness in Death Metal vocalizations is measured from acoustic analysis of harmonics-to-noise ratio, jitter, and shimmer, all evidence points towards a poor quality speech signal. When compared to the spectral features of normally sung speech, the growls and screams in Death Metal vocalizations are significantly noisier, with little-to-no harmonic structure and extremely low values of harmonics-to-noise ratio (Kato & Ito, 2013; Tsai et al., 2010). Furthermore, these vocalizations are characterized by high levels of jitter and shimmer (Kato & Ito, 2013).

In the natural environment, vocalizations with high levels of roughness and noise-like qualities may signal an auditory warning designed to elicit rapid sound-source localization and subsequent adaptive behavior (e.g., avoid or retreat) (Arnal et al., 2015; McCarthy & Olsen, 2017; Tsai et al., 2010). However, this biological function comes at the expense of linguistic intelligibility. Thus, the intelligibility of sung lyrics in Death Metal vocalizations should be relatively low. Nevertheless, we expect that intelligibility should still be achieved at better-than-chance accuracy because some of the elements necessary for vocal communication are retained in the signal, albeit in a degraded form.

THE ROLE OF EXPERTISE: EXPERT LISTENERS AND EXPERT MUSICIANS

Given that intelligibility of Death Metal vocalizations is predicted to be relatively low, we now turn to the question of whether listener expertise may be associated with the intelligibility of sung lyrics. Two forms of expertise are investigated here, operationally defined by their mode of knowledge acquisition. The first is the implicit acquisition of genre-specific musical knowledge that results from being a "fan" or "expert listener" of a specific musical genre. The second is the acquisition of general musical expertise through active and formal instrumental training that leads to expert musicianship.

Genre-specific expert listeners (fans). A "fan" is broadly defined as a passionate devotee of a particular type of consumptive object (Hunt, Bristol, & Bashaw, 1999). As such, fans are highly motivated to frequently engage in passionate behaviors related to that object (e.g., attend every game of their favorite football team or listen repetitively to their favorite musical group). Such passion can function in a positive and harmonious way in one's life, or become obsessive to the point of maladaptive behavior (Vallerand et al., 2003; Vallerand et al., 2008). In the context of music, fans of a particular genre will have engaged with music in a way that increases exposure, experience, and familiarity with genre-specific characteristics that demarcate one genre from another (e.g., characteristic instrumental timbres, lyrical themes, vocal styles, and so on). Through the mechanism of *perceptual learning*, this genre-specific expertise should be associated with enhanced intelligibility of noisy and degraded vocalizations (Gibson, 1963, 1971; Goldstone, 1998; Watanabe & Sasaki, 2015).

The concept of perceptual learning is built on the premise that perception is shaped by one's knowledge and past experience (Samuel & Kraljic, 2009). With repeated exposure to a certain type of stimulus, perceivers develop an increased ability to discriminate specific elements of that stimulus; elements that may not be perceptible by those without equivalent experience-based knowledge. Changes to stimulus processing that result from perceptual learning are long lasting and are underpinned by four key mechanisms: (1) *attentional weighting*—increased attention to important unitary elements of a stimulus; (2) *stimulus imprinting*—detectors (receptors) developed to specialise in processing particular elements of a stimulus; (3) *differentiation*—the ability to separate important elements from the whole; and (4) *unitization*—the ability to collect important elements to combine into a functional whole (Goldstone, 1998). In the context of speech perception, perceptual learning enhances intelligibility of speech sounds that "deviate" from the norm (Norris, McQueen, & Cutler, 2003; Samuel & Kraljic, 2009). For example, intelligibility of degraded speech such as speech with artificial compression or vocoding is enhanced by perceptual learning, and this enhancement generalizes across multiple speakers (Kraljic & Samuel, 2006; Samuel & Kraljic, 2009).

It is likely that Death Metal fans are genre-specific expert listeners who have undergone a process of perceptual learning through repeated exposure to specific characteristics of the genre. As a result, fans will have implicitly learned to extract the elements of Death Metal vocalizations necessary for relatively accurate perception of degraded speech (perhaps initially in combination with the aid of written material, such as songbooks). If perceptual learning does indeed underpin this process, then enhanced intelligibility should generalize to any Death Metal vocalizations that have acoustic profiles that are typical of the genre, even if those vocalizations are unfamiliar. To test this prediction, the present study ensured that all stimuli were unfamiliar to fans and non-fans of Death Metal music.

Expert musicians and enhanced auditory skills. The second form of expertise under investigation is the acquisition of knowledge derived from formal music training. In general, musicians who have received many years of formal music training show enhanced auditory skills associated with long-term changes in the brain (Kraus & Chandrasekaran, 2010; Musacchia et al., 2007). Some of these enhanced skills may have pre-existed the training, given that individuals with excellent auditory skills are more likely to persist with music lessons than individuals with poor auditory skills. However, it is also likely that the process of training itself enhances the ability to perceive fine-grained acoustic structures used to convey meaning in music, such as pitch, timing, and timbre. These acoustic structures are also important for communicating meaning in speech (Kraus & Chandrasekaran, 2010). Indeed, training in music may lead to enhanced sensitivity to speech prosody (Thompson, Schellenberg, & Husain, 2004) and enhanced ability to discriminate small acoustic fluctuations in speech (Parbery-Clark, Tierney, Strait, & Kraus, 2012). Musicians also outperform nonmusicians on speech-in-noise intelligibility tasks (Coffey et al., 2017). In the context of the present study, we expect that a "musician's advantage" for speech-in-noise intelligibility will transfer to enhanced intelligibility of noisy speech. If this indeed the case, then intelligibility of noisy degraded speech found in Death Metal vocalizations should be greater for musicians than nonmusicians. This advantage should be most evident for non-fans of Death Metal music, given that fans may already demonstrate enhanced intelligibility processing from their genre-specific listening expertise.

