Original Article

Trends in Hearing
Volume 27: 1–19
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23312165221141142
journals.sagepub.com/home/tia
SAGE

Musical Emotion Categorization with Vocoders of Varying Temporal and Spectral Content

Eleanor E. Harding[1,2,3], Etienne Gaudrain[1,4], Imke J. Hrycyk[1,2], Robert L. Harris[1,3], Barbara Tillmann[4], Bert Maat[1,2,5], Rolien H. Free[1,2,5] and Deniz Başkent[1,2]

Abstract

While previous research investigating music emotion perception of cochlear implant (CI) users observed that temporal cues informing tempo largely convey emotional arousal (relaxing/stimulating), it remains unclear how other properties of the temporal content may contribute to the transmission of arousal features. Moreover, while detailed spectral information related to pitch and harmony in music — often not well perceived by CI users— reportedly conveys emotional valence (positive, negative), it remains unclear how the quality of spectral content contributes to valence perception. Therefore, the current study used vocoders to vary temporal and spectral content of music and tested music emotion categorization (joy, fear, serenity, sadness) in 23 normal-hearing participants. Vocoders were varied with two carriers (sinewave or noise; primarily modulating temporal information), and two filter orders (low or high; primarily modulating spectral information). Results indicated that emotion categorization was above-chance in vocoded excerpts but poorer than in a non-vocoded control condition. Among vocoded conditions, better temporal content (sinewave carriers) improved emotion categorization with a large effect while better spectral content (high filter order) improved it with a small effect. Arousal features were comparably transmitted in non-vocoded and vocoded conditions, indicating that lower temporal content successfully conveyed emotional arousal. Valence feature transmission steeply declined in vocoded conditions, revealing that valence perception was difficult for both lower and higher spectral content. The reliance on arousal information for emotion categorization of vocoded music suggests that efforts to refine temporal cues in the CI user signal may immediately benefit their music emotion perception.

Keywords
music emotion perception, vocoders, arousal, valence, cochlear implants

Received 14 January 2022; Revised 27 September 2022; accepted 11 October 2022

Introduction

A cochlear implant (CI) helps regain hearing function for hard-of-hearing or deaf populations, by replacing (residual) acoustic hearing with electrical hearing (for an introduction to cochlear implants, see Clark, 2004, or Macherey & Carlyon, 2014). While CIs partially restore hearing, the spectro-temporal details of the sounds transmitted are reduced compared to normal hearing (NH) (for a review of relevant factors related to electric stimulation of the nerve, the nerve-electrode interface, and physiological/clinical aspects, see Başkent et al., 2016). Thus while music[1] enjoyment has been reported to contribute to quality of life in CI users (Fuller et al., 2021; Lassaletta et al., 2008), the difficulties in hearing and appreciating music still persist within this

[1]Department of Otorhinolaryngology/Head and Neck Surgery, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands
[2]Graduate School of Medical Sciences, Research School of Behavioural and Cognitive Neurosciences, University of Groningen, Groningen, The Netherlands
[3]Prins Claus Conservatoire, Hanze University of Applied Sciences, Groningen, The Netherlands
[4]Lyon Neuroscience Research Center, CNRS UMR5292, Inserm U1028, Université Lyon 1, Université de Saint-Etienne, Lyon, France
[5]Cochlear Implant Center Northern Netherlands, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Corresponding Author:
Eleanor E. Harding, Department of Otorhinolarynology, University Medical Center Groningen, Hanzeplein 1 9713 GZ, Groningen, The Netherlands.
Email: e.e.harding@rug.nl

population (e.g., Caldwell et al., 2017; Fuller et al., 2019; Gfeller et al., 2000; Lassaletta et al., 2008).

One aspect of music that contributes to listener enjoyment is the emotion it conveys (Garrido & Schubert, 2011; Grewe et al., 2005; Mazaheryazdi et al., 2018). Emotion is conveyed by a range of musical features inherent to the composition, for example, information based on pitch and time dimensions as written in the score, as well as features inherent to the production of the sound by performers, for example, variation in loudness, tempo and micro-timing (Gabrielsson & Lindström, 2001) (though these latter features can be indicated by the composer).

Music Emotion Perception with Normal Hearing

Emotional affect can be described along the planes of valence and arousal (Feldman, 1995). Valence is the positive or negative association, for example joy has positive valence and sadness has negative valence. Arousal is the visceral association, how excited one becomes physiologically, for example fear has high arousal, whereas serenity has low arousal. These four emotions were previously plotted along valence and arousal axes (see Table 1 in Materials; e.g., Bigand et al., 2005), and classical music excerpts have been consistently classified into these four emotion categories (Filipic et al., 2010; Lévêque et al., 2018).

Table 1. Participant demographic information. For participants who reported to have musical background and where participants listed multiple instruments, the instrument with the most years played is reported. Years of formal training and years active refer to years where one or more instruments were played, though only the primary instrument is reported for each participant in the table.

		Musical background	
Participant ID	Age (years)	Yes/no (self report)	Primary instrument (age onset, years formal training, total years active)
1	25	yes	Piano (7, 13, 17)
2	63	no	
3	26	yes	Bass guitar (12, 6, 14)
4	38	no	Clarinet (15, 4, 4)
5	22	yes	Guitar (12, 2, 3)
6	25	yes	Violin (4, 14, 14)
7	51	no	
9	29	no	
10	29	yes	Drums (8, 1, 1)
11	28	no	
12	49	no	
13	23	no	
14	26	yes	Guitar (8, 13, 17)
15	23	yes	Keyboard (16, 0, 7)
16	22	yes	Guitar (17, 0.3, 1)
17	21	no	
18	32	yes	Violin (6, 11, 11)
19	23	no	
20	25	no	
21	36	no	
22	29	yes	Guitar (11, 0, 18)
23	22	no	

The positive or negative valence of an emotion in music is often conveyed by pitch relationships, for example: 'happy' and 'sad' emotions are typically associated with the major third and minor third pitch intervals prevalent in major and minor mode, respectively (Nieminen et al., 2012), and ironic, angry or melancholic music often makes use of pitch relationships from e.g., pentatonic interval relationships of the Dorian mode (e.g., rock music; Temperley & de Clercq, 2013). Pitch relationships that form simple integer ratios (e.g., the same note one octave apart) are moreover perceived as consonant and those that form complex relationships (e.g., chromatically sequential notes) are perceived as dissonant (Berg & Stork, 2004). The dissonance is also less saliently reflected in roughness in the acoustic envelope (Tramo et al., 2001); consonance and dissonance are globally perceived as exhibiting positive and negative valence, respectively, reflected in distinct activations in emotion networks in the brain (i.e., the amygdala; Koelsch et al., 2005). The degree of emotional arousal conveyed by music, on the other hand, is consistently linked to perceived tempo and loudness. When normal-hearing listeners continuously rated emotion at frequent fixed points throughout classical music excerpts, acoustic analysis of the excerpts revealed that sudden loudness (measured as weighted intensity) had immediate and most pronounced influence on higher arousal ratings, followed by an increased tempo (Schubert, 2004). Consistent with this, faster and louder music is consistently rated as more arousing than slow and soft music (Ilie & Thompson, 2006).

Music Emotion Perception with CI Hearing

Compared to normal hearing, CI hearing has reduced spectral and temporal resolution of sound, as well as a smaller dynamic range, which affects the perception of pitch, envelope and loudness of music. In the implant, frequency-specific information is primarily transmitted in the form of temporal modulations of electric pulse trains sent to different electrodes, mimicking the tonotopic organization of the healthy ear (Loizou, 1998). A typical modern implant has 12 to 22 electrodes (Macherey & Carlyon, 2014). One of the main factors that typically limits the potential benefit from increasing the number of electrodes in the implant is the spread of current, leading to channel interaction (de Balthasar et al., 2003; Shannon, 1983): when the current spreads far around each stimulating electrode, the populations of neurons of the spiral ganglion excited by each electrode become less distinct from each other. This reduces the capacity to discriminate different frequency profiles, i.e. limits the effective spectral resolution of the implant.

Furthermore, a secondary consequence of current spread is that the envelope information carried by each electrode – or channel – becomes effectively contaminated by the surrounding electrodes. Thus fine-grained frequency and temporal information is often lost to the CI listener, which means that their experience of music is qualitatively different from that of a NH listener. Impaired pitch perception might result in impaired perception of interval relationships that normally would inform mode, and in turn emotional valence in music. In a paradigm with short musical bursts improvised by musicians to convey the intended emotion, participant categorization of the stimuli was analyzed according to acoustic features including brightness (spectral energy above 3000 Hz), root mean square energy, roughness, and pitch. NH listeners correctly identified happy, sad, threat and neutral emotions while CI users were only able to identify the happy musical emotion above chance (Paquette et al., 2018) —interestingly, among CI users, spectral information did not inform valence ratings while roughness in the envelope and intensity cues did. Contrastingly, valence ratings of NH listeners exclusively correlated with pitch-related cues. A study investigating the perception of mode and tempo in music found that when tempo was held constant, implant hearing could not utilize mode cues when judging emotional valence (D'Onofrio et al., 2020). Similarly a study with children reported that changes in mode did not affect happy and sad ratings in CI user children whereas NH peers assigned emotion ratings according to differences in mode (Hopyan et al., 2013). Moreover, compared to NH listeners, CI users are also reported to be less sensitive to changes in musical consonance and dissonance (Caldwell et al., 2016), which also convey valence. In that study, dissonance was created by sequentially playing tonally distant chords, and CI users ratings of pleasantness were independent of consonance and dissonance while NH listeners consistently rated progressions with dissonant chords as unpleasant.

Valence has been reported to be perceived by CI-users (Ambert-Dahan et al., 2015). In one study, CI-users categorized musical excerpts into emotional categories happiness, fear, sadness and peacefulness and additionally rated the valence and arousal. Valence was accurately rated. However, the majority of participants possessed residual hearing in their contralateral ear, which has been previously reported to greatly improve music perception (El Fata et al., 2009; Gfeller et al., 2006).

Implants have been developed with the primary clinical aim to perceive speech, and early research reported that the temporal resolution available to implants was sufficient for speech intelligibility (in ideal listening conditions), even with few frequency channels (Shannon et al., 1995), and particularly slow-varying temporal information (Shannon, 1992). This preserved perception of the envelope may have consequences for CI users' music listening. Namely, despite reduced temporal resolution compared to NH listeners, CI user and NH perception of information encoded in the envelope of music may still be comparable — at least performance regarding tempo portions of music perception tests are similar across NH and CI groups (Cooper et al., 2008). With regard to music emotion perception, the consequence would be that emotional arousal cues could still be available to CI users enough to complete music emotion perception tasks, though the CI users' performance is still significantly below that of NH controls (e.g., Caldwell et al., 2015; D'Onofrio et al., 2020; Hopyan et al., 2013). Adult CI users rating classical music excerpts as happy or sad seem to have based their judgements on tempo alone, while NH counterparts seem to have used both tempo and mode cues (Caldwell et al., 2015). Also in CI children, tempo seems to be used almost exclusively to judge whether emotion in classical music excerpts is happy or sad in both original form and when tempo is changed to faster or slower, while their NH peers in contrast seem to use tempo as well as mode to make these emotion judgements (Giannantonio et al., 2015; Hopyan et al., 2013).

CI outcomes for speech perception largely vary across individual CI users (Blamey et al. 2013). This large inter-individual variability suggests that CI users would differ in their abilities in hearing both temporal and spectral properties of CI-transmitted music as well. Two questions emerge with respect to CI users hearing music and perceiving music emotion. On the one hand, it remains unclear whether and to what degree improved spectral resolution would contribute to emotional valence perception in CI users at all, considering that in the literature they seem to rely almost exclusively on arousal features captured by temporal properties of the envelope and dynamic changes. In other words, given the limitations from electric stimulation, the spectral resolution may not easily reach a precision that can transmit valence features. On the other hand, it is not clear how good the temporal resolution of the signal needs to be before emotional arousal features are transmitted to the listener. The above paragraphs outlined that regarding perception of music emotion, temporal cues are typically more available than spectral cues in CI hearing (e.g., Giannantonio et al., 2015; Hopyan et al., 2013), however, it is not clear how dependent the arousal features could be on temporal resolution abilities of CI users, given their large inter-participant variability (Blamey et al., 2013).

