Trends in Hearing
Volume 27: 1–18
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23312165221148022
journals.sagepub.com/home/tia
SAGE

Optimization of Sound Coding Strategies to Make Singing Music More Accessible for Cochlear Implant Users

Sina Tahmasebi[1,2], Manuel Segovia-Martinez[3], and Waldo Nogueira[1,2]

Abstract

Cochlear implants (CIs) are implantable medical devices that can partially restore hearing to people suffering from profound sensorineural hearing loss. While these devices provide good speech understanding in quiet, many CI users face difficulties when listening to music. Reasons include poor spatial specificity of electric stimulation, limited transmission of spectral and temporal fine structure of acoustic signals, and restrictions in the dynamic range that can be conveyed via electric stimulation of the auditory nerve. The coding strategies currently used in CIs are typically designed for speech rather than music. This work investigates the optimization of CI coding strategies to make singing music more accessible to CI users. The aim is to reduce the spectral complexity of music by selecting fewer bands for stimulation, attenuating the background instruments by strengthening a noise reduction algorithm, and optimizing the electric dynamic range through a back-end compressor. The optimizations were evaluated through both objective and perceptual measures of speech understanding and melody identification of singing voice with and without background instruments, as well as music appreciation questionnaires. Consistent with the objective measures, results gathered from the perceptual evaluations indicated that reducing the number of selected bands and optimizing the electric dynamic range significantly improved speech understanding in music. Moreover, results obtained from questionnaires show that the new music back-end compressor significantly improved music enjoyment. These results have potential as a new CI program for improved singing music perception.

Keywords

cochlear implant, music perception, sound coding strategy, dynamic range, compression

Received 15 December 2021; Revised received 7 December 2022; accepted 11 December 2022

Introduction

The cochlear implant (CI) is a surgically implanted neuroprosthesis that can partially restore the sense of hearing to people with severe to profound sensorineural hearing loss. Although most CI users obtain very good speech understanding in quiet, they experience limited speech understanding in noisy environments and obtain poor music perception (Limb & Roy, 2013; McDermott, 2004). Previous research in the field of music and CIs has focused on making music more accessible for CI users by reducing the spectral complexity of music (Gauer et al., 2019; Nagathil et al., 2016, 2017) or enhancing the singing voice in popular music (Buyens et al., 2014; Pons et al. 2016; Gajecki & Nogueira, 2018; Tahmasebi et al., 2020) through pre-processing or front-end algorithms. The present work investigates the optimization of the CI sound coding strategy to reduce the complexity and enhance the singing voice of music without the need for pre-processing algorithms.

CIs transmit a low number of frequency channels, ranging from 12–22 depending on the CI manufacturer, corresponding with the number of available electrodes. Moreover, the CI electrodes are surrounded by a highly conductive fluid, which causes large spread of electric current in the cochlea when stimulated, leading to channel interactions (Bierer, 2007; Chatterjee et al., 2006; McKay et al., 1996; Stickney

[1]Department of Otolaryngology, Hannover Medical School, Hannover, Germany
[2]Cluster of Excellence Hearing4all, Hannover, Germany
[3]Oticon Medical, Vallauris, France

Corresponding authors:
Sina Tahmasebi, Karl-Wiechert-Allee 3, 30625 Hannover, Germany.
Email: tahmasebi.sina@mh-hannover.de

Waldo Nogueira, Karl-Wiechert-Allee 3, 30625 Hannover, Germany.
Email: nogueiravazquez.waldo@mhhannover.de

et al., 2006). Furthermore, there often exists a mismatch between the transmitted frequencies by the CI sound processor and the corresponding locations of the stimulated electrodes, leading to the so-called tonotopic mismatch (Greenwood, 1991; Stakhovskaya et al., 2007). Lastly, the upper limit of temporal pitch (Carlyon et al., 2008; Macherey, 2010) in CI users**,** conveyed via the electrical pulse rate or periodic fluctuations in the temporal envelope of the pulse amplitudes, has been shown to be significantly lower (~300 Hz) than in normal hearing (NH) listeners (~1500 Hz; Verschooten et al., 2019). These distortions cause limited spectral resolution and pitch perception in CI users (there are however exceptions; Goldsworthy & Shannon, 2014) leading to poor melody and timbre perception (Gfeller et al., 1998; Hsiao & Gfeller, 2012; Schulz & Kerber, 1994). Moreover, CIs provide only a narrow electric dynamic range (EDR) to encode the wide acoustic dynamic range (DR) of speech and music. While the DR of acoustic hearing in NH listeners is up to around 120 dB (Zeng, 2004), CI users have an electric DR of around 6–20 dB (see Skinner et al., 1997; Zeng & Galvin, 1999). This inherent compression has multiple potential negative effects on music perception (Drennan & Rubinstein, 2008).