Method

PARTICIPANTS

Sixty-four Australian participants took part in the investigation (26 females and 38 males; $M = 23.50$ years, $SD = 6.43$, range = 18–41 years). There were four groups of equal sample size ($n = 16$): fans and non-fans of violent music who are either trained or untrained in music. Non-fans were recruited from Macquarie University's pool of first-year psychology students and completed the task for course credit. Participants who self-identified as fans of violent music were recruited from: (1) Macquarie University's pool of first-year psychology students; (2) a snowball sampling technique targeting students currently attending private music colleges in Sydney; and (3) self-selection by responding to advertisements on social media Death Metal fan pages. The inclusion criterion for participants in the 'musician' group was $\geq$ 5 years of instrumental music training, and for the 'nonmusician' group, < 5 years of music training. Although the nonmusician group did include participants with some music training, the demographic data reported in Table 1 clearly show that two distinct groups were achieved: musicians reported an average of 10.38 years of music training, whereas nonmusicians reported an average of 0.97 years of training. All participants reported normal hearing.

STIMULI

Stimuli comprised 24 words sung in a Death Metal vocal timbre from the song *Hammer Smashed Face* by the band *Cannibal Corpse*. The original vocal track of the song with no instrumental backing (i.e., the vocal stem) was exported from the publicly available video game software *Rock Band* (produced by Harmonix) into the music editing software *Audacity* (Version 2.1.2). Individual words from the vocal stem were segmented in *Audacity*, resulting in 24 words sung in a Death Metal timbre but without musical accompaniment (see online supplemental material for sound files of all 24 items used in the study). A number of factors informed the choice of the final 24 items used as target words in the study. First, we restricted our choice of sung words to those for which it was possible to determine word boundaries. Second, we ensured that our sample of words included both context-congruent words expected in the Death Metal genre, such as "Kill," "Die," "Pulverize," and "Suffer," in addition to context-neutral words, such as "Me," "I," "Out," and "It" (see Table 2 for complete list of context-congruent and context-neutral words). In choosing the foils that accompanied each target word in a trial, we ensured they were matched, where possible, on the number of syllables, onset phoneme category, coda, and rhyming characteristics.

TABLE 1. *Demographic Information*

Experimental Group		n	Male | Female	Group Means (SD): Age	Group Means (SD): Years of Training
Fans					
	Musicians	16	13 | 3	24.69 (7.41)	12.56 (4.76)
	Nonmusicians	16	9 | 7	25.88 (6.74)	1.25 (1.34)
Non-Fans					
	Musicians	16	5 | 11	21.88 (6.01)	8.19 (2.81)
	Nonmusicians	16	11 | 5	21.56 (4.77)	0.63 (1.36)

Note. "Years of Training" refers to years of formal instrumental music training.

TABLE 2. *List of Context-Congruent and Context-Neutral Stimuli*

Context-Congruent	Context-Neutral
Cold	Being
Coming	Bidding
Cortex	Contents
Cranial	Continues
Crushing	I
Die	Inside
Feel	It's
Kill	Like
Killing	Me
Pulverized	Out
Suffer	Something
Tissue	You

Note. The term "context" refers to Death Metal music

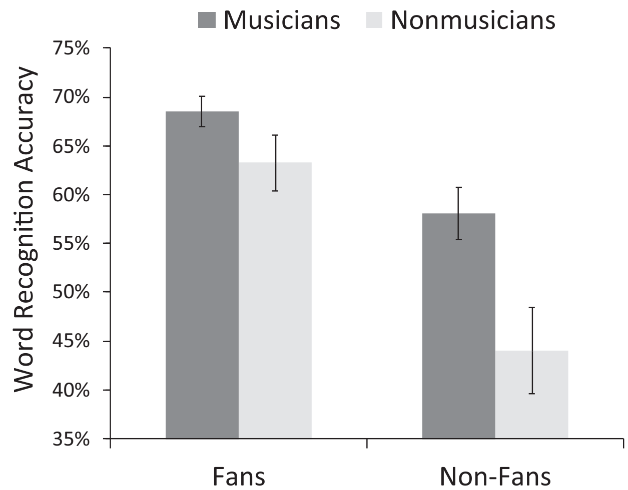

FIGURE 1. Group mean word recognition accuracy for fans and non-fans of violent death-metal music, reported for the musician and nonmusician groups. Error bars report standard error of the mean.

PROCEDURE

Participants completed the study through the online platform Qualtrics. After consenting to participate, participants completed an online demographic questionnaire about age, gender, Death Metal fan status and music training, followed by the word recognition intelligibility task. This task comprised 24 trials using a four-alternative forced-choice protocol. In each trial, participants heard a randomly selected word item. After each item was played, participants were instructed to "choose one of four options from the list that best represents the sung word in the audio example" (see Appendix A for a list of trials, including the four response options in each trial). Of the four options, one option was correct (randomly placed in one of the four possible positions) and three were incorrect (foils). Two practice trials were presented to familiarize participants with the task. After the listening phase of the study, participants were asked to name the song and band from which the stimuli were derived. None of the participants correctly identified the song or band, confirming that stimuli were unfamiliar to participants.