Vocoders with Temporal and Spectral Smearing

CI users comprise a heterogeneous population, thus it would prove cumbersome to satisfactorily create groups with systematically matched perception abilities in spectrotemporal resolution to address the current questions. And even if we managed to achieve such matching in performance, we still would not be able to know with certainty that the underlying hearing-related factors causing this functional match would be the same across different implant users. Vocoder technique provides a tool such that the spectrotemporal resolution of the signal can be systematically manipulated to approximate the heterogenous range of acoustic features offered by implant transmission of acoustic features, to be used with NH listeners, such that the hearing-related factors remain homogenous. During vocoding, a signal is reduced to several frequency bands, similar to the number of electrodes in a modern implant (12–22), as well as the range of the number of independent frequency channels (4–16) that CI users seem to be able to make use of (Friesen et al., 2001; Fu & Nogaki, 2005). The envelope is extracted at each frequency band; then the envelope is superimposed on a carrier signal at each frequency band, and the resulting envelope modulated carrier signals are summed across all frequency bands (e.g, Davis et al., 2005). Broadband noise carriers simulate less temporal content from the original signal, as intrinsic fluctuations of the noise result in a non-flat envelope, which means that the envelope transmitted with that carrier will be contaminated by the carrier's own envelope. Broadband noise carriers are moreover very effective at simulating current spread because their spectral shape can easily be adapted with the appropriate filter to simulate current spread in the cochlea (e.g., Bingabr et al., 2008). Sinewave carriers comprise a signal with more temporal information: because the envelope of a sinewave is intrinsically flat, it does not alter the channel's envelope and transmits it faithfully. With each carrier type, the spread of excitation in electrical current can in turn be simulated with the filter order used in bandpass filters that determine the distinct frequency bands. A sharper, higher order filter will simulate less spread of excitation among electrodes, whereas a less steep, lower order filter increases the spectral smearing simulating more spread of excitation. Note that spectral and temporal modifications to the signal are not wholly independent of each other, see 'Methods'.

Vocoders have successfully simulated CI users' music emotion ratings in previous studies with NH participants (Giannantonio et al., 2015; Paquette et al., 2018). Giannantonio et al. (2015) investigated musical emotion categorization using a happy-sad forced choice paradigm. Excerpts were composed as happy (major mode, fast tempo 80–255 bpm), sad (minor mode, slow tempo 20–100 bpm), or ambiguous (same excerpts transposed to the opposite mode or a neutral tempo of 80 bpm) and vocoded with either 22 or 32 channels or noise-masked (pink, white) conditions. It was found that judgments incorporated both tempo and modal cues in the non-vocoded/non-masked condition, while in the CI-simulated conditions, judgments aligned with tempo changes but ignored modal changes. This finding informatively showed the same pattern as CI users, that tempo cues but not modal cues were utilized for music emotion judgments. Paquette et al. (2018) investigated the categorization of happiness, sadness, fear and neutrality in musical stimuli that were short bursts of notes (<2 s), as well as assessing ratings along valence and arousal dimensions. It was found that NH listeners with vocoded stimuli used the timbral features of energy and roughness to inform both arousal and valence ratings. This study used one type of vocoder (8 channels with a noise carrier). Thus while these studies showed that music emotion perception was possible in vocoded conditions, the chosen vocoders did not investigate the degrees of temporal or spectral content available in the signal, and how this might contribute to emotion categorization.

Table 2. Emotion categories from music stimuli inhabit the four quadrants of valence and arousal. In parenthesis, the musical work from which an iconic example of the emotion was played during training. Famous musical works were avoided for actual stimuli in the experimental phase (Bigand et al., 2005).

	Low arousal	High arousal
Positive valence	Serenity (Monoman's *Meditation)*	Joy (Vivaldi's *Four Seasons: Spring*)
Negative valence	Sadness (Beethoven's *Moonlight Sonata*)	Fear (Mussorgsky's *Night on Bald Mountain*)

The present study will add to the literature by systematically manipulating the vocoding parameters along temporal and spectral dimensions, as well as including four emotion categories that allow the assessment of valence and arousal perception in music emotion. In order to assess how varying temporal and spectral resolution content across the heterogeneous range of CI user hearing contributes to music emotion categorization, we asked NH participants to categorize emotions from musical excerpts that cover four corners of the valence-arousal plane: joy, fear, serenity, sadness (see Table 2 in Materials). The musical excerpts were vocoded using conditions that offered greater or less temporal information (sinewave and noise carriers, respectively), and greater or less spread of excitation, or spectral smearing (low vs. high order synthesis filters, respectively). Our hypothesis was that overall emotion categorization would improve as the quality of both temporal and spectral content improved. Moreover improved transmission of arousal features would correspond to improved temporal content, while improved transmission of valence features would correspond to improved spectral content.

Materials and Methods

Participants

Participants were recruited by an advertisement on social media or word-of-mouth. Twenty-four participants responded, but one did not complete the experiment. As a result, 23 (mean age 31.39 years, range 21–63 years, standard deviation 11.90 years; 12 self-reported female, 11 self-reported male) self-reported normal-hearing adults participated in the study. All participants were residing in the Netherlands or Flemish Belgium, and all grew up in countries with a Western Tonal music tradition (Netherlands, United Kingdom, Germany) and could understand Dutch sufficiently to follow experimental directions (self-reported). Demographic information is listed in Table 1. Participants were given the option to self-report whether they had a musical background (yes/no), and if so listed the instrument they played, amount of years of formal training (lessons), and total years played. This information was not factored into the main analysis (see S3. supplementary analysis addressing musicianship differences) but rather demonstrated that participants experienced a wide range of musical backgrounds, thus reflecting varying musical experience occurring in the general population. Participants received 8€ per hour and were tested in accordance with the University Medical Center Groningen medical ethics protocol PICKA-XL (NL66549.042.18).

Materials

Stimuli consisted of 40 classical music excerpts previously determined to represent one of four emotions each: joy, fear, serenity, and sadness (Bigand et al., 2005; Lévêque et al., 2018; Liégeois-Chauvel et al., 2014; a full list of the stimuli material can be found in supplementary table S3 of Pralus et al., 2020 https://doi.org/10.1016/j.cortex.2020.05.015.). While original excerpts were 20 s each in duration (Lévêque et al., 2018; Pralus et al., 2020), feedback from participants during an informal pilot of the online testing version of this experiment indicated that the excerpts were too long to keep attentional vigilance, especially in their vocoded versions. Therefore, we divided each of the 20-s excerpts into two 10-s excerpts presented consecutively. The split halves of each excerpt were always played sequentially, with smooth fade-out and fade-in over 1 s to avoid abrupt onsets and offsets. The categorization of first and second halves was compared in a supplementary analysis (see S4).

The musical excerpts were taken from various recording sources, as a result the relationship between the physical intensity of the excerpts and their intended loudness was not consistent across the excerpts of the set (e.g., an orchestra clearly playing excitedly was the same intensity in the recording as an acoustic guitar solo). While NH listeners can easily deduce loudness cues independently of intensity from spectrotemporal features in the signal (McKenna & Stepp, 2018), degradations introduced by the vocoders may disrupt these cues. As loudness is one of the acoustic features that greatly contribute to the emotional content of music, the loudness of the musical excerpts was adjusted in a separate experiment, described in supplementary materials (S1). The purpose of this loudness manipulation was to make loudness consistent with intended loudness, and hence ensure it is a reliable cue across excerpts despite their various origins. This is particularly relevant for the cases where the physical degradations imposed by vocoding or electric hearing might prevent the perception of acoustic details that would provide the listener information about the intended loudness independently from physical intensity.

Vocoder Processing: Varying Temporal and Spectral Content.

Channels. While earlier research shows voice and speech perception performance in CI users best overlaps with the range of 4–8 channel vocoders in typical implementations (e.g., Friesen et al., 2001; Gaudrain & Başkent, 2018), here we chose 16 channels for the vocoder implementations. This overlaps with the number of electrodes in modern implants, and this choice gives more options for systematic changes in vocoder parameters for manipulating channel interactions (Bingabr et al., 2008). Moreover, 16 channels were used in a previous vocoder study with music materials (short melodies) that manipulated the current spread similarly to the current study (Crew & Galvin, 2012). Each loudness-adjusted excerpt was thus processed with a 16-channel vocoder implemented in Matlab 9.9 (R2020b, The MathWorks, Inc., Natick, Massachusetts, United States). The band partitioning was based on Greenwood's cochlear place-frequency mapping function (Greenwood, 1990) between 150 and 7000 Hz. The analysis filters were 12th order zero-phase Butterworth filters.

Carrier. The carrier was either a **sinewave** or **Gaussian noise**. Sinewave carriers have a flat temporal envelope that matches the envelope of the pulse trains used in actual implants, and makes them well suited to carry information in the form of amplitude modulation. In contrast, noise carriers contain intrinsic modulations that hinder access to the modulation information that the carrier is meant to support. As far as voice pitch perception is concerned, which at least partially relies on amplitude modulation perception in implants (Carlyon, 1997), sinewave and noise carriers seem to encompass, in NH listeners (Gaudrain & Başkent, 2015) the performances observed in CI listeners (Gaudrain & Başkent, 2018). Yet, the use of sinewaves introduces two complications. First, modulating a sinewave results in spectral sidebands, which can potentially be resolved by the normal auditory system. In other words, this simulation potentially introduces fine spectral cues that are not available to CI users. This can be mitigated by making sure all sidebands remain unresolved, that is, by limiting the frequency of modulation to 1/10th of the center frequency (see below). Second, sinewaves are, by definition, restricted to a single frequency. That means that, in NH listeners, they excite a restricted segment of the cochlea. In the case of noise**,** the carrier was a broadband white noise. Broadband noise carriers are very effective at simulating current spread because their spectral shape can easily be adapted to match physical measurements of electrical activation in the cochlea (Bingabr et al., 2008). Simulating current spread with sinewave carriers is thus less straightforward than with noise carriers and requires extra steps in the vocoding process. To this effect, Crew and Galvin (2012) proposed that current spread can still be approximated by smearing the envelopes across channels before they are used to modulate the carriers.

Channel interaction/Spread of excitation. The envelope in each band was extracted using half-wave rectification and an 8th order zero-phase Butterworth low-pass filter with a cutoff frequency that was at least 30 Hz, and at most one tenth of the center frequency of the channel. These values were chosen to limit envelope cues to those that could not result in resolved sidebands with the sinewave carrier. Synthesis was achieved by multiplying the envelope with the carrier. The spread of excitation of each channel informs the quality of the spectral content of the signal, but also determines the amount of channel interaction, which, in turn, also affects the quality of the temporal modulation cues in each channel. In our experiment, the spread corresponded to either a 4th or a 12th order zero-phase Butterworth filter (corresponding to slopes of 24 and 72 dB/oct, respectively). We chose these orders to build from a previous study: Crew and Galvin (2012) modulated channel interaction with 1st, 2nd and 4th order filters. Aside from a condition with no channel interaction, the condition with a 4th order filter was considered to have optimally reduced channel interaction. In their experiment, participants tasked with identifying the melodic contour of vocoded melodies performed significantly better with 4th order compared to 1st or 2nd order filters. In our study, we took the 4th order as our starting point and improved quality of spectral content even further by using a 12th order filter; informal piloting within the lab determined that 12th order filters were still degraded but noticeably provided more spectral information in our musical excerpts.