Previous studies have shown that the ability to recognize songs by CI users mainly depends on understanding the lyrics (Fujita & Ito, 1999). Gfeller et al. (2005) also showed that CI users rely on lyrics in the music to recognize melody or to identify pitch. Lyrics can help CI recipients compensate for poor pitch and melody perception (Gfeller, 2009) and appreciate music (Sorkin & Zombek, 2021). Hsiao (2008) showed that pediatric CI recipients performed with greater accuracy in melody recognition when lyrics were available. Buyens et al. (2014) showed that CI users enjoy popular music more when the vocals are enhanced by 6 dB with the respect to the background instruments. In this context, different pre-processing approaches or front-end algorithms have been proposed to enhance the vocals-to-instruments ratio (VIR) within an audio mixture (Gajecki & Nogueira, 2018; Pons et al., 2016; Tahmasebi et al., 2020). Recently, source separation using deep neural networks (DNNs) has been optimized to improve music enjoyment for CI listeners based on objective measures. Tahmasebi et al. (2020) used objective measures to design and optimize a DNN model based on a multilayer perceptron (MLP; Murtagh, 1991) that separates the singing voice from the background instruments under realistic sound scenarios and allowed the listener to apply the desired amount of enhancement on the singing voice in real-time. However, limitations such as high computation demand, reduced battery life of the speech processor, latency introduced by DNNs, and the poorer performance on unseen data make these algorithms too challenging and complicated to be implemented in current CI sound processors.

A more practical approach to make music more accessible for CI users is to optimize the sound coding strategy to reduce the complexity and enhance the singing voice in music. One possibility to reduce the complexity of the accompaniment in music and enhance the vocals is to strengthen the noise reduction algorithm (NRA) in the CI sound coding strategy. Kim et al. (2020) showed that using a NRA did not improve music perception and sound quality and suggested disabling NR algorithms for hearing aid (HA) users. However, this was investigated using one instrumental and one vocal music piece and the specific effect of the NRA on singing music was not investigated. Nevertheless, Kim et al. (2020) also showed that a NRA improved music listening comfort. Kohlberg et al. (2016) reported that the use of a single channel NRA might attenuate constantly played components of a music piece and consequently increase music enjoyment in CI users. Kam et al. (2012) reported that around 50% of the CI users participating in their study found the NRA ClearVoice (Advanced Bionics, Valencia, USA) useful to improve the understanding of words in singing music. These findings lead to the consideration of NRAs as means to attenuate constantly played instruments of the accompaniment.

A second possibility to make music more accessible for CI users is to reduce the complexity of the electrical stimulation patterns transmitted by the CI by selecting fewer bands for stimulation. The so-called NofM band selection algorithm in the CI sound coding strategy selects N bands from M possible ones for stimulation. Selecting fewer bands for stimulation can be interpreted as a method to reduce complexity of the music transmitted by the CI. It has been shown that neglecting the least significant spectral components can be beneficial for CI users (e.g., Kludt et al., 2021; Nogueira et al., 2005). Moreover, selecting fewer bands for stimulation can enhance the VIR of singing music, at least when the vocals are higher in intensity than the background instruments, which is the case for the majority of Western pop music (e.g., Man et al., 2014).

Finally, the back-end compressor in the CI sound coding strategy may be used to optimize the EDR for singing music. Most compression systems used in CIs are designed and optimized for speech and in particular for speech in quiet (Langner et al., 2020). Therefore, designing a back-end compressor for music in the CI sound coding to improve music appreciation may be needed (Moore & Sęk, 2016). It is necessary to first characterize the music and then optimize the back-end compressor to provide a wider EDR to the CI user. It is widely accepted that music has a wider acoustic DR than speech. However, this applies to live acoustic music and not for commercially recorded music (Eargle, 2005). Unlike live acoustic music, commercially recorded music is normally compressed and tends to have a narrower DR (Kirchberger & Russo, 2016). Another factor that needs to be considered while designing a back-end compressor is the average intensity of music. Music is typically played and listened to higher intensity levels than speech (Chasin & Russo, 2004). This difference might be even greater when the average intensity of music is compared to speech in quiet, for which typically the CI compression systems are designed. While the typical level of music exceeds 80 dB SPL (Chasin, 2006), the average level of speech is considered to be 65 dB SPL. An additional factor is the standard used to calculate the DR. There exist different percentile recommendations for higher and lower dynamics in the DR calculation. Another factor that influences the DR is the music genre. Depending on the music genre, the DR of music pieces may differ drastically (Kirchberger & Russo, 2016). In this work, we aimed at designing a back-end compressor optimized for popular singing music containing background instruments that has narrower DR than, for instance, classical music.