Results

A 2 × 2 between-subjects ANOVA was conducted to investigate the effects of genre-specific listening expertise (fans and non-fans of violent Death Metal music) and general musical expertise (musicians and nonmusicians) on intelligibility of sung Death Metal lyrics. As can be seen in Figure 1, word recognition accuracy was better than chance (> 25%) for all four participant groups, with a grand mean of 58.46% accuracy ($SD = 15.09$).

However, differences between groups were observed. First, there was a significant main effect of fan group, $F(1, 60) = 23.41$, $p < .001$, $\eta^2_p = .28$. As predicted, word recognition accuracy was significantly greater for fans of violent Death Metal music ($M = 65.88\%$, $SD = 9.50$) than for non-fans ($M = 51.04\%$, $SD = 16.09$). Second, there was a significant main effect of musician group, $F(1, 60) = 9.86$, $p < .01$, $\eta^2_p = .14$. As predicted, word recognition accuracy was significantly greater for musicians ($M = 63.27\%$, $SD = 10.18$) than for nonmusicians ($M = 53.64\%$, $SD = 17.63$). There was no significant fan group × musician group interaction, $F(1, 60) = 2.07$, $p > .05$, $\eta^2_p = .03$. We next conducted planned contrasts to compare intelligibility between musicians and nonmusicians for fans and non-fans separately. As shown in Figure 1, word recognition accuracy was significantly greater for musicians relative to nonmusicians in the non-fan group, $t(30) = 2.71$, $p < .05$, 95% CI [.03, .25], but not in the fan group, $t(30) = 1.59$, $p > .05$, 95% CI [−.01, .12].[1] Due to the relatively large variability in years of music training, we also investigated whether years of music training correlated with word recognition accuracy within the fan group and the non-fan group. Results show a significant correlation within the non-fan group, $r(32) = .34$, $p = .05$, but not within the fan group, $r(32) = .16$, $p > .05$. Thus, among non-fans the significant enhancement of word recognition accuracy observed for musicians relative to nonmusicians was corroborated by a significant positive correlation between years of music training and word recognition accuracy. Conversely, among fans, the lack

[1] These results were also observed when using the more conservative Bonferroni adjusted alpha of .025 to control for inflated Type I error due to multiple comparisons.

TABLE 3. *Word Recognition Accuracy for Individual Items: Fans, Non-Fans, and Both Groups Combined*

Fans ($n = 32$)		Non-Fans ($n = 32$)		Combined ($n = 64$)	
Word	Accuracy	Word	Accuracy	Word	Accuracy
Continues	96.88%	Coming	90.63%	Coming	89.06%
You	96.88%	Suffer	87.50%	Suffer	89.06%
Suffer	90.63%	Continues	78.13%	Continues	87.50%
Being	87.50%	Crushing	78.13%	You	85.94%
Coming	87.50%	You	75.00%	Being	78.13%
Kill	87.50%	Being	68.75%	Crushing	78.13%
Inside	84.38%	Cortex	68.75%	Cortex	73.44%
Feel	81.25%	Die	68.75%	Die	73.44%
Cortex	78.13%	Something	65.63%	Inside	71.88%
Crushing	78.13%	Feel	59.38%	Feel	70.31%
Die	78.13%	Inside	59.38%	Kill	70.31%
Killing	78.13%	Killing	56.25%	Something	70.31%
Something	75.00%	Kill	53.13%	Killing	67.19%
It‘s	71.88%	Like	46.88%	It‘s	56.25%
Cold	62.50%	Cold	40.63%	Cold	51.56%
Contents	53.13%	It‘s	40.63%	Like	42.19%
Tissue	50.00%	Pulverized	40.63%	Pulverized	42.19%
Bidding	43.75%	Tissue	31.25%	Contents	40.63%
Pulverized	43.75%	Contents	28.13%	Tissue	40.63%
I	40.63%	I	28.13%	I	34.38%
Me	40.63%	Me	25.00%	Me	32.81%
Like	37.50%	Out	18.75%	Bidding	28.13%
Out	21.88%	Bidding	12.50%	Out	20.31%
Cranial	15.63%	Cranial	3.13%	Cranial	9.38%

of a reliable difference in word recognition accuracy between musicians and nonmusicians was echoed by a non-significant correlation between years of music training and word recognition accuracy.

Next, recognition accuracy was calculated for each individual word item to evaluate whether words congruent to violent Death Metal music were recognized by fans and non-fans more accurately than words with no obvious association with the genre. These data help determine whether participants tend to respond to the "most likely" option when unsure of the correct option (e.g., choosing a violent word instead of a neutral word when unsure, because Death Metal vocalizations are generally aggressive). Table 3 presents fans' and non-fans' group mean word recognition accuracy for each individual target word, with the most accurately perceived item situated at the top of each list. As shown in Table 3, participants did not tend to choose the most context-congruent word in each trial. If participants had selected the "most likely" option, then recognition accuracy for words such as "Kill," "Killing," "Pulverized," and "Cranial" should be situated at the top of each list in Table 3. This was not the case.