In the case of the sinewave carrier, the spread of excitation was simulated following the method used by Crew and Galvin (2012). In each channel, the frequency response of the 4th and 12th order was used to add scaled copies of the channel's envelope to the surrounding channels. In the case of the noise carrier, the resulting envelope-modulated noise was filtered in the band with the synthesis filter, which was a 4th or 12th order zero-phase Butterworth filter, before the bands were summed together. The RMS of the vocoded stimulus was adjusted such as to match that of the original stimulus in the 150–7000 Hz frequency band. Spectrograms of selected conditions are shown in Figure 1; see supplementary materials for spectrograms of the remaining conditions (S2).

Procedure

All testing was conducted online using jsPsych 6.1 software (de Leeuw, 2015) on a secure server. Participants were emailed an invitation with general instructions: they were asked to conduct the experiment in a quiet background on a computer, laptop, or tablet with an updated internet browser and stable internet connection. Participants were further recommended to use headphones but told that built-in speakers were acceptable. The email contained a unique link where they first read a consent form and agreed with the terms before proceeding to the experiment. Participants were informed that self-paced breaks would occur approximately every 10 min where they could continue the experiment with a button-press, but to please complete the experiment in one sitting. The experiment lasted approximately 1.5 h determined by informal piloting within the lab and the estimate for online participants was rounded up to 2 h; remuneration was therefore €16 (8€/hr) or, upon incompletion, a flat rate per trial completed.

The experimental task was embedded in a graphically illustrated story (Figure 2) about an alien planet ("The aliens need participants' help to understand emotions in earth's music!").

The task was to listen to the excerpts and to judge which emotion was evoked by the music. Response options given to participants were four buttons with the options joy, fear, serenity and sadness ("vreugde", "angst", "sereniteit", "droefheid" respectively). After task instructions, an 8-trial training with feedback was conducted (e.g., this emotion was "joy") with each vocoder carrier per emotion, but only with the high filter orders. After having categorized the item into the emotion category, participants gave confidence judgments using scale 1–7 (1 defined as little confidence, 7 defined as utmost confidence) with instruction to rate the confidence of their previous emotion categorization. Training items were musical excerpts with iconic emotion status: Joy (Vivaldi's *Four Seasons: Spring*) – Serenity (Monoman's *Meditation*) – Fear (Mussorgsky's *Night on Bald Mountain*) – Sadness (Beethoven's *Moonlight Sonata*) that were not included in the experiment (Table 2). The experiment consisted of three blocks: one for the two carrier conditions (sinewave and noise, 80 trials each) followed by a block for the non-vocoded condition (40 trials). Within each of the vocoded blocks, the filter order randomly alternated between the two selected values (4th and 12th order). Each filter order was assigned to an equal number of trials. Filter order remained the same between first and second halves of full stimuli excerpts. In these blocks, a self-paced break was added after about 40 trials. The order of carrier conditions was counterbalanced across participants. The third block was always non-vocoded, to minimize learning of the experimental stimuli. Before each block, a sample sentence from the Vrije Universiteit sentence corpus (Versfeld et al., 2000) introduced the vocoder carrier to familiarize the participant with a new vocoder before the trial. The sentence was presented visually (in text) and auditorily at the same time, a method that is known to facilitate rapid perceptual learning of degraded speech (Benard & Başkent, 2014; Davis et al., 2005). No responses to the sample vocoder sentences were recorded or analyzed. To reduce testing time and potential testing fatigue, each participant was presented half of the stimuli set. The full stimuli set was counterbalanced across participants. During the experiment there was no control for participants closing the experiment and resuming later; if this happened, the 8-trial training always opened the new session to re-familiarize the participant with the task. It was not possible to return to earlier trials once responses were registered. A progress bar showed participants how much of the experiment they had completed. On average, excluding the time spent on the breaks, the whole experiment lasted about 1h, which corresponds to about 19 s per trial. Correct and incorrect answers per trial were recorded and formed the basis of the raw data confusion matrix.

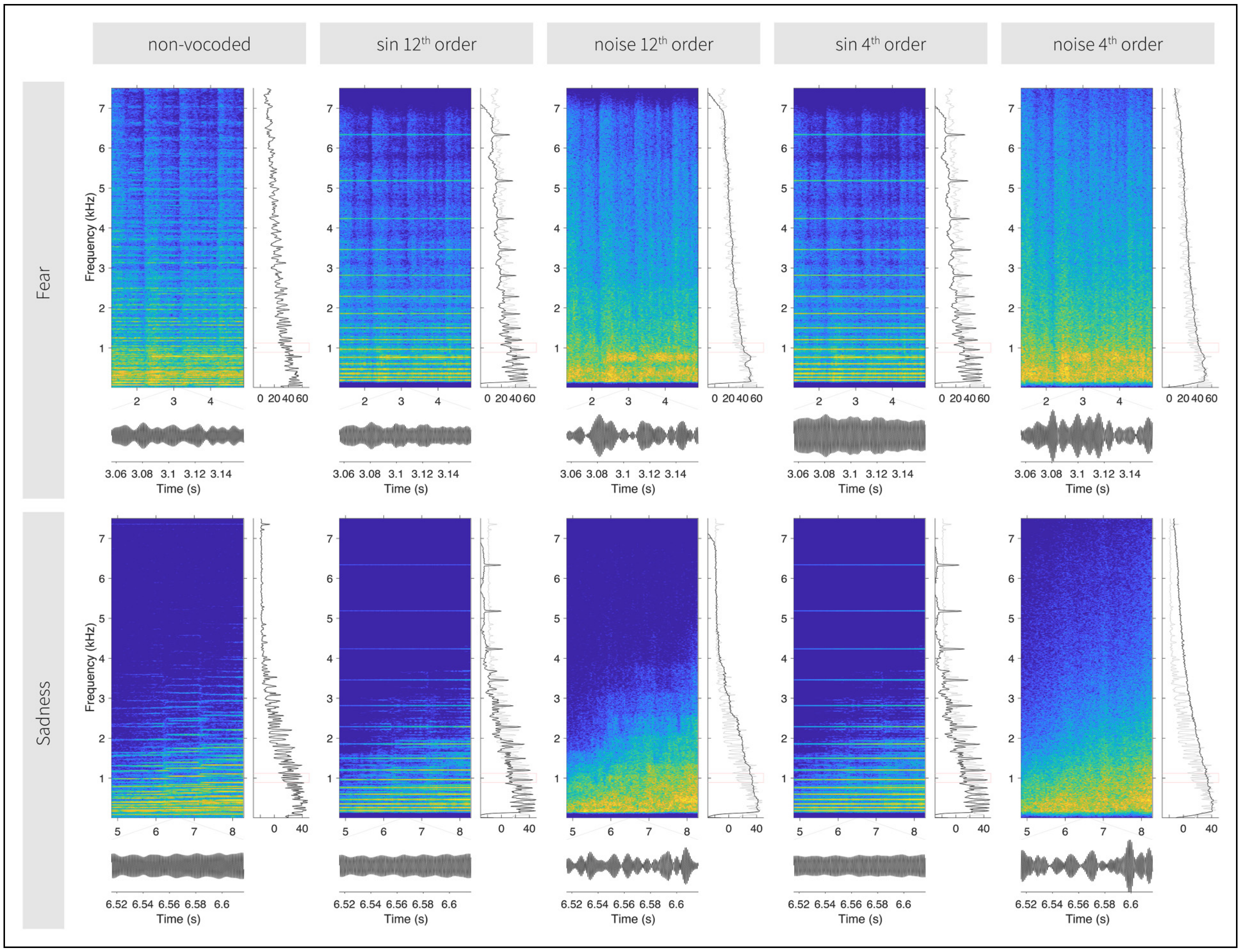

Figure 1. Example stimuli spectrograms shown in non-vocoded and vocoded conditions. Fear, a high-arousal emotion, shows more energy at frequencies across the spectrum in all conditions. Sadness, in comparison, is lower arousal and accordingly displays lower spectral energy across frequencies, especially compared to higher frequencies in the fear condition. Compared to non-vocoded spectrograms, vocoders with sinewave carriers contain energy more localized to channel frequencies whereas in the noise carrier, energy is smeared across multiple frequencies.

Analysis

Scores were first analyzed across vocoder conditions using a Generalized Mixed Effect Model (gLMM) based on a binomial distribution, i.e., using logit as a link function. The analysis was implemented in R (v4.0.3, R Core Team, 2020) using the lme4 package (v1.1.27.1, Bates et al., 2015) The model had binary score (correct or incorrect) as dependent variable, Vocoder (non-vocoded, sinewave 12th order filter, sinewave 4th order filter, noise 12th order filter, noise 4th order filter) as fixed effect, and Participant and Presented Emotion as random factors. Note that the Presented Emotion is the emotion label that was assigned to the musical excerpt by the experimenter during the design and selection of the material by Bigand et al., (2005). An analysis of deviance was performed on the fitted model using the car package (v3.0.11, Fox & Weisberg, 2019). Post-hoc pairwise comparisons were performed using the emmeans package (v1.7.0, Lenth et al. 2021). Multiple comparisons were here corrected with the Tukey method.

Using raw scores, while informative, does not make the sensitivity of the participants apparent. For example, if a participant always answered one emotion, this one category would score 1 (all correct) while not taking into account that other emotion categories also incorrectly received that answer. Therefore, sensitivity, d', from the Signal Detection Theory (Green & Swets, 1988; Macmillan and Douglas Creelman, 2004) was derived from the raw data for each emotion category. For a given emotion, the d' was calculated by considering the correct categorization responses as hits (for a given row of the confusion matrix, the entries term on the diagonal), and incorrect categorization responses, i.e., trials where other emotions were presented but the considered emotion was responded, as false-alarms (the terms off the diagonal on the same row). In order to determine whether sensitivity was above chance-level ($d' = 0$), a 1-sample t-test was performed for each emotion category and each vocoder condition. A comparison of responses to first and second halves of each original excerpt is conducted in supplementary

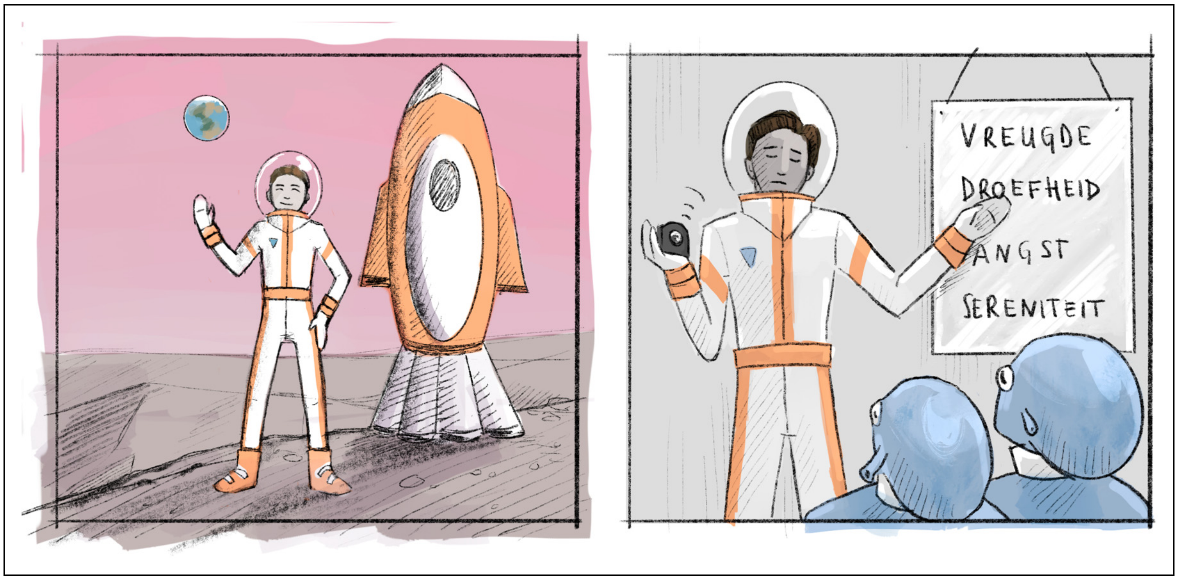

Figure 2. Participants were presented with the scenario that they were an astronaut on an alien planet who was helping aliens to understand earth music. (Illustration by Kristin Hrycyk, image published under the CC BY NC 4.0 license, https://creativecommons.org/licenses/by-nc/4.0/).

analysis S4. The *d'* values were then analyzed with a Repeated Measures ANOVA using the 'ez' package (Lawrence, 2016) in R (R Core Team, 2020). The RM ANOVA had Presented Emotion (joy, fear, serenity, sadness) and Vocoder (non-vocoded, sinewave 12th order filter, sinewave 4th order filter, noise 12th order filter, noise 4th order filter) as repeated factors. A separate RM ANOVA designed to investigate effects of vocoder type on emotion categorization was also conducted, with factors Presented Emotion (joy, fear, serenity, sadness), Carrier (sinewave, noise) and Filter order (Low-4th, High-12th). When the sphericity hypothesis was violated, the degrees of freedom were adjusted using the Greenhouse-Geisser method, and the adjusted p-value was reported as p_{GG}. All ANOVAs used a Type 3 sum of squares. Generalized eta-squared (η^2_G) are reported as effect sizes (Bakeman, 2005). Interactions in the ANOVAs were followed up with paired *t*-tests.