Music perception can be measured by means of many different perceptual tests including timbre recognition (e.g., Gfeller et al., 2002; McDermott, 2004), melody identification (e.g., Galvin et al., 2007), and rhythm discrimination (e.g., Gfeller et al., 1997; Looi et al., 2008). Moreover, by utilizing subjective questionnaires, quality ratings for music (Looi et al., 2007), as well as paired- and multiple-comparison tests (Landsberger et al., 2020), the overall impression of the music piece and the intelligibility of the lyrics can be assessed. Crew et al. (2016) introduced the sung speech corpus as a tool to measure the ability of CI users to identify melodies conveyed by sung speech using a melodic contour identification (MCI) task. The present study takes a first step towards a tool to measure lyrics understanding. An adaptive matrix test based on the German Oldenburg sentence test (OLSA, Wagener et al., 1999a), has been adapted for measuring lyrics understanding in music, with background noise substituted with background instruments.

This study investigates the effect of reducing the number of selected bands in NofM, the effect of strengthening the NRA, and the effect of a novel music back-end multiband compressor on speech understanding in the presence of background instruments, MCI, the clarity of the singing voice, and singing music enjoyment in CI users. We tested five parameterizations of the CI sound coding strategy that differ in number of selected bands, strength of the NRA, the compression applied by the back-end compressor, and the combination of all previous manipulations. Moreover, we define two objective measures to optimize sound coding strategies for improving singing music perception of CI users. The first one is based on the electrical vocals-to-instruments ratio ($\mathrm{VIR_{el}}$) and the second one is based on the electrical vocals-to-instruments ratio enhancement ($\mathrm{VIR_{enh}}$), both in dB, by estimating the energy contribution of the vocals and the background instruments to the corresponding electrodograms. In addition, we created a speech recognition test with background instruments to simulate speech understanding of singing music and an MCI test based on singing voice with and without background instruments to assess the perception of the melody conveyed by the singing voice. Lastly, we evaluated the five parameterizations in ten CI users through three perceptual evaluation tests. These tests include the new designed speech recognition test with background instruments, the newly MCI tests using sung speech and background instruments, and a subjective appreciation and perception questionnaire. The last test was based on two aspects: the intelligibility of the lyrics and the overall impression of the music piece.

Methods

Subjects

Ten CI users of which six were bilateral and four were bimodal (CI in one ear and HA on the other ear) participated in the study. Table 1 presents their demographic data. At the time of testing, the participants were between 20 and 75 years old and had at least 9 months of experience listening with the CI in the tested ear. All participants were implanted with a Neuro Zti implant (Oticon Medical, Vallauris, France) and wore a NEURO 2 sound processor (Oticon Medical, Vallauris, France) using the Crystalis sound coding strategy. Bimodal CI users used a soft foam earplug in the ear canal of the untested acoustic hearing ear. Additionally, they covered the acoustic hearing ear with an over-ear earmuff to minimize acoustic leakage. Table 1 presents the pure-tone audiometry (PTA) hearing threshold of the subjects with residual hearing on the contralateral side. The PTA was calculated as mean of the audiometric thresholds at 500 Hz, 1000 Hz, 2000 Hz, and 3000 Hz (Carlson et al., 2017).

Crystalis Sound Coding Strategy

CI sound coding strategies are responsible for converting the acoustic signals captured by the microphone into the electrical stimulation patterns or electrodograms delivered by the CI electrodes. Figure 1 shows the block diagram of the Crystalis XDP sound coding strategy, which is a multiband spectral extraction strategy. Crystalis XDP is described as an NofM strategy because in each stimulation cycle, N out of M bands or channels are selected for stimulation. The Crystalis XDP incorporates a multiband back-end compression system instead of a conventional automatic gain control (AGC). Crystalis XDP, in contrast to other sound coding strategies, keeps the pulse amplitude fixed and modulates the pulse duration to code the sound level. The pulse duration can be adjusted in one microsecond [μs] steps from 10 to 120 μs.