To investigate whether differences in target word accuracy were due to differences in their relative frequency in the English language, we calculated word frequencies of all 24 target words using the MCWord Orthographic Wordform Database (Medler & Binder, 2005), a corpus of approximately 16,600,000 written examples and 1,300,000 spoken examples. If word recognition accuracy in the present study were biased on target words' frequency of occurrence in the English language, then we would expect to see a positive correlation between word frequency and word accuracy. This was not the case. There was a small and non-significant negative correlation between target word frequency and target word accuracy for the fan group, $r(22) = -.21$, $p > .05$, and also the non-fan group, $r(22) = -.18$, $p > .05$. To investigate the possible influence of word frequency on recognition accuracy across all words in the study, regardless of whether they were a target or a foil, we also correlated word frequency data for all targets and foils with the number of times each word was chosen by fans and non-fans across the entire study. There was a small and non-significant positive correlation between word frequency and the number of times each word was chosen by the fan group, $r(82) = .11$, $p > .05$, and by the non-fan group, $r(82) = .06$, $p > .05$.[2] See

[2] From the full list of 96 words used in the study (24 targets and 72 foils), word frequency data for 12 foils were not available from the MCWord database. Therefore, the correlation between word frequency data (targets and foils) and the number of times each word was chosen by participants was calculated on 84 words from the full list of 96.

Appendix A for the MCWord database word frequency data of all items in the study.

Finally, owing to the finite number of words in the song from which stimuli were derived, it was not possible to match a priori the context-congruent and context-neutral words (see Table 2) on various dimensions such as length or frequency of occurrence of each word's orthographic neighbor(s). Nevertheless, we analyzed whether recognition accuracy of context-congruent words and context-neutral words in Table 2 correlates with the frequency of occurrence in the English language of each word's orthographic neighbor(s). If this issue were relevant to our findings, then we would expect that words with high-frequency orthographic neighbors should be more readily accessed from memory, and thus recognized more accurately, than words with low-frequency orthographic neighbors. On the other hand, it is also possible that words with high-frequency orthographic neighbors are less distinct and hence more easily confused with other words, resulting in lower word recognition accuracy than words with "rare" or low-frequency orthographic neighbors.

To investigate these competing hypotheses, we calculated each word's orthographic neighborhood size, defined as a set of words that can be formed by changing one letter while preserving letter positions (Coltheart's 'N'; see Coltheart, Davelaar, Jonasson, & Besner, 1977), and then averaged the frequency of occurrence of each word's orthographic neighbor(s) in the English language, calculated using the MCWord Orthographic Wordform Database (Medler & Binder, 2005). We then conducted two separate correlational analyses comparing: 1) word recognition accuracy of context-congruent words with the frequency of occurrence of each word's orthographic neighbor(s); and 2) word recognition accuracy of context-neutral words with the frequency of occurrence of each word's orthographic neighbor(s). Results show that there was no significant correlation between word recognition accuracy of context-congruent words and the frequency of occurrence of each word's orthographic neighbor(s), $r(12) = .15$, $p > .05$, and word recognition accuracy of context-neutral words and the frequency of occurrence of each corresponding word's orthographic neighbor(s), $r(12) = -.43$, $p > .05$.

Discussion

The results of this investigation indicate that participants who identified as fans of Death Metal music were significantly more accurate in word recognition than those who identified as non-fans. From experience and familiarity with the particular genre of music, fans have likely developed a form of genre-specific listening expertise and perceptual learning from increased exposure to the acoustic features of violent Death Metal vocalizations and the characteristic growl-like timbre (Tsai et al., 2010). Drawing from the mechanisms outlined in Goldstone (1998), *attentional weighting* and *differentiation* are most applicable here. Fans' experience listening to growl-like vocalizations likely increased the effective weighting of attentional resources to features of the acoustic signal that are optimal for intelligibility. For example, experienced listeners may be able to fine-tune their attention to acoustic cues that encode the identity of phonological segments (consonants, vowels, stress, etc.) (Poeppel & Hackl, 2008). As a result, fans may have better access to a *semantic* level of processing.

On the other hand, non-fans with little-to-no experience-based perceptual learning are likely restricted to a more *affective* level of processing. Linguistic information in Death Metal vocalizations is carried in an aggressive growl-like timbre. In the natural environment, growls function to communicate aggression and elicit fear (Arnal et al., 2015; Tsai et al., 2010). Non-fans with little experience listening to growling vocalizations may attend closely to the aggressive and aversive nature of the signal at the expense of other details of the signal, analogous to the phenomenon of "weapon focus" in visual attention (Loftus, 1979), and may experience high arousal, negative affect, and feelings of tension, fear, and anger in response to such vocalizations (Thompson et al., 2018). This perceptual bias and emotional response may distract attention away from the linguistic cues associated with an underlying verbal message. Conversely, through repeated exposure to this genre of music, fans may have become desensitized to the inherently aggressive nature of growling vocalizations, allowing them to allocate more attention to linguistic cues. Such a possibility is compatible with research suggesting that chronic exposure to violent video games may also result in desensitization to aggression and violence (e.g., Bartholow, Bushman, & Sestir, 2006; Carnagey, Anderson, & Bushman, 2007; Gentile, Swing, Anderson, Rinker, & Thomas, 2016; Ivarsson, Anderson, Åkerstedt, & Lindblad, 2013). Quite possibly, an optimal level of desensitisation to the aggressive high-arousal/negative-valenced characteristics of growling vocalizations constitutes an important underlying factor in the appreciation of Death Metal music and its lyrics.