In order to account for the degree that vocoded materials may have rendered participants unsure of their responses, and therefore confound their sensitivity, confidence ratings were collected after each trial, and scaled to the mean and *SD*, of each participant before being correlated with *d'* (Pearsons, 2-tailed). Confidence ratings from both correct and incorrect categorization trials were used. Confidence ratings were further entered into the same analysis as sensitivity in order to assess any differences between the two metrics.

Multiple comparisons (post-hocs and sequential *t*-tests and correlations) were corrected using the false-discovery-rate method (Benjamini & Hochberg, 1995).

The raw data of emotion categorization were also compiled as confusion matrices. In a confusion matrix, a 'perfect score' would be represented as one diagonal line with points only on the same emotion category for both 'presented' (x-axis) and 'responded' (y-axis). Errors are represented by data points off the diagonal, or 'confusions', where a different category was responded from what was presented. The confusion matrices were followed up with an analysis of feature information transmission (Miller & Nicely, 1955). In feature information transmission analysis (FITA), considered features were the valence and arousal classes of each emotion category: each emotion was quantified as having either positive or negative valence, and low or high arousal (see Table 1). The analysis allowed us to estimate how much of the information associated with the feature was effectively received by the listener, which is particularly useful to compare acoustic features in the stimuli (van Wieringen & Wouters, 1999). The outcome measure was the relative quantity of transmitted information in proportion to the total available information (T_{rel},). As this was a continuous variable bound between 0 and 1, it was logit-transformed before being entered in a RM ANOVA with Feature (arousal, valence) and Vocoder (carrier, filter order) as repeated factors.

Results

The proportion correct per vocoder and per presented emotion are reported in Table 3.

The gLMM analysis on raw scores showed a significant effect of the vocoder condition [$\chi^2(4) = 311.7$, $p < 0.0001$]. All pairwise comparisons between vocoder conditions were significant ($p < 0.001$) except the two sinewave vocoders (4th vs. 12th order, $p = 0.71$), the two noise vocoders (4th vs. 12th order, $p = 0.13$), and the sine-4th order compared to the noise-12th condition ($p = 0.05$).

Sensitivity index d'

In order to account for false alarms in the categorization of participants (see ‘Analysis’ above), the sensitivity index (*d'*) was calculated. The sensitivity data are reported in Figure 3. Results with the sensitivity analysis were the same as with the proportion-correct scores.

d' values for each presented emotion and vocoder were all greater than zero, i.e., the performance was better than chance [$t(22)>2.42$, $p_{\mathrm{FDR}}<0.05$], except for serenity in the noise, 4th order condition [$t(22)=1.81$, $p_{\mathrm{FDR}}=0.08$].

The repeated measure ANOVA with Presented Emotion and Vocoder as repeated factors confirmed that both Emotion [$F(3,66)=25.1$, $p<0.0001$, $\eta^2_{\mathrm{G}}=0.12$] and Vocoder [$F(4,88)=59.3$, $p<0.0001$, $\eta^2_{\mathrm{G}}=0.44$] had a significant effect. The Emotion and Vocoder effects did not interact significantly [$F(12,264)=0.69$, $p_{\mathrm{GG}}=0.67$, $\eta^2_{\mathrm{G}}=0.01$].

Emotion: The main effect of Presented Emotion showed that joy (average $d'=1.15$) was the best recognized, followed by fear ($d'=0.98$), then serenity and sadness (both $d'=0.71$).

Non-vocoded vs vocoded: The main effect of Vocoder showed that, again, emotions were best categorized in the non-vocoded condition. The non-vocoded condition yielded the highest sensitivity ($d'=1.74$), but in the vocoded conditions, the performance was relatively higher for sinewave vocoders ($d'=0.89$ for 12th order; $d'=0.78$ for 4th order) than for noise vocoders ($d'=0.60$ for 12th order; $d'=0.42$ for 4th order).

Table 3. Proportion correct per vocoder and per presented emotion.

		Proportion correct	
Vocoder	Emotion	mean	*SD*
non-vocoded	joy	0.87	0.15
non-vocoded	fear	0.80	0.20
non-vocoded	serenity	0.72	0.25
non-vocoded	sadness	0.75	0.19
sin 12th order	joy	0.63	0.25
sin 12th order	fear	0.70	0.25
sin 12th order	serenity	0.44	0.26
sin 12th order	sadness	0.43	0.26
noise 12th order	joy	0.62	0.24
noise 12th order	fear	0.53	0.23
noise 12th order	serenity	0.32	0.21
noise 12th order	sadness	0.36	0.21
sin 4th order	joy	0.62	0.24
sin 4th order	fear	0.58	0.24
sin 4th order	serenity	0.49	0.18
sin 4th order	sadness	0.40	0.24
noise 4th order	joy	0.52	0.27
noise 4th order	fear	0.53	0.23
noise 4th order	serenity	0.30	0.21
noise 4th order	sadness	0.28	0.23

Vocoded carrier type and filter order: In order to examine the effect of carrier type and filter order within vocoded-only conditions, we analyzed only the vocoded data in a separate ANOVA with factors Emotion (joy, fear, serenity, sadness), Carrier (sinewave, noise) and Filter order (low-4th, high-12th). The effect of Emotion was again significant [$F(3,66)=20.74$, $p<0.0001$, $\eta^2_{\mathrm{G}}=0.13$]. In addition, the analysis confirmed that Carrier type had a significant effect on sensitivity [$F(1,22)=28.2$, $p<0.0001$, $\eta^2_{\mathrm{G}}=0.09$] with the sinusoidal carrier resulting in a higher sensitivity ($d'=0.83$) than the noise carrier ($d'=0.51$). Finally, the analysis also showed that filter order had a small, but significant impact on emotion categorization [$F(1,22)=7.17$, $p<0.05$, $\eta^2_{\mathrm{G}}=0.02$], whereby the 12th order filters resulted in larger sensitivity values ($d'=0.74$) compared to 4th order filters ($d'=0.60$). None of these factors interacted with one another (all $p>0.2$).

Confidence Ratings

Confidence ratings were entered into correlations with sensitivity. Combining all non-vocoded and vocoded conditions, correlations ranged between .54 and .64 (all *p*'s < .0001) across all four emotions, indicating that confidence increased with sensitivity with each emotion judgement. Correlations are separated by vocoder in Table 4 (and shown in Figure 4). All correlations were significant except the non-vocoded condition.

Furthermore, confidence ratings were entered into the same ANOVAs as the sensitivity values. The same main effects of Presented Emotion and Vocoder were observed as with sensitivity ANOVA (*F*'s>18.61, *p*'s<.oo1, $\eta^2_{\mathrm{G}}>.140$). An additional Presented Emotion × Vocoder interaction was found [$F(12,264)=3.06$, $p<.001$, $\eta^2_{\mathrm{G}}=.056$], however follow-up separate one-way ANOVAs per vocoder with the factor Emotion (joy, fear, serenity, sadness) revealed that the interaction was driven by the fact that there was no Emotion effect in the non-vocoded condition [$F(3,66)=0.74$, $p=.530$, $\eta^2_{\mathrm{G}}=.019$] while Emotion was significant in all vocoder conditions (*F*'s>7.46, *p*'s<.oo1, η^2_{G} >.160).

Raw Data – Confusion Matrices. The raw data presented in the form of confusion matrices (Figure 5) show that, in vocoded conditions, emotions were systematically confused within arousal type but confused across valence type. In other words, high-arousal joy and fear were most often mistaken for each other, and low-arousal serenity and sadness were most often mistaken for each other, despite the opposing valences. The non-vocoded condition shows relatively good performance as most of the responses can be found on the diagonal. The confusions of the diagonal appeared to grow from sinewave to noise carrier and from 12th to 4th order filter. Moreover, the number of confusions also seemed to depend on the emotions themselves: high arousal emotions (joy and fear) seemed to be better identified than low arousal emotions (serenity and sadness) with systematic confusion of the valence

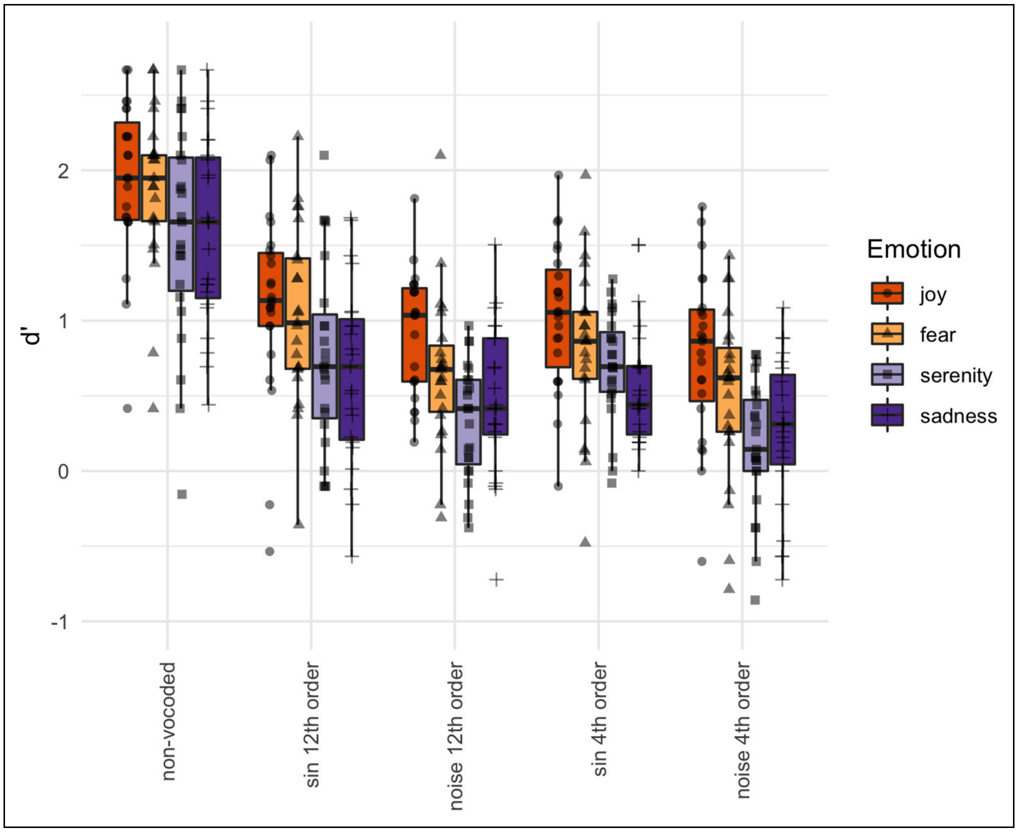

Figure 3. Sensitivity in perception of presented emotion categories (*d'*) as a function of vocoder manipulations (x-axis), and shown for different emotions (color and symbol). The horizontal line in the boxplot shows the median sensitivity across participants. The box extends from the 25th to the 75th percentile, and the whiskers extend to the value most remote from the median within 1.5 times the interquartile range. Individual data points are overlaid on top of the boxplots. *d'* = 0 denotes the chance level.