NofM Channel Selection Algorithm

The Crystalis XDP sound coding strategy, processes and separates the sound signal using 20 bands. Following this, N out of the 20 frequency bands with the highest spectral density are selected for stimulation in each stimulation cycle. A related point to consider is that some Oticon Medical CI

Table 1. Subject Data With ID, Gender, Age at Testing for the Present Study, Cause of Deafness in the Implanted Ear, Gender, Tested Side, Cochlear Implant (CI) Experience, and the Existence of Residual Hearing on the Contralateral Side to the CI, Duration of Hearing Loss on the CI Side.

Subject ID	Age (years)	Cause of deafness	Gender	Tested side	CI experience on tested side	Residual hearing on contralateral side	Pure-tone audiometry (dB HL)	Duration of deafness (years)
P01	75	Unknown	M	R	5	–	–	5
P02	47	Unknown	F	L	1.5	–	–	2
P03	75	Unknown	M	L	2	+	64	5
P04	66	Unknown	M	R	2	+	64	9
P05	63	Meningitis	M	L	4	+	78	1
P06	71	Unknown	M	R	0.7	+	90	16
P07	59	Genetic	F	L	3	–	–	5
P08	40	Unknown	F	R	3	–	–	0
P09	20	MD	F	R	2	–	–	0
P10	55	ISSHL	F	L	1.5 –	–	.	5

MD, mitochondrial diseases; ISSHL, idiopathic sudden sensorineural hearing loss.

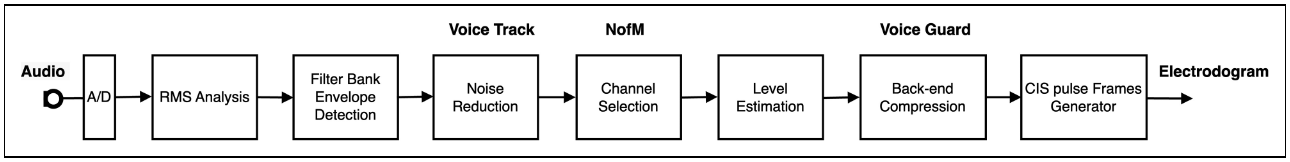

Figure 1. Signal processing chain of the Crystalis XDP sound coding strategy in Oticon Medical cochlear implants.

users have one or more electrodes deactivated. For those CI users, the sound signal will be processed not into 20 bands but into fewer bands identical to the number of active electrodes.

Voice Track Noise Reduction Algorithm

NRAs are widely used in sound coding strategies. In the Crystalis strategy, a modified version of a single microphone NRA based on the Wiener filter is used. This algorithm is called Voice Track (VT) and has four different settings (off, soft, medium, and strong). VT operates in 20 independent frequency bands with a 2-s attack time and an 18-ms release time aiming at reducing the background noise (Guevara et al., 2016). The output signal of this algorithm G_{nc} is obtained by applying a gain $\mathrm{G_w}$ to the input signal, that is,

$$G_{nc}(t, m) = 1 - G_w \cdot (R_i(t, m) + R_{mean}(t, m)), \qquad (1)$$

with t and m being the time frame and the frequency band, respectively. The instantaneous noise-to-signal ratio, $\mathrm{R_i}$, and the average noise-to-signal ratio, $\mathrm{R_{mean}}$, are defined as

$$R_i(t, m) = \frac{P_{noise}(t, m)}{P(t, m)}, \qquad (2)$$

$$R_{mean}(t, m) = \frac{P_{noise}(t, m)}{P_{mean}(t, m)}, \qquad (3)$$

where P_{noise}, P, P_{mean} denote the instantaneous power of the noise, instantaneous power of the signal, and the average power of the signal, respectively. Strengthening the VT settings will increase the gain $\mathrm{G_w}$ applied to the input signal where an attenuation coefficient of 0, 0.35, 0.5, and 0.75, respectively, will be applied to each frequency band.