Along similar lines, it is possible that non-fans allocated more attentional resources to the threatening affective qualities of growling vocalizations (*task-irrelevant* information), rather than to *task-relevant* information such as the phonological segments necessary for accurate encoding. For non-fans, growling vocalizations may function as a high-arousal/negative-valenced "emotional distractor" that has the potential to elicit poor task-relevant attention and reduced cognitive-perceptual performance (Sussman, Heller, Miller, & Mohanty, 2013). This conjecture awaits further investigation.

Another possible explanation of the fan expertise effect is that adept listeners with experience listening to Death Metal may become accustomed to rule-governed manipulations of the vocal tract that transform the natural sound of language into characteristic Death Metal vocalizations. Specifically, the acoustic characteristics of Death Metal vocalizations are exaggerations of the latent possibilities of the human vocal tract, and unmarked vocalizations can be transformed in various ways to achieve the unnaturally low frequencies and high levels of distortion and intensity. For example, to produce the characteristic "heavy" style of death metal vocalizations, the larynx can be lowered, the head tilted forward, and the lips rounded in order to extend the length of the vocal tract; a process that mimics the physiological properties of a large beast (Smialek, Depalle, & Brackett, 2012). Indeed, humans tend to perceive an increase in sound-source body size as the vocal tract lengthens (Fitch, 1994). When the vocal tract is lengthened, the acoustic signal undergoes a systematic rule-governed transformation from unmarked human speech into highly marked Death Metal vocalizations. The rule-governed nature of this process means that experienced listeners such as fans may gain expertise by learning the phonological transformation that takes an underlying linguistic representation and changes this representation into the "accent" of Death Metal. Future research could investigate this hypothesis by conducting spectral analyses of growling vocalizations to validate the existence of phonological transformations.

To increase the generalizability of this research, it would be valuable in future studies to examine whether other forms of genre-specific listener expertise have a comparable benefit for intelligibility of sung vocalizations in that genre. Classical singing offers one possible comparison. Classical singing comprises vocalizations that are not growled, yet sometimes still include low fundamental frequencies (e.g., a male bass voice) and elicit relatively low listener intelligibility ($\sim 48\%$ correct recogntion of Classical singing in Condit-Schultz & Huron, 2015; the grand mean recognition accuracy in the present study was $\sim 58\%$ correct). Would a similar fan effect to that reported here be observed for intelligibility of Classical male bass vocalizations? Furthermore, to what extent can the benefits of genre-specific expertise extend beyond intelligibility for sung words in that genre to sung words in another genre? For example, do fans of violent music have an intelligibility advantage for sung vocalizations in music genres outside of Death Metal, or is this expertise effect restricted to sung Death metal vocalizations? These questions await future investigation.

BENEFITS OF MUSIC TRAINING FOR WORD INTELLIGIBILITY

In addition to the benefit of genre-specific listener expertise, a similar intelligibility benefit was observed for the more general yet formalized acquisition of musical expertise that characterizes expert musicians. Here, musicians with a mean of 10.38 years of instrumental music training were significantly more accurate in recognizing words than nonmusicians, who on average reported less than one year of formal music training. Research consistently reports enhanced auditory skills in musicians, although the frequent reliance on correlational evidence leaves questions of causation uncertain. For example, musicians have an enhanced ability over nonmusicians to perceive fine-grained acoustic structures such as pitch, timing, and timbre—important cues for communicating meaning in music and speech (Kraus & Chandrasekaran, 2010). Musicians can also discriminate small acoustic fluctuations (Parbery-Clark et al., 2012) and outperform nonmusicians when speech intelligibility is measured in the presence of background noise (Coffey et al., 2017). The present findings contribute to this body of research by illustrating that the enhanced auditory skills observed in musicians include heightened intelligibility of *noisy speech*, rather than just speech in noise. Whether such a musician's advantage transfers to other forms of noisy speech, such as foreign-accented speech, awaits further investigation.

As predicted, the intelligibility benefit for musicians relative to nonmusicians was strongest in the non-fan group of participants. For fans of Death Metal music, there was no significant difference in intelligibility between musicians and nonmusicians. To explain this finding in more detail, it is possible that the relatively high intelligibility results observed from the fan/musician group reflects a maximal level of accuracy, or a kind of perceptual ceiling that is defined by what is perceptually possible given the noisy acoustic features of the stimulus set. Mean word recognition accuracy in the fan/musician group was 68.48% ($SD = 6.27$). The next most accurate group was the fan/nonmusician group, with mean word recognition at 63.27% ($SD = 11.51$). The difference between the fan/musician group and the fan/nonmusician group was not statistically significant, with an increase of only 5.21% accuracy for the musicians. Thus, the musician's advantage for the fan group was small, most likely because intelligibility observed from the genre-specific expert listeners (fans) had already reached a close-to-maximal level of accuracy. On the other hand, mean word recognition accuracy in the non-fan/musician group was 58.06% (SD = 10.82), significantly higher than mean word recognition accuracy in the non-fan/nonmusician group (M = 44.02%, SD = 17.68). Thus, the musician's advantage in the non-fan group was relatively large, with an increase of 14.04% accuracy for the musicians. If a ~69% word recognition accuracy reflects a maximal point of accuracy for the most expert of perceivers in the present study (fan/musician group), then the ~44% word accuracy of the non-fan/nonmusician group reflects the abilities of people in the general population who neither listen to Death Metal music or have received a substantial degree of music training. With such a relatively poor level of intelligibility, it is not surprising that a common anecdotal description of Heavy Metal/Death Metal music is that of an aversive and unintelligible "wall of noise."