Table 4. Correlations between *d'* and confidence ratings (*df* = 458).

Vocoder	Pearson's *R*	*p*
noise 12th order	0.33	0.00368
noise 4th order	0.46	2.06E-05
sinewave 12th order	0.28	0.0102
sinewave 4th order	0.22	0.0495
non-vocoded	−0.10	0.32

within both arousal conditions. The FITA analysis below statistically addresses the visual pattern in the confusion matrices.

Feature Information Transmission Analysis (FITA)

The raw-data confusion matrices (Figure 5) indicated that confusions primarily occur within arousal class (joy confused with fear and vice versa, and serenity confused with sadness and vice versa). This pattern was analyzed with FITA using the features arousal and valence (Miller & Nicely, 1955; see 'analysis'). The FITA outcomes (T_{rel},) are shown in Figure 6. Note that variability was high across participants, with individual participant means ranging from 0 to 100% in most conditions. The RM ANOVA with factors Feature (arousal, valence) and Vocoder (non-vocoded, sinewave 12th order filter, sinewave 4th order filter, noise 12th order filter, noise 4th order filter) showed that both Feature [$F(1,22) = 83.8$, $p < 0.0001$, $\eta^2_G = 0.40$] and Vocoder [$F(4,88) = 26.11$, $p_{GG} < 0.0001$, $\eta^2_G = 0.36$] had significant effects on the relative transmission, but, importantly, the two factors also interacted weakly [$F(4,88) = 3.61$, $p < 0.01$, $\eta^2_G = 0.05$]. The ANOVA was followed up with post-hoc comparisons corrected with the fdr method.

Arousal. Post-hoc comparisons showed that for arousal, the non-vocoded and sinewave vocoders did not significantly differ in loss of transmission (p_{FDR}>0.32 for 12th order, p_{FDR}>0.06 for 4th order) while both had higher T_{rel} values than the noise vocoders (p_{FDR}<0.001 for both orders). The noise 4th order vocoder was worse at transmitting arousal than the sin 12th order vocoder [$t(22) = 3.25$, p_{FDR}<0.05]. All other comparisons for arousal transmission were non-significant (p_{FDR}>0.06).

Valence. Further post-hoc comparisons show that, for valence, the non-vocoded condition yielded higher T_{rel} values than all the vocoded conditions (p_{FDR}<0.0001). Furthermore,

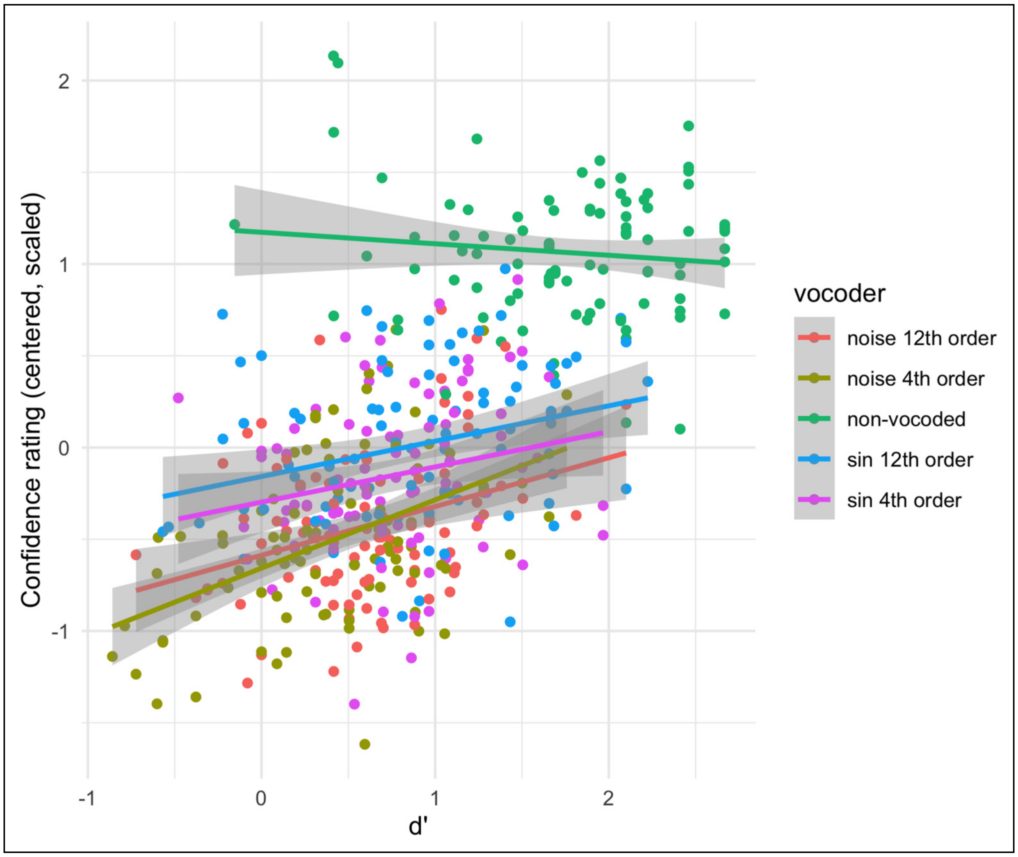

Figure 4. Confidence ratings per vocoder. In all except the non-vocoded condition, confidence was reported higher as sensitivity increased.

the sin 12th order vocoder was better at transmitting the valence than the noise 4th order vocoder [$t(22) = 2.56$, $p_{FDR} < 0.05$]. All other comparisons for valence were non-significant ($p_{FDR} > 0.06$).

Discussion

Our study investigated to what degree changes in temporal and spectral content of vocoded musical excerpts would impact music emotion categorization in NH listeners. For this purpose, we varied the carrier type and filter order of vocoders, and hypothesized that emotion categories would be better perceived in vocoded conditions that offered better temporal- (e.g., sinewave carriers vs noise carriers) and spectral (e.g., high-order vs low-order filters) content. The four emotions used covered the four quadrants of the valence and arousal plane. We hypothesized that categorization across arousal classes would improve with better temporal content, and that categorization across valence classes would improve with better spectral content. We intentionally introduced these systematic manipulations with the aim to observe how the differing content might affect CI hearing of musical emotion. This allowed a more controlled variation than can be implemented with actual CIs.

Supporting our hypothesis, the vocoder type influenced the categorization of emotions, such that better temporal content (sinewave vocoders) and better spectral content (high filter orders) both improved music emotion categorization. In addition to these main effects, the following pattern emerged: improved quality of temporal content was associated with improved categorization across emotional arousal classes, while improved quality of spectral and temporal content transmitted valence significantly better than reduced quality of spectral and temporal content. Within the parameters we have selected for spectral manipulation there was little effect of spectral content only.

Music Emotion Categorization with Normal Hearing

Our sensitivity findings with non-vocoded conditions indicated that musical emotions were successfully categorized far above chance-level. This aligns with an earlier study (Bigand et al., 2005) where participants freely sorted the same longer excerpts from existing classical works (we halved these excerpts in duration for our stimuli; see Methods) into four emotion categories that were labeled joy, fear, serenity and sadness during analysis. Bigand et al. evaluated consistency of the categorization across

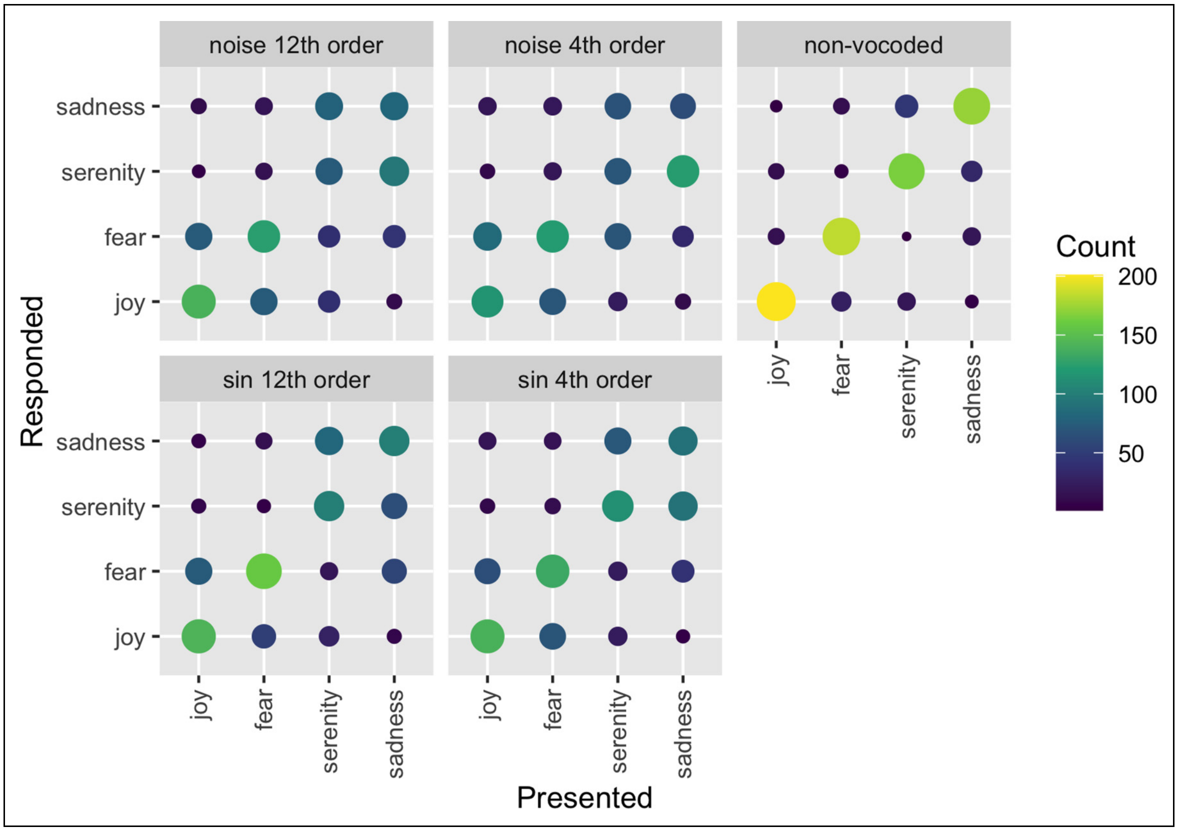

Figure 5. Confusion matrices for emotion categorization shown for each vocoder condition. The size and color of each dot is proportional to the number of relevant responses. The presented emotion categories are listed on the x-axis, while the responded emotion categories are listed on the y-axis.

participants, and it was shown that the categorization was highly stable across both musicians and non-musicians, across two experiments conducted a week apart. Moreover, categorization in turn roughly corresponded to arousal and valence dimensions (see Table 1), as well as a third, weakly emerging dimension of melodic movement. These categories were also stable even when the first few seconds of excerpts were used, though in the shortest condition the melodic movement ceased to factor into the emotion categorization. Thus we replicate the general stability of music emotion categories from this original study with the same musical excerpts in longer and shorter lengths, in participants with a range of musical experience (based on supplemental analyses presented in S3 that show no difference in sensitivity across our participants with self-reported musical vs. non-musical backgrounds).

In addition to the sensitivity, we have also collected confidence ratings. Interestingly, in our study, confidence was high for all emotion categories in the non-vocoded condition, independently of accuracy. This is in line with previous literature for NH music emotion recognition where confidence did not increase with increased accuracy (Vieillard et al., 2008). This may also be due to the fact that non-vocoded blocks were always presented after vocoded blocks, thus lowered confidence in preceding blocks was at once improved, perhaps inflating confidence, once the signal was no longer vocoded.