Voice Guard Back-End Compressor

The back-end compressor (Langer et al., 2020) in the Crystalis XDP sound coding strategy called Voice Guard (VG) consists of four-bands with fixed knee points where the compression function of each band can be separately adjusted. Input dynamic range (IDR) refers to sound intensities between the softest and loudest sounds that can be captured by the microphone for further processing and it is by default set to 70 dB. The IDR is compressed and mapped to the EDR of each CI user, which is the difference between the largest and smallest pulse duration (the most comfortable and threshold level) of the stimulation pulse in dB units. VG operates as a transfer function where an interval of 70 dB (from 23 dB HL to 93 dB HL) is mapped to an EDR that is subject-dependent. For most of the CI users, the EDR varies between $10 - 20$ dB (Skinner et al., 1997; Zeng & Galvin, 1999). The compression is performed by dividing the IDR (x-axis) using a knee-point into two intervals and the EDR (y-axis) using a middle point that we call M-point into two intervals (Figure 2). Finally, the two intervals on the IDR are mapped to the two corresponding intervals on the EDR. The M-point is set to 75% of the EDR of each CI user. Theoretically, the number of knee points and M-points can be increased, consequently having more than two intervals to be mapped.

Auto Voice Guard. In the currently used clinical settings, an adaptive version of VG is used, hence the sound coding strategy is called Crystalis coordinated adaptive processing (CAP). Crystalis CAP switches automatically between predefined knee points based on the RMS input level. The predefined knee points are the same for Crystalis CAP and for Crystalis XDP. Crystalis CAP differs from Crystalis XDP only in that the former uses an adaptive version of VG called Auto VG. This compression system has been designed and optimized to maximize the DR of speech without background noise and in return improve speech understanding in the environment where no noise is present. Auto VG adjust automatically the knee points based on the acoustic environment. Table 2 shows the knee points for three predefined environments in the in Crystalis CAP and corresponding input loudness levels. These values have been set to maximize EDR having speech in quiet as input of the sound coding strategy aiming at maximizing speech understanding in quiet background environments. For this purpose, a large database of speech signals from western languages was used (Segovia-Martinez et al., 2016).

Music Voice Guard. In this work, we present a new back-end compressor that we term Music VG. It has been designed to maximize the EDR of music signals instead of maximizing the EDR of speech signals as it is done in the Crystalis XDP sound coding strategy. Similar to Auto VG, Music VG is a four band back-end compressor with one M-point. In contrast to Auto VG, it uses two knee points for compression. A histogram analysis of a large music dataset called MUSDB (Rafii et al., 2017) consisting of popular music pieces from different genres was used to set the knee points of the Music VG. The lower knee point of each band is mapped to the M-point that corresponds with 20% of the EDR of the CI users and the upper knee point is mapped to the C level. This is the comfort level of the stimulation charge and its CI user dependent. This compressor has been optimized for an input level of 75 dB SPL. It is worthwhile to mention that using the volume buttons on the NEURO 2 sound processor applies a gain to the input. This in return modifies the presented input level to the backend compressor (a shift on the x-axis; see Figure 2). Although the knee points are not affected by adjustment of the physical buttons, the back-end compressor is. Therefore, a 6 dB margin was introduced to the knee points of Music VG, anticipating the volume changes by the CI user and compensating its effect on the music VG Table 3.

Objective Measurement

A MATLAB (Mathworks, Natick, MA, USA) based environment called Oticon Medical Cochlear Implant (OMCI) simulation chain was used to simulate the electrodograms of the Crystalis CAP sound coding strategy. Objective measures based on the vocals-to-background instruments ratio (VIR_{el}) were estimated from the energy of the electrodogram of each signal. Moreover, percentile analysis of was used to quantize the DR with different back-end compressors. In the objective measures, a virtual CI user with 20 and 100 µs T and C levels, respectively for all electrodes was simulated. These C and T levels correspond to a 14 dB EDR.

Music Pieces used for the Objective Measures

100 music pieces from the MUSDB and the iKala (Chan et al., 2015) datasets were used to obtain the objective measures. Twenty-second excerpts of each music piece in which ~~the~~ vocals were present were used. The iKala dataset contains 30-s 2-channel music tracks of singing music with