RESPONSE BIAS AND WORD FREQUENCY

If participant responses were biased toward the "most likely" option in a trial, such as context-congruent words related to violence or antisocial themes, then word recognition accuracy for context-congruent words in Table 3 would be expected to be situated towards the top of each word list with highest accuracy. However, this was not the case. Word recognition accuracy for context-congruent words and aggressive words in particular were reasonably well dispersed across the accuracy range of each group's response list in Table 3. This finding provides evidence that participants did not just choose the most likely option in a context-congruent trial.

Two limitations of the study are notable. First, the entire list of 24 target words and their 72 foils in the four-alternative forced-choice procedure were not explicitly matched for word frequency in the English language. Such an endeavor would be near impossible to complete with systematic accuracy, given the constraints associated with choosing context-congruent and context-neutral target words, and then matching their foils on the number of syllables and onset, coda, and rhyming characteristics. Nevertheless, the item analysis provided strong evidence that differences in word recognition accuracy in the present study could not be explained by differences in their relative frequency in the English language. Second, the trial-by-trial listening conditions in the present study do not represent real-world listening experience. For example, words are usually sung in the context of surrounding words or sentences, rather than in isolation. By design, this contextual information was withdrawn from listeners to control for the potential influence of learned expectations based on surrounding linguistic context of a given word. Therefore, future research should not only investigate word intelligibility in the context of a full vocal line in which it is embedded, but also in the context of full musical accompaniment.

CONCLUSIONS

A number of researchers have suggested that exposure to violent media, including violent music, may lead to aggressive cognitions and anti-social behaviors. The present study was designed to shed light on this issue by investigating intelligibility of growl-like vocalizations in violent Death Metal music. The findings show that word recognition of vocalizations in Death Metal music is significantly greater in two groups of listeners with different kinds of expertise: genre-specific listening expertise that results from being a fan and thus an experienced listener of the genre, and a more general kind of musical knowledge that is associated with expert musicians. Fans of Death Metal music understand the lyrics significantly more accurately than non-fans; musicians understand the lyrics of Death Metal vocalizations significantly more accurately than nonmusicians. However, the "musician's advantage" was only observed for those who identified as a non-fan of Death Metal. For fans, intelligibility was relatively good, regardless of musicianship. Not surprisingly, participants with little-to-no music training or Death Metal experience—arguably representing a majority of the general public—were very poor at the word recognition task.

It is still an open question whether violent words and messages in Death Metal music (and Heavy Metal more broadly) *cause* increased levels of aggressive cognitions and behavior. Similar to studies investigating desensitisation to aggression from exposure to violent video games, it is possible that exposure to violent music leads to desensitization of the inherently aggressive and aversive elements of the growl-like vocal timbre. Future studies should investigate whether there are causal connections between short- and long-term exposure to violent music and desensitization to violence more broadly, as well as the social, psychological, and neurological mechanisms that underpin causal connections along the pathway between perception, cognition, and action.

Author Note

This research was supported by an Australian Research Council Discovery Project grant (DP160101470) held by the second author. We thank Dr. Sachiko Kinoshita for advice on computing word frequencies and members of the Macquarie University Music, Sound, and Performance Research Group for helpful comments on an earlier draft.

Correspondence concerning this article should be addressed to Kirk N. Olsen, Department of Psychology, Macquarie University, New South Wales, 2109, Sydney, Australia. E-mail: kirk.olsen@mq.edu.au

References

Anderson, C. A., Carnagey, N. L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology, 84*, 960–971.

Arnal, L. H., Flinker, A., Kleinschmidt, A., Giraud, A.-L., & Poeppel, D. (2015). Human screams occupy a privileged niche in the communication soundscape. *Current Biology, 25*, 2051–2056.

Arnett, J. (1991). Heavy metal music and reckless behavior among adolescents. *Journal of Youth and Adolescence, 20*, 573–592.

Bartholow, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event-related brain potential data. *Journal of Experimental Social Psychology, 42*, 532–539.

Berger, H. M., & Fales, C. (2005). "Heaviness" in the perception of heavy metal guitar timbres: The match of perceptual and acoustic features over time. In P. D. Green & T. Porcello (Eds.), *Wired for sound: Engineering and technologies in sonic cultures* (pp. 181–197). Middletown, CT: Wesleyen University Press.

Billboard. (2017). *Cannibal Corpse*. Retrieved from http://www.billboard.com/artist/298634/cannibal-corpse/chart

Blabbermouth.net. (2017). *Cannibal Corpse awarded with plaque signifying sales of more than two million*. Retrieved from http://www.blabbermouth.net/news/cannibal-corpse-awarded-with-plaque-signifying-sales-of-more-than-two-million/

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real-life violence. *Journal of Experimental Social Psychology, 43*, 489–496.