In the current study, sensitivity was better than chance across almost all (non-vocoded and vocoded) conditions. High arousal emotions (joy and fear) were categorized with more sensitivity and confidence than low-arousal emotions (serenity and sadness), and when mistakes in categorization were made, they tended to be within arousal class (e.g., serenity confused with sadness or joy confused with fear). In other words, high arousal emotions were categorized with more sensitivity and confidence than low-arousal emotions, independently of valence. This phenomenon was reported previously as well. One study investigated arousal and valence dimensions with the stimuli of Vieillard et al. (2008): a study on NH adults and children found a main effect of arousal, where joy and fear were more accurately identified than sadness or peacefulness, though the effect was much larger in children (Hunter et al., 2011). Results were interpreted such that cultural assignment of mode cues is less available to children than overtly salient tempo. In our case, this explanation is not likely as all participants were adults. One possible explanation pointed out by Vieillard et al. (2008) is that more musical events occur

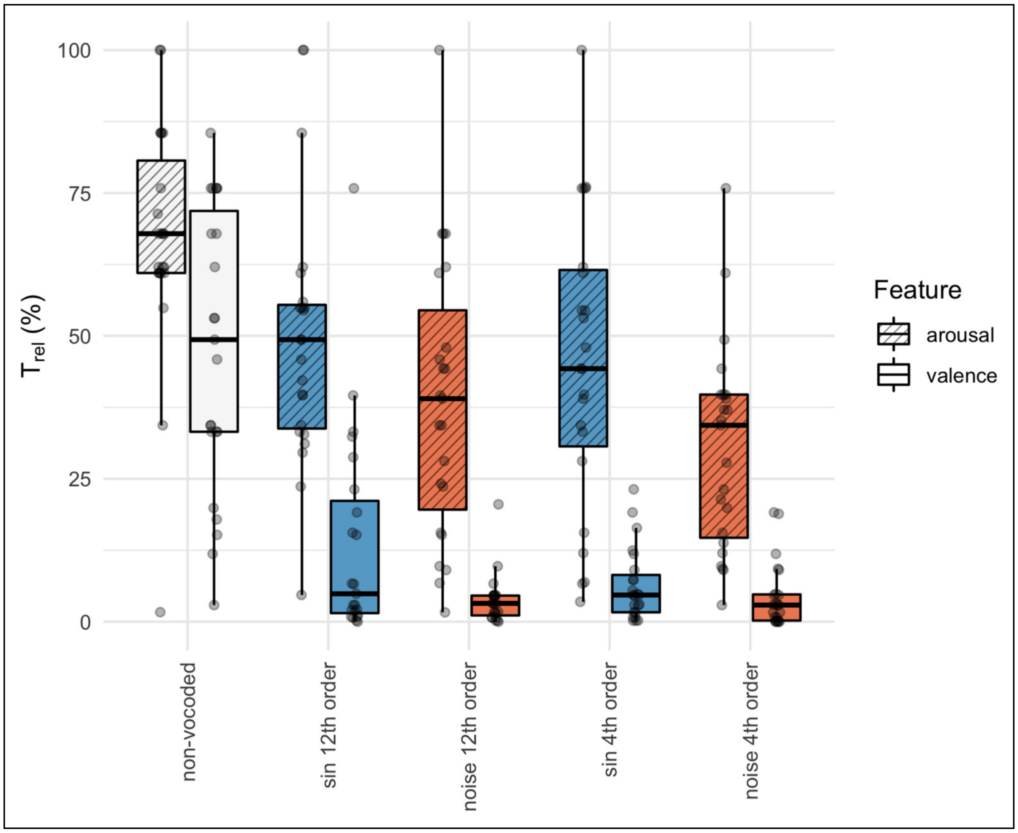

Figure 6. Relative transmission in percent of the total available information, shown as a function of feature (hatched vs. solid) and vocoder (x-axis). The sinewave and noise carriers are further marked by color (sinewave: blue; noise: red). The signification of the boxplot details is identical to that of Figure 3.

over a shorter time in high arousal conditions, thus, more information is provided to meet the decision threshold. However, Vieillard et al. (2008) did not find the same result in adults, rather categorization accuracy was greatest in happy and sad excerpts compared to threat and peaceful excerpts. Thus this effect in the current study, though comprised of non-vocoded materials, may heavily be influenced by the vocoded conditions and not interpretable per se from the perspective of music emotion categorization with NH.

Music Emotion Categorization with Vocoded Materials

Our results globally indicated that while music emotion categorization accuracy was well above chance with non-vocoded materials, vocoded conditions demonstrated significantly lower, but still above-chance performance. These differences between non-vocoded and vocoded conditions reflect previous comparisons between non-vocoded and vocoded music emotion categorization accuracy rates (Giannantonio et al., 2015; Paquette et al., 2018), validating our approach. Confidence ratings correlated with sensitivity, indicating that as sensitivity increased, so did confidence. This is intuitive and suggests that participants were aware of the missing acoustic information that they otherwise would have used to judge the emotion category.

The feature transmission analysis showed that arousal features of the signal were conveyed better across all conditions. Non-vocoded and sine-wave-vocoded conditions were collectively better at transmitting arousal compared to noise vocoders, with little influence of filter order. Thus perhaps the noise vocoders disrupted the temporal content beyond some threshold that efficiently conveyed emotional arousal, which was still discernible in the sine and non-vocoded conditions. On the other hand, valence feature transmission, though significantly better in high order filter with sinewave carrier compared to low order filter with noise carrier conditions, was perceived significantly worse than arousal features in all vocoded conditions. In other words valence transmission only improved with combined improvement of spectral and temporal content quality. Within the parameters selected for spectral manipulation, there was little effect of spectral content alone on valence perception.

This finding is consistent with a previous music emotion study that investigated vocoded and noise-masked CI-simulated materials in NH listeners (Giannantonio et al., 2015). The paradigm used piano excerpts (~10 s long) taken from Western classical music with happy (major mode, fast tempo 80–255 bpm) and sad (minor mode, slow tempo 20–100 bpm) emotions, and changed the same excerpts to the opposite mode or a neutral tempo (80 bpm). They then compared the categorization of original and changed excerpts in order to assess which features, tempo or mode, informed the categorization. It was found that modal cues informed non-vocoded/non-masked music emotion perception but exclusively tempo cues informed vocoded and masked music emotion perception. Thus spectral cues were not used as well as temporal cues during music emotion categorization with degraded acoustic signals. Our findings were able to generalize this phenomenon to a paradigm with four emotions. Moreover, our shift along temporal and spectral dimensions in the vocoders further highlight the role of temporal information informing music emotion categorization with reduced acoustic input.

Another music emotion perception study approximating CI listening with vocoded materials and NH participants found somewhat contrasting results (Paquette et al., 2018), where valence and arousal were rated during categorization of simple melodies of bursts of notes of short duration (<2 s) containing the emotions happiness, sadness, fear and neutrality. It was found that NH listeners with vocoded stimuli used the timbral features of energy and roughness to inform both arousal and valence ratings. In that study, stimuli were taken from a set of improvised emotional content played on clarinet and violin instruments, consisting of only a few notes (Paquette et al., 2013). While the stimuli description does not indicate whether chords were played on the violin, at least half of the stimuli were monophonic (played on the clarinet, which can only play one note at a time), in contrast to our fully orchestrated excerpts or classical piano excerpts in Giannantonio et al. (2015). It may be that vocoded monophonic bursts in their materials provided clearer spectral cues than the fully orchestrated excerpts from our study or Giannantonio et al. (2015), offering more salient spectral information. It would be interesting to apply the current vocoders with varying spectral and temporal content to the Paquette et al. (2018) paradigm, to evaluate how valence and arousal cues in short bursts are modulated as the quality of spectral and temporal content varies. If results still diverge from vocoded conditions in Giannantonio et al. (2015) and the current study, it would add an intriguing nuance to future directions for music emotion perception with CI (simulated) hearing.

Limitations to the Current Paradigm

Vocoder Parameters. The current study systematically varied the quality of temporal and spectral content by way of changing carrier and filter order, respectively. While we chose these parameters based on previous studies that manipulated spread of excitation (Bingabr et al., 2008; Crew & Galvin, 2012), the choices are still only capturing a limited range of temporal and spectral content variation. Moreover, it cannot be said that the two parameters have equivalent effects, ie., that our current spread contrast (4th vs. 12th order) is meant to yield a physical or perceptual difference equivalent to that induced by our carrier contrast (sinewave vs. noise). Instead, we can only argue that for each parameter, we had one condition providing more information than the other. This means that any interaction, or lack thereof, as it happened, could be very different if we had chosen different parameter values (e.g., if we had contrasted 2nd order to 24th order, or if we had compared sinewave carriers to pulse-spreading harmonic complexes).

This being said, the parameter values were not chosen arbitrarily, but were instead based on reports in the literature in order to either realistically mimic certain aspects of electrical stimulation, or to allow comparison to previously collected data. We used 16 channels in all conditions, as this is an average of what modern implants offer (12–22 channels) and 16 channels were moreover used in Bingabr et al. (2008) and Crew and Galvin (2012). We moreover took the best spectral content that previously aided melodic contour identification in Crew and Galvin (2012), 4th order filters, and treated it as our lowest spectral content here. Our best spectral content was achieved with a 12th order filter, which we found during informal piloting to offer the best noticeable contrast in spectral content. Future studies could add more conditions, or a different range; for example, while spectral quality seemed to have a small effect on music emotion categorization in the current study, perhaps a larger difference would be observed if we degraded content even further, e.g., 1st order filter).

Furthermore, while we branded the two parameters as temporal vs. spectral, their effect on the stimuli is not as orthogonal as one may wish, and we cannot rule out that spectral differences might have resulted from our manipulation in the temporal domain, and vice-versa, which could in turn influence emotion categorization. Indeed, decreased filter order could also potentially result in smearing temporal information in each channel as it gets mixed up more with neighboring channels. Nevertheless, one can expect this effect to remain relatively limited compared to the drastic change of carrier we introduced: while the noise carrier introduces an exogenous source of noise in the temporal envelopes, current spread only mixes envelope information coming from the same original signal, and reduces envelope information in as much as the envelope is uncorrelated across bands (for music, one could expect it to be relatively well correlated). As for the reciprocal influence of carrier nature on spectral resolution, specific steps have been taken to limit its extent, notably by making sure that the sinewave carrier is not generating unresolved side-bands. Therefore, here as well, we can expect that the manipulation of filter order should be the primary drive of spectral resolution, and that carrier type should only have a secondary effect.

Since we found here that the responses to first and second halves of the stimuli were highly correlated (see supplementary analysis in S4), our experiment can be conducted with half the amount of items, leaving room to test more parameters in future paradigms. In order to preserve fidelity of emotion during performance, intensity in our stimuli was kept ecological within and across emotion categories by our loudness adjustment experiment (see supplementary materials). This was mostly aimed at preserving the ecological cues in the vocoded condition, as this is the first instance that the current stimuli were presented with a vocoder. Loudness, along with tempo, is informative of arousal, and in turn arousal cues are the most useful type of cue for CI users. Ultimately, increased dynamic range may allow larger differences in intensity that could provide more salience to loudness and in turn arousal. Thus spectral energy may prove an interesting parameter to systematically adjust in vocoder conditions in future research.

Interpretation of Observed Effects. We furthermore cannot assume that the vocoders introduce all of the confusion between emotions in the stimuli. Indeed, in the non-vocoded condition (Figure 5), there was some categorical confusion between joy and fear, and to a stronger extent serenity and sadness. While vocoders might have amplified an existing confusion between categories, the acoustic cues to valence were perhaps more subtle than the acoustic cues to arousal and therefore more susceptible to degradations.