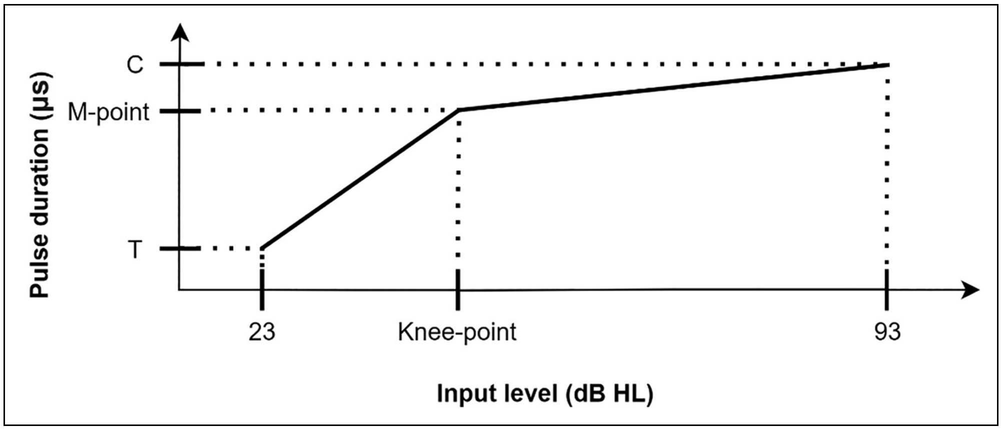

Figure 2. Back-end compression system in the Crystalis sound coding strategy.

Table 2. Knee Points for the Backend Compressor Auto Voice Guard (VG) for Three Predefined Sound Environments. The Values of the Knee Points Are Expressed in Decibel Hearing Level (dB HL). The Corresponding RMS Input Level for Quiet, Medium and Loud Environments Are 60 dB SPL, 70 dB SPL, and 80 dB SPL, respectively.

	Environment		
Band	Quiet	Medium	Loud
Band 1 (195 – 846 Hz)	52 dB HL	61 dB HL	70 dB HL
Band 2 (846 – 1497 Hz)	52 dB HL	61 dB HL	70 dB HL
Band 3 (1497 – 3451 Hz)	47 dB HL	57 dB HL	66 dB HL
Band 4 (3451 – 8000 [Hz])	41	50	58

Table 3. Knee Points for the New Back-End Compressor Music VG. The Values Are Expressed in dB HL.

Band	Lower knee point (dB HL)	Upper knee point (dB HL)
Band 1 (195 – 846 [Hz])	42	68
Band 2 (846 – 1497 [Hz])	37	67
Band 3 (1497 – 3451 [Hz])	36	67
Band 4 (3451 – 8000 [Hz])	30	61

one channel for the singing voice and the other for background music. All music tracks were performed by professional musicians and six singers, of which three were female and three male. The MUSDB dataset consists of popular and commercially available music pieces from different genres that are professionally mixed. Each includes four stereo sources (bass, drums, vocals and a group of other instruments).

Electrical Vocals-to-Instruments Ratio

The VIR$_{\mathrm{el}}$ used in this study is based on the electrical signal-to-noise ratio used in Nogueira et al. (2016). The VIR using electrodograms was used to calculate VIR$_{\mathrm{el}}$ for various parameterizations. To calculate VIR$_{\mathrm{el}}$, the singing voice and the background instruments were processed through the OMCI (Figure 3) and the energy of electrodograms in dB was used to obtain the corresponding VIR$_{\mathrm{el}}$. Equations (4) and (5) with m[n] being the singing music sound signal and v[n] the vocals sound signal, i[n] instruments sound signal and n the samples denote the VIR$_{\mathrm{ac}}$ calculation. Equation (6) with E$_{\mathrm{V}}$ and E$_{\mathrm{I}}$ being the electrodograms for the vocals and the instruments, the time frame, and m the electrode in the electrodogram denote the VIR$_{\mathrm{el}}$ calculation. The equation assumes linearity in the signal processing algorithms used in the CI sound coding strategy. However, this assumption is not valid for two signal processing stages. Hence, during the processes using the OMCI simulation chain, the non-linearity of the two signal processing algorithms namely NofM and VG in the sound coding strategy was taken into consideration. In order to do that, the mixture audio was first processed and the corresponding bands in NofM algorithm and knee points of VG back-end compressor were stored. Second, while processing the singing voice and background instruments, the stored bands and knee points of the mixture were used.