Coffey, E. B. J., Mogilever, N. B., & Zatorre, R. J. (2017). Speech-in-noise perception in musicians: A review. *Hearing Research, 352*, 49–69.

Collister, L. B., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review, 3*, 109–125.

Coltheart, M., Davelaar, E., Jonasson, T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance IV* (Vol. 79, pp. 345–359). New York: Academic Press.

Condit-Schultz, N., & Huron, D. (2015). Catching the lyrics: Intelligibility of twelve song genres. *Music Perception, 32*, 470–483.

Condit-Schultz, N., & Huron, D. (2017). Word intelligibility in multi-voice singing: The influence of chorus size. *Journal of Voice, 31*, 121/e121–121/e128.

Eddins, D. A., Kopf, L. M., & Shrivastav, R. (2015). The psychophysics of roughness applied to dysphonic voice. *Journal of the Acoustical Society of America, 138*, 3820–3825. DOI: 10.1121/1.4937753

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics: Facts and models* (3rd ed.). New York: Springer.

Fine, P. A., & Ginsborg, J. (2014). Making myself understood: Perceived factors affecting the intelligibility of sung text. *Frontiers in Psychology, 5*, 809.

Fitch, W. T. (1994). *Vocal tract length perception and the evolution of language* (Unpublished doctoral dissertation). Brown University.

Gentile, D. A., Swing, E. L., Anderson, C. A., Rinker, D., & Thomas, K. M. (2016). Differential neural recruitment during violent video game play in violent-and nonviolent-game players. *Psychology of Popular Media Culture, 5*, 39–51.

Gibson, E. J. (1963). Perceptual learning. *Annual Review of Psychology, 14*, 29–56.

Gibson, E. J. (1971). Perceptual learning and the theory of word perception. *Cognitive Psychology, 2*, 351–368.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612.

Gowensmith, W. N., & Bloom, L. J. (1997). The effects of heavy metal music on arousal and anger. *Journal of Music Therapy, 34*, 33–45.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*, 393–402.

Hunt, K. A., Bristol, T., & Bashaw, R. E. (1999). A conceptual approach to classifying sports fans. *Journal of Services Marketing, 13*, 439–452.

Huron, D., & Condit-Schultz, N. (2015). *Catching the words: The science of vocal intelligibility.* Unpublished manuscript.

Ivarsson, M., Anderson, M., Åkerstedt, T., & Lindblad, F. (2013). The effect of violent and nonviolent video games on heart rate variability, sleep, and emotions in adolescents with different violent gaming habits. *Psychosomatic Medicine, 75*, 390–396.

Kato, K., & Ito, A. (2013). Acoustic features and auditory impressions of death growl and screaming voice. In *Proceedings of the Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (pp. 460–463). New York: Institute of Electrical and Electronics Engineers.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review, 13*, 262–268.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience, 11*, 599–605.

Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *Journal of the Acoustical Society of America, 117*, 2201–2211.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science, 243*, 489–494.

Loftus, E. F. (1979). *Eyewitness testimony.* Cambridge, MA: Harvard University Press.

Mast, J. F., & McAndrew, F. T. (2011). Violent lyrics in heavy metal music can increase aggression in males. *North American Journal of Psychology, 13*, 63–64.

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Language and Cognitive Processes, 27*, 953–978.

McCarthy, L., & Olsen, K. N. (2017). A "looming bias" in spatial hearing? Effects of acoustic intensity and spectrum on categorical sound source localization. *Attention, Perception, and Psychophysics, 79*, 352–362.

Medler, D. A., & Binder, J. R. (2005). *MCWord: An on-line orthographic database of the English language*. http://www.neuro.mcw.edu/mcword/

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*, 15894–15898.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*, 204–238.

Parbery-Clark, A., Tierney, A., Strait, D. L., & Kraus, N. (2012). Musicians have fine-tuned neural distinction of speech syllables. *Neuroscience, 219*, 111–119.

Poeppel, D., & Hackl, M. (2008). The functional architecture of speech perception. In J. R. Pomerantz (Ed.), *Topics in Integrative Neuroscience: From Cells to Cognition* (pp. 154–180). Cambridge, UK: Cambridge University Press.

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: Where is it? *Trends in Cognitive Sciences, 9*, 271–276.

Russo, F. A., & Pichora-Fuller, M. K. (2008). Tune in or tune out: Age-related differences in listening to speech in music. *Ear and Hearing, 29*, 746–760.

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, and Psychophysics, 71*, 1207–1218.

Scotto Di Carlo, N. (2007). Language and diction: Effect of multifactorial constraints on intelligibility of Opera singing (I). *Journal of Singing-The Official Journal of the National Association of Teachers of Singing, 63*, 443–455.

Sharman, L., & Dingle, G. A. (2015). Extreme metal music and anger processing. *Frontiers in Human Neuroscience, 9*, 272.

Smialek, E., Depalle, P., & Brackett, D. (2012). A spectrographic analysis of vocal techniques in extreme metal for musicological analysis. In *Proceedings of the 36th International Computer Music Conference*. Ljubljana, Slovenia: International Computer Music Association.

Smiljanic, R., & Chandrasekaran, B. (2013). *Processing speech of varying intelligibility.* Paper presented at the Meeting of the International Congress on Acoustics. Montreal, Canada: Acoustical Society of America.