We also note that we are not per se investigating the emotion experienced by participants when they hear the music, rather we asked them to identify which emotion corresponds to the music heard. Participants were not asked to distinguish between felt and perceived emotions, a distinction that can be ambiguous to perform in an explicit way (e.g., Scherer, 2004), or even impossible according to theories of embodied cognition (e.g., Niedenthal, 2007). Moreover, the acoustic manipulation in vocoded conditions may have distanced recognition of the content from the actual emotional experience, in particular for NH participants who are not used to listening to music with reduced spectrotemporal content. Our investigation thus is limited to the recognition of intended emotional content of music by the composer/performers, also with reduced spectrotemporal content, can generally contribute to overall music listening pleasure and experienced emotion (e.g., Sachs et al., 2015; Fuller et al., 2019, 2021). Future studies could moreover explore the perceived/expressed emotion judgements with either open choice responses or more fine-grained response options.

Implications for Music Emotion Categorization with Cochlear-Implant Hearing

The two previous studies investigating music emotion categorization with vocoded conditions described above (Giannantonio et al., 2015; Paquette et al., 2018) also incorporated experiments with actual CI users. Giannantonio et al. (2015) found that CI users rely more strongly on tempo cues than NH listeners, and that when modal/pitch cues factored in ratings, these were linked to residual hearing abilities in the implant users. Our findings from vocoded conditions align with these results, indicating greater reliance on tempo to judge music emotion without spectral cues (e.g., vocoded simulations have less spectral information than might be available in residual hearing). We moreover show that this is the case with listeners with a broad range of musical training, though this influence was assessed only in a supplementary analysis here (see S3). Our study supports the findings from Giannantonio et al. (2015), which in turn suggests that the findings may extend to stimuli with natural tempo and modal changes, and into emotional dimensions of valence and arousal, in CI users.

Paquette et al. (2018) found that CI users used the timbral features of energy and roughness to inform arousal and valence such that increased energy and roughness were associated with lower valence and arousal ratings. Interestingly, the minimum pitch in the burst also informed arousal (though other spectral cues did not inform categorization; brightness, maximum- and mean pitch), contrary to our results. However; CI users have previously demonstrated limited ability to perceive the contour of a short succession of notes (Galvin et al., 2007), which may be more accessible than in longer musical passages such as fully orchestrated classical music excerpts used in our study.

While the literature is consistent that valence is informed strongly by spectral cues among normal-hearing listeners, the literature is also consistent in reporting that CI users are not able to efficiently make use of such spectral cues. This was echoed by our vocoded condition results, where valence was not efficiently transmitted in the vocoded signal in either of the spectral content conditions. An alternative to using spectral cues to inform valence could be to use sensory dissonance from envelope roughness as a musical device: A previous study found that CI users rated dissonant-chord accompaniment to be as pleasant as consonant-chord accompaniment, whereas NH listeners rated dissonant-chord accompaniment to be less pleasant than consonant-chord accompaniment (Caldwell et al., 2016). All chords in the stimuli were either dissonant or consonant, with no within-trial variation. Thus if CI users can detect envelope roughness, it is possible that they might still be able to detect differences in dissonance and consonance if played in succession within the same passage. Future research could investigate how sensitive CI users are in perceiving envelope roughness, and whether they are inclined to attribute this cue to valence or whether a valence association with roughness might be created by cultural instruction to inform musical emotion perception. That could in turn allow tension and relaxation patterns associated with music enjoyment (Bigand et al., 1996; Lehne & Koelsch, 2014) to become available to CI users.

It might be argued that while our results seem to align with previous studies with a happy-sad paradigm, our findings differ from a previous report of CI music emotion categorizations with more nuanced valence and arousal ratings and more ecological (e.g., no manipulated mode or tempo) stimuli similar to ours (Ambert-Dahan et al., 2015). On a visual analog scale, CI-users rated the emotion expressed in musical excerpts in terms of the presence of happiness, fear, sadness and peacefulness, as well as the arousal and valence of each excerpt. Overall, it was found that CI users had preserved valence perception but impaired arousal perception, and that 'peacefulness' was the most accurately rated emotion. However, an explanation why this study may have found preserved valence in CI hearing compared to NH hearing was that 11 of the 13 CI user participants had residual hearing in the contralateral (hearing-aided) ear. Residual hearing significantly benefits music perception in CI users (El Fata et al., 2009; Gfeller et al., 2006), including, as previously indicated (Giannantonio et al., 2015), that CI users could use modal cues to perceive emotion when they have residual hearing. Thus we speculate that results may have resembled our findings with vocoded materials (and the CI-alone condition in D'Onofrio et al., 2020) if CI users were using CI hearing alone with no acoustic hearing. Moreover, the analysis was carried out differently than in our study, assessing the difference between CI user and NH responses within each emotion category but not what the responses were relative to other emotion categories within groups. For example, visual inspection of the results indicates that the pattern of responses was perhaps similar to our findings: arousal ratings seemed to be higher for happy and fear than for peaceful and sad in both groups, but with globally lower arousal ratings for CI compared to NH listeners. It is furthermore not clear whether valence ratings were different at all among the categories within the groups, especially CI users.

Finally, a main finding of our vocoded conditions was that arousal was still conveyed efficiently with less temporal content. Moreover, changing spectral content among vocoders had only a small impact on valence transmission, and then in combination with the quality of temporal content. When considering the implication of this for CI users, it is important to note that manipulation of temporal information via different types of vocoding is different from what happens in cochlear implants. However, our results could indicate that even implants that transduce signals with less optimal temporal content could still be capable of conveying emotion in music, reaffirming generalizability of previous music emotion research with CI users, especially arousal. Considering that we found the quality of temporal content to impact both arousal and valence in vocoded conditions, future technology efforts aimed at improving the quality of temporal content in implanted hearing may directly benefit music emotion perception for CI users.

Conclusion

Music emotion perception in CI users is limited by the acoustic cues that can be transmitted by the electrical signal of the implant. By systematically varying availability of acoustic cues along temporal and spectral dimensions, our vocoder results with NH participants complement previous results that arousal cues informed by temporal content are likely most reliably available to the CI listeners and form the basis of their music emotion recognition. In vocoded conditions, arousal with both higher and lower temporal content was perceived better than valence, even in the condition with better spectral content. While these observations may be specific to our study and the specific parameters we have used in vocoding, the results of the present study seem in line with previous results from CI users. Based on this, we speculate that while increasing spectral resolution in implants might certainly help with music emotion perception when technically possible, our findings suggest that a promising direction for efforts to improve music emotion perception for CI users may also lie in the temporal domain. Future research could focus on CI users' perception of envelope roughness in musical passages, for example; an interesting direction for CI-designed music composition might be to create musical tension-relaxation patterns with alternating degrees of envelope roughness.

Acknowledgments

We would like to thank Dr. Christina Fuller for support in conceptual planning and proofreading the manuscript, and Ms. Jennifer Breetveld for administrative support.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by the Dorhout Mees Foundation, the Gratama Foundation, Heinsus Houbolt Funds, and the VICI grant (grant number 016.VICI.170.111) from the Netherlands Organization for Scientific Research (NWO) and the Netherlands Organization for Health Research and Development (ZonMw). The team Auditory Cognition and Psychoacoustics of the Lyon Neuroscience Research Center is part of the LabEx CeLyA (Centre Lyonnais d'Acoustique, ANR-10-LABX-60). The data that support the findings of this study will be openly available in the DataverseNL repository via https://doi.org/10.34894/1REVFB.

ORCID iDs

Eleanor E. Harding https://orcid.org/0000-0002-3244-9625
Etienne Gaudrain https://orcid.org/0000-0003-0490-0295
Bert Maat https://orcid.org/0000-0001-9856-7010
Deniz Başkent https://orcid.org/0000-0002-6560-1451

Supplemental Material

Supplemental material for this article is available online.

Note

[1] The scope of our study is limited to Western tonal music.

References

Ambert-Dahan, E., Giraud, A. L., Sterkers, O., & Samson, S. (2015). Judgment of musical emotions after cochlear implantation in adults with progressive deafness. *Frontiers in Psychology*, *6*, 181. https://doi.org/10.3389/fpsyg.2015.00181

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*(3), 379–384. https://doi.org/10.3758/BF03192707

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H., Singmann, H., & Dai, B. (2015). lme4: Linear mixed-effects models using Eigen and S4[computer program]. *R package version 1*.1–7. 2014.

de Balthasar, C., Boëx, C., Cosendai, G., Valentini, G., Sigrist, A., & Pelizzone, M. (2003). Channel interactions with high-rate biphasic electrical stimulation in cochlear implant subjects. *Hearing Research*, *182*(1), 77–87. https://doi.org/10.1016/S0378-5955(03)00174-6

Başkent, D., Gaudrain, E., Tamati, T., & Wagner, A. E. (2016). Perception and Psychoacoustics of Speech in Cochlear Implant Users. In A. T. Cacace, E. de Kleine, A. G. Holt, & P. van Dijk (Eds.), *Scientific Foundations of Audiology: Perspectives from Physics, Biology, Modeling, and Medicine* (pp. 285–319). Plural Publishing.

Benard, M. R., & Başkent, D. (2014). Perceptual learning of temporally interrupted spectrally degraded speech. *The Journal of the Acoustical Society of America*, *136*(3), 1344–1351. https://doi.org/10.1121/1.4892756

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berg, R., & Stork, D. (2004). *The Physics of Sound* (3rd ed). Pearson.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*(1), 125–141. https://doi.org/10.3758/BF03205482

Bigand, E., Vieillard, S., Madurell, F., Marozeau, J., & Dacquet, A. (2005). Multidimensional scaling of emotional responses to music: The effect of musical expertise and of the duration of the excerpts. *Cognition & Emotion*, *19*(8), 1113–1139. https://doi.org/10.1080/02699930500204250

Bingabr, M., Espinoza-Varas, B., & Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hearing Research*, *241*(1–2), 73–79. https://doi.org/10.1016/j.heares.2008.04.012

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., Dillier, N., Dowell, R., Fraysse, B., Gallégo, S., Govaerts, P. J., Green, K., Huber, A. M., Kleine-Punte, A., Maat, B., Marx, M., Mawman, D., Mosnier, I., O'Connor, A. F., O'Leary, S., … Lazard, D. S.. 2012. Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients. *Audiology & Neurotology* *18*(1):36–47. https://doi.org/10.1159/000343189

Caldwell, M. T., Jiam, N. T., & Limb, C. J. (2017). Assessment and improvement of sound quality in cochlear implant users. *Laryngoscope Investigative Otolaryngology*, *2*(3), 119–124. https://doi.org/10.1002/lio2.71

Caldwell, M. T., Jiradejvong, P., & Limb, C. J. (2016). Impaired perception of sensory consonance and dissonance in cochlear implant users. *Otology & Neurotology*, *37*(3), 229–234. https://doi.org/10.1097/MAO.0000000000000960

Caldwell, M. T., Ranklin, S. K., Jiradejvong, P., Carver, C., & Limb, C. J. (2015). Cochlear implant users rely on tempo rather than on pitch information during perception of musical emotion. *Cochlear Implants International*, *16*(3), S114–S120. https://doi.org/10.1179/1467010015Z.000000000265

Carlyon, R. P. (1997). The effects of two temporal cues on pitch judgments. *The Journal of the Acoustical Society of America*, *102*, 1097–1105. https://doi.org/10.1121/1.419861

Clark, G. (2004). *Cochlear implants. In: Speech Processing in the Auditory System. Springer Handbook of Auditory Research, vol 18*. Springer, New York, NY. https://doi.org/10.1007/0-387-21575-1_8

Cooper, W. B., Tobey, E., & Loizou, P. C. (2008). Music perception by cochlear implant and normal hearing listeners as measured by the montreal battery for evaluation of amusia. *Ear & Hearing*, *29*(4), 618–626. https://doi.org/10.1097/AUD.0b013e318174e787

Crew, J. D., & Galvin, J. J. (2012). Channel interaction limits melodic pitch perception in simulated cochlear implants. *The Journal of the Acoustical Society of America*, *132*(5), EL429–EL435. https://doi.org/10.1121/1.4758770

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*(2), 222–241. https://doi.org/10.1037/0096-3445.134.2.222

D'Onofrio, K. L., Caldwell, M., Limb, C., Smith, S., Kessler, D. M., & Gifford, R. H. (2020). Musical emotion perception in bimodal patients: Relative weighting of musical mode and tempo cues. *Frontiers in Neuroscience*, *14*, 114. https://doi.org/10.3389/fnins.2020.00114

Fata, El, James, F., Laborde, C. J., & & Fraysse, M. L., B. (2009). How much residual hearing is 'useful' for music perception with cochlear implants?. *Audiology & Neuro-otology*, *14*(Suppl 1), 14–21. https://doi.org/10.1159/000206491. Epub 2009 Apr 22. PMID: 19390171.