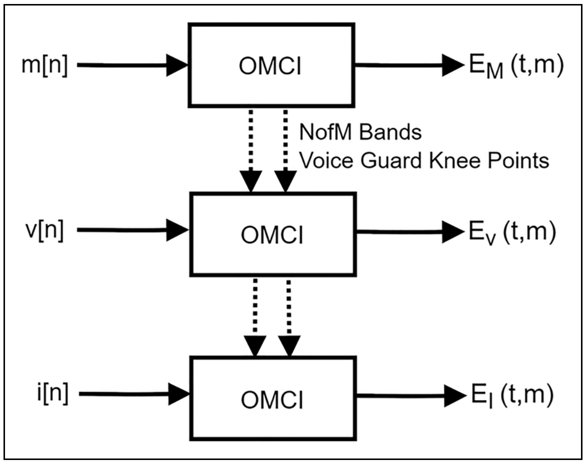

Figure 3. A visualization of the methods used to calculate VIR$_{\mathrm{el}}$ in objective measures. In the figure, m[n] and v[n] and i[n] denote the mixture, singing music, and the instruments' sound signal with n representing the samples. E$_{\mathrm{M}}$, E$_{\mathrm{V}}$ and E$_{\mathrm{I}}$ denote the electrodograms for the mixture, vocals, and instruments, respectively.

$$m[n] = v[n] + i[n] \tag{4}$$

$$VIR_{ac}[dB] = 10 \cdot \log_{10}\left[\frac{\sum_{n=1}^{N} v[n]^2}{\sum_{n=1}^{N} i[n]^2}\right] \tag{5}$$

$$VIR_{ac}[dB] = 10 \cdot \log_{10}\left[\frac{\sum_{t=1}^{T}\sum_{m=1}^{M} E_V(t,m)^2}{\sum_{t=1}^{T}\sum_{m=1}^{M} E_I(t,m)^2}\right] \tag{6}$$

$$VIR_{enh}[dB] = VIR_{el}(Baseline) - VIR_{el}(Parametrization) \tag{7}$$

Measure of Enhancement

We calculate and report VIR$_{\mathrm{enh}}$ as an evaluation metric for vocals enhancement within two parameterizations shown in Equation (7). The amount of electrical enhancement is the arithmetical difference between the corresponding VIR$_{\mathrm{el}}$ of each parameterization. VIR$_{\mathrm{enh}}$ was used to assess the enhancement applied to the singing voice with the respect to the background instruments in music.

Percentile Analysis

Percentile analysis is a useful tool to characterize the dynamics of a certain quantity and has been used to characterize the DR of hearing aid and CI compressors (Kirchberger & Russo, 2016; Langer et al., 2020; Ma et al. 2015; Moore, 2008). Values in a set of samples that are greater than a specific percentage of all samples are called percentiles and are used in DR calculation. We used percentile analysis to investigate the effect of the back-end compressor on the EDR of singing music as input of the sound coding strategy. Accordingly, the OMCI simulation chain was used to simulate the electrodograms of music pieces mentioned in Music Pieces used for Objective Measures section. Amplitude percentiles of the electrodogram for Auto VG and music VG were calculated for further analysis. In this analysis, 33th, 66th, and 99th amplitude percentiles of the electrodograms were calculated as recommended by IEC 60118-15(2012) that were developed to characterize HA signal processing algorithms. The DR of each back-end compressor is defined as the difference between the 99th and 33th percentile. Here, the percentile indicates the value that is greater than a specific percentage of all the samples in a dataset (e.g., 33th percentile is the value that is greater than 33% of all samples).

Parameterization of the Sound Coding Strategy

Based on the results obtained from the objective measures (see section Results Objective Measures) five parameterizations were defined to be investigated using a perceptual evaluation.

Baseline: The typical clinical parameterizations of the Crystalis CAP used by Oticon Medical CI users were defined as the baseline for comparison with other parameterizations of the sound coding strategy. In this parameterization, N was set to 8, NRA was set to soft and VG was set to auto.

6of20: Results obtained from $\mathrm{VIR}_{\mathrm{enh}}$ showed that singing voice can be enhanced by reducing N in NofM algorithm. This parameterization differs from the baseline reducing N in the NofM band selection from 8 to 6.

NRA Medium: Results from $\mathrm{VIR}_{\mathrm{enh}}$ objective measure showed that strengthening the NRA setting enhances the singing voice as observed by an increase in the $\mathrm{VIR}_{\mathrm{el}}$. This parameterization differs from the baseline in that the NRA setting was strengthened from soft to medium.

Music VG: Based on the percentile analysis objective measure, the newly designed Music VG improves the EDR of music signals. This parameterization differs from the baseline in that it uses Music VG instead of Auto VG.

Combination: Based on the combination of 6of20, NRA Medium, and Music VG used to not only enhance the $\mathrm{VIR}_{\mathrm{el}}$ but also improve the EDR of music.