Sussman, T. J., Heller, W., Miller, G. A., & Mohanty, A. (2013). Emotional distractors can enhance attention. *Psychological Science, 24*, 2322–2328. DOI: 10.1177/0956797613492774

Thompson, W. F., Geeves, A. M., & Olsen, K. N. (2018). Who enjoys listening to violent music and why? *Psychology of Popular Media Culture*. Advanced online publication. DOI: 10.1037/ppm0000184

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46–64.

Tsai, C.-G., Wang, L.-C., Wang, S.-F., Shau, Y.-W., Hsiao, T.-Y., & Auhagen, W. (2010). Aggressiveness of the growl-like timbre: Acoustic characteristics, musical implications, and biomechanical mechanisms. *Music Perception, 27*, 209–222.

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., et al. (2003). Les passions de l'ame: On obsessive and harmonious passion. *Journal of Personality and Social Psychology, 85*, 756–767.

Vallerand, R. J., Ntoumanis, N., Philippe, F. L., Lavigne, G. L., Carbonneau, N., Bonneville, A., et al. (2008). On passion and sports fans: A look at football. *Journal of Sports Sciences, 26*, 1279–1293.

Walser, R. (2014). *Running with the devil: Power, gender, and madness in heavy metal music* (2nd ed.). Middletown, CT: Wesleyan University Press.

Warburton, W. A., Roberts, D. F., & Christenson, P. G. (2014). The effects of violent and antisocial music on children and adolescents. In D. Gentile (Ed.), *Media violence and children: A complete guide for parents and professionals* (pp. 301-328). Santa Barbara, CA: Praeger.

Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. *Annual Review of Psychology, 66*, 197–221.

Wendahl, R. (1966). Some parameters of auditory roughness. *Folia Phoniatrica et Logopaedica, 18*, 26–32.

Whalen, D., Benson, R. R., Richardson, M., Swainson, B., Clark, V. P., Lai, S., et al. (2006). Differentiation of speech and nonspeech processing within primary auditory cortex. *Journal of the Acoustical Society of America, 119*, 575–581.

Yorkston, K. M., Strand, E. A., & Kennedy, M. R. (1996). Comprehensibility of dysarthric speech: Implications for assessment and treatment planning. *American Journal of Speech-Language Pathology, 5*, 55–66.

Appendix A

LIST OF TRIALS WITH WORD FREQUENCIES FROM THE MCWORD DATABASE

Experiment Trials	Word Frequency (per Million)	Experiment Trials	Word Frequency (per Million)	Experiment Trials	Word Frequency (per Million)
Q1		Q9		Q17	
a) Slumping	0.357	a) Gold	91.619	a) Dry	92.987
b) Something	645.258	b) Coal	37.837	b) Die	81.148
c) Stunting	0.297	c) Toad	3.332	c) Dime	3.927
d) Summing	NA	d) Cold	186.212	d) Day	772.038
Q2		Q10		Q18	
a) Implied	13.088	a) Peeing	0.476	a) Continue	86.146
b) Inside	189.187	b) Teeing	0.297	b) Continues	26.831
c) In stride	NA	c) Being	850.746	c) Continued	123.150
d) Incite	0.714	d) Meaning	71.451	d) Contusion	0.119
Q3		Q11		Q19	
a) Me	1884.020	a) Context	38.789	a) Kill	80.315
b) Be	6137.930	b) Cortex	1.725	b) Pill	13.802
c) We	3317.080	c) Contest	11.780	c) Chill	11.601
d) Knee	31.234	d) Complex	67.286	d) Quill	1.547
Q4		Q12		Q20	
a) Oh	302.759	a) Bidding	3.748	a) Paralysed	NA
b) Ah	55.745	b) Brimming	1.428	b) Polarized	1.785
c) Oo	NA	c) Billing	1.309	c) Pulverized	0.892
d) Out	2521.240	d) Digging	13.386	d) Penalized	2.082
Q5		Q13		Q21	
a) Oi	NA	a) Crushing	5.414	a) Issue	90.905
b) Bi	NA	b) Crunching	1.666	b) Miss You	NA
c) I	20145.900	c) Crashing	6.723	c) Diss' You	NA
d) Wry	2.677	d) Crusting	NA	d) Tissue	10.173
Q6		Q14		Q22	
a) Feel	357.432	a) Cranial	0.476	a) Crumbing	NA
b) Fear	164.022	b) Cradle	5.414	b) Drumming	2.737
c) Fetal	NA	c) Kill	80.315	c) Gumming	0.119
d) Fail	37.124	d) Cable	11.661	d) Coming	212.151
Q7		Q15		Q23	
a) Kidding	6.485	a) Context	38.789	a) Hike	2.142
b) Kicking	13.743	b) Condense	0.952	b) Like	1884.020
c) Chilling	5.235	c) Contents	21.774	c) Bike	8.269
d) Killing	35.577	d) Contest	11.780	d) Reich	3.570
Q8		Q16.		Q24	
a) Mew	0.357	a) Snuffer	0.059	a) Bits	30.163
b) View	205.369	b) Stuffer	NA	b) Hits	7.437
c) You	6334.670	c) Shudder	5.176	c) Its	1563.350
d) Knew	473.086	d) Suffer	41.169	d) Is	9055.390

Note. In the study, the order of trials and the order of response options within each trial were randomized for each participant.