Feldman, L. A. (1995). Valence focus and arousal focus: Individual differences in the structure of affective experience. *Journal of Personality and Social Psychology*, *69*(1), 153. https://doi.org/10.1037/0022-3514.69.1.153

Filipic, S., Tillmann, B., & Bigand, E. (2010). Judging familiarity and emotion from very brief musical excerpts. *Psychonomic Bulletin & Review*, *17*(3), 335–341. https://doi.org/10.3758/PBR.17.3.335

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed). Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Friesen, L., Shannon, R. V., Başkent, D., & Wang, Z. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *The Journal of the Acoustical Society of America*, *110*(2), 1150–1163. https://doi.org/10.1121/1.1381538

Fu, Q. J., & Nogaki, G. (2005). Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing. *JARO - Journal of the Association for Research in Otolaryngology*, *6*(1), 19–27. https://doi.org/10.1007/s10162-004-5024-3

Fuller, C., Başkent, D., & Free, R. (2019). Early Deafened, Late Implanted Cochlear Implant Users Appreciate Music More Than and Identify Music as Well as Postlingual Users. *Frontiers in Neuroscience*, *13*(October), Article 1050. https://doi.org/10.3389/fnins.2019.01050

Fuller, C., Free, R., Maat, B., & Başkent, D. (2021). Self-Reported music perception is related to quality of life and self-reported hearing abilities in cochlear implant users. *Cochlear Implants International*, *0*(0), 1–10. https://doi.org/10.1080/14670100.2021.1948716

Gabrielsson, A., & Lindström, E. (2001). The influence of musical structure on emotional expression. In *Music and emotion: Theory and research, series in affective science* (pp. 223–248). Oxford University Press.

Galvin, J. J., Fu, Q.-J., & Nogaki, G. (2007). Melodic contour identification by cochlear implant listeners. *Ear & Hearing*, *28*(3), 302–319. https://doi.org/10.1097/01.aud.0000261689.35445.20

Garrido, S., & Schubert, E. (2011). Individual differences in the enjoyment of negative emotion in music: A literature review and experiment. *Music Perception*, *28*(3), 279–296. https://doi.org/10.1525/mp.2011.28.3.279

Gaudrain, E., & Başkent, D. (2015). Factors limiting vocal-tract length discrimination in cochlear implant simulations. *The Journal of the Acoustical Society of America*, *137*(3), 1298–1308. https://doi.org/10.1121/1.4908235

Gaudrain, E., & Başkent, D. (2018). Discrimination of voice pitch and vocal-tract length in cochlear implant users. *Ear & Hearing*, *39*(2), 226–237. https://doi.org/10.1097/AUD.0000000000000480

Gfeller, K., Christ, A., Knutson, J. F., Witt, S., Murray, K. T., & Tyler, R. S. (2000). Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients. *Journal of the American Academy of Audiology*, *11*, 390–406. https://doi.org/10.1055/s-0042-1748126

Gfeller, K. E., Olszewski, C., Turner, C., Gantz, B., & Oleson, J. (2006). Music perception with cochlear implants and residual hearing. *Audiology & Neuro-otology*, *11*(Suppl 1), 12–15. https://doi.org/10.1159/000095608. Epub 2006 Oct 6. PMID: 17063005.

Giannantonio, S., Polonenko, M. J., Papsin, B. C., Paludetti, G., & Gordon, K. A. (2015). Experience changes how emotion in music is judged: Evidence from children listening with bilateral cochlear implants, bimodal devices, and normal hearing. *PLoS ONE*, *10*(8), 1–29. https://doi.org/10.1371/journal.pone.0136685

Green, D. M., & Swets, J. A. (1988). *Signal Detection Theory and Psychophysics* (Reprint edition). Peninsula.

Greenwood, D. D. (1990). A cochlear frequency–position function for several Species—29 years later. *The Journal of the Acoustical Society of America*, *87*(6), 2592–2605. https://doi.org/10.1121/1.399052

Grewe, O., Nagel, F., Kopiez, R., & Altenmüller, E. (2005). How does music arouse 'chills'? *Annals of the New York Academy of Sciences*, *1060*(1), 446–449. https://doi.org/10.1196/annals.1360.041

Hopyan, T., Gordon, K. A., & Papsin, B. C. (2013). Identifying emotions in music through electrical hearing in deaf children using cochlear implants. *Cochlear Implants International*, *12*(1), 21–26. https://doi.org/10.1179/146701010X12677899497399

Hunter, P. G., Schellenberg, G., E., & Stalinski, S. M. (2011). Liking and identifying emotionally expressive music: Age and gender differences. *Journal of Experimental Child Psychology*, *110*(1), 80–93. https://doi.org/10.1016/j.jecp.2011.04.001

Ilie, G., & Thompson, W. F. (2006). A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, *23*(4), 319–330. https://doi.org/10.1525/mp.2006.23.4.319

Koelsch, S., Fritz, T., V Cramon, D. Y., Müller, K., & Friederici, A. D. (2005). Investigating emotion with music: An FMRI study. *Human Brain Mapping*, *27*(3), 239–250. https://doi.org/10.1002/hbm.20180

Lassaletta, L., Castro, A., Bastarrica, M., Pérez-Mora, R., Herrán, B., Sanz, L., de Sarriá, M. J., & Gavilán, J. (2008). Musical perception and enjoyment in post-lingual patients with cochlear implants. *Acta Otorrinolaringologica (English Edition)*, *59*(5), 228–234. https://doi.org/10.1016/s2173-5735(08)70228-x

Lawrence, M. A. (2016). *Package "ez"*. [computer program]. Version 4.

de Leeuw, J. R. (2015). Jspsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2021). Emmeans: Estimated marginal means, aka least-squares means[computer program]. *R Package Version 1* (2018).

Lehne, M., & Koelsch, S. (2014). Tension-Resolution patterns as a key element of aesthetic experience: Psychological principles and underlying brain mechanisms. In *Art, aesthetics, and the brain* (pp. 545). Oxford University Press.

Lévêque, Y., Teyssier, P., Bouchet, P., Bigand, E., Caclin, A., & Tillmann, B. (2018). Musical emotions in congenital amusia: Impaired recognition, but preserved emotional intensity. *Neuropsychology*, *32*(7), 880–894. https://doi.org/10.1037/neu0000461

Liégeois-Chauvel, C., Bénar, C., Krieg, J., Delbé, C., Chauvel, P., Giusiano, B., & Bigand, E. (2014). How functional coupling between the auditory Cortex and the amygdala induces musical emotion: A single case study. *Cortex*, *60*, 82–93. https://doi.org/10.1016/j.cortex.2014.06.002

Loizou, P. C. (1998). Mimicking the human ear. *IEEE Signal Processing Magazine*, *15*(5), 101–130. https://doi.org/10.1109/79.708543

Macherey, O., & Carlyon, R. P. (2014). Cochlear implants. *Current Biology*, *24*(18), R878–R884. https://doi.org/10.1016/j.cub.2014.06.053

Macmillan, N. A., & Douglas Creelman, C. (2004). *Detection Theory: A User's Guide*. Psychology Press.

Mazaheryazdi, M., Aghasoleimani, M., Karimi, M., & Arjmand, P. (2018). Perception of musical emotion in the students with cognitive and acquired hearing loss. *Iranian Journal of Child Neurology*, *12*(2), 41–48. https://doi.org/10.22037/ijcn.v12i2.14598

McKenna, V. S., & Stepp, C. E. (2018). The relationship between acoustical and perceptual measures of vocal effort. *The Journal of the Acoustical Society of America*, *144*(3), 1643–1658. https://doi.org/10.1121/1.5055234

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *The Journal of the Acoustical Society of America*, *27*(2), 338–352. https://doi.org/10.1121/1.1907526

Niedenthal, P. M. (2007). Embodying emotion. *Science (New York, N.Y.)*, *316*(5827), 1002–1005. https://doi.org/10.1126/science.1136930

Nieminen, S., Istók, E., Brattico, E., & Tervaniemi, M. (2012). The development of the aesthetic experience of music: Preference, emotions, and beauty. *Musicae Scientiae*, *16*(3), 372–391. https://doi.org/10.1177/1029864912450454

Paquette, S., Ahmed, G. D., Goffi-Gomez, M. V., Hoshino, A. C. H., Peretz, I., & Lehmann, A. (2018). Musical and vocal emotion perception for cochlear implants users. *Hearing Research*, *370*, 272–282. https://doi.org/10.1016/j.heares.2018.08.009

Paquette, S., Peretz, I., & Belin, P. (2013). The 'Musical Emotional Bursts': A Validated Set of Musical Affect Bursts to Investigate Auditory Affective Processing. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00509

Pralus, A., Belfi, A., Hirel, C., Lévêque, Y., Fornoni, L., Bigand, E., Jung, J., Tranel, D., Nighoghossian, N., Tillmann, B., & Caclin, A. (2020). Recognition of musical emotions and their perceived intensity after unilateral brain damage. *Cortex*, *130*, 78–93. https://doi.org/10.1016/j.cortex.2020.05.015

Sachs, M. E., Damasio, A., & Habibi, A. (2015). The pleasures of sad music: A systematic review. *Frontiers in Human Neuroscience*, *9*, 404. https://doi.org/10.3389/fnhum.2015.00404

Scherer, K. R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of new Music Research*, *33*(3), 239–251. https://doi.org/10.1080/0929821042000317822

Schubert, E. (2004). Modeling perceived emotion with continuous musical features. *Music Perception*, *21*(4), 561–585. https://doi.org/10.1525/mp.2004.21.4.561

Shannon, R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction. *Hearing Research*, *12*(1), 1–16. https://doi.org/10.1016/0378-5955(83)90115-6

Shannon, R. V. (1992). Temporal modulation transfer functions in patients with cochlear implants. *The Journal of the Acoustical Society of America*, *91*(4), 2156–2164.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.)*, *270*(5234), 303–304. https://doi.org/10.1126/science.270.5234.303

Temperley, D., & de Clercq, T. (2013). Statistical analysis of harmony and melody in rock music. *Journal of New Music Research*, *42*(3), 187–204. https://doi.org/10.1080/09298215.2013.788039

Tramo, M. J., Cariani, P. A., Delgutte, B., & Braida, L. D. (2001). Neurobiological foundations for the theory of harmony in western tonal music. *Annals of the New York Academy of Sciences*, *930*(1), 92–116. https://doi.org/10.1111/j.1749-6632.2001.tb05727.x

Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). Method for the selection of sentence materials for efficient measurement of the speech reception threshold. *The Journal of the Acoustical Society of America*, *107*(3), 1671–1684. https://doi.org/10.1121/1.428451

Vieillard, S., Peretz, I., Gosselin, N., Khalfa, S., Gagnon, L., & Bouchard, B. (2008). Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion*, *22*(4), 720–752. https://doi.org/10.1080/02699930701503567

van Wieringen, A., & Wouters, J. (1999). Natural vowel and consonant recognition by laura cochlear implantees. *Ear and Hearing*, *20*(2), 89–103. https://doi.org/10.1097/00003446-199904000-00001