Perceptual Experiments

This study consisted of three perceptual experiments. In the first experiment, speech understanding with background instruments was assessed. In the second experiment, melody identification of sung music with and without background instruments was measured. In the last perceptual experiment, music appreciation was assessed using a questionnaire that was completed by the participants. The perceptual evaluations were performed in a double-walled sound-treated room using an active (self-amplified) loudspeaker (Genelec 8090B, Helsinki, Finland). The tests were performed using two NEURO2 sound processors, since on each speech processor only four parameterizations can be stored. On the first sound processor, the Baseline, 6of20, and NRA medium parameterizations were stored and on the second sound processor, Music VG, and the combined parameterization (6of20, NRA Medium, and Music VG) were stored to be tested in CI users.

Speech Understanding Test

The OLSA test was used to assess the speech understanding of CI users in the presence of background instruments. The test is a very simple model of singing voice understanding with accompaniment. The test was calibrated at 75 dB SPL and for each condition, two lists each consisting of 20 sentences were used. During the speech understanding test, a continuous background instrument of a music piece from the iKala dataset was played to keep the NRA activated.

Melodic Contour Identification

In the second experiment, the ability of CI users to perceive melody contours of the singing voice was assessed using an adapted version of the MCI test by Galvin et al. (2007). Figure 3 depicts the melodic contour patterns used in this test. Five patterns with three notes were used to assess the MCI of CI users. The melodic contour patterns designed by Galvin et al. (2007) consisted of nine possible contours with five notes each. However, it has been shown this number of patterns can be very challenging and demanding for CI users (Omran et al., 2010; Tabibi et al., 2016). Some studies have used a simplification of the MCI test with five melodic contours created with five notes (Tabibi et al., 2016). All notes in the melodic contour patterns were from a sung speech corpus that was created for this study and was sung by a trained female singer. The sung speech corpus contains monosyllabic words from the Freiburger monosyllabic speech test (Hahlbrock, 1953). The fundamental frequency of the mid-note in all contours was Bb3 (233

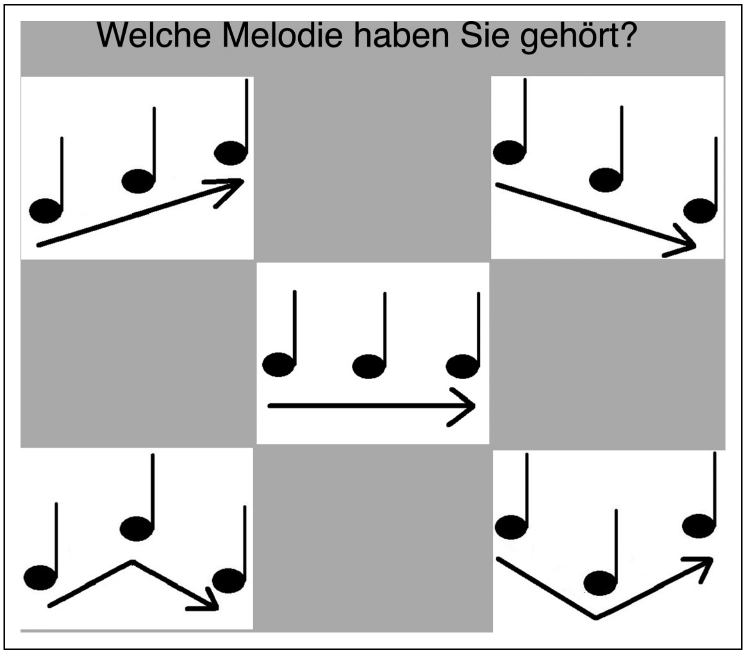

Figure 4. Melodic contour patterns used in melodic contour identification (MCI) test. In German: "Welche Meldodie haben Sie gehört?" translates to "Which melody did you hear?" in English.

Hz). Six conditions were tested. In the first and second conditions, four semitone spacing between notes using a single word (the same word in all three notes e.g., "Maus – Maus - Maus" German word for mouse) with and without background instruments respectively was used. Kasturi and Loizou (2007) showed a significant effect of semitone-spacing on the MCI test. Hence in the third and fourth conditions, we used two-semitone spacing between notes in the melodic contours. In the fifth and sixth conditions, multiple words (different words in each possible melodic contour e.g., "Bett - Kamm – Milch," the German words for bed, comb, and milk) with two semitones spacing with and without background instruments respectively were used. Each condition consisted of 15 melodic contours and was presented at 75 dB SPL (Figure 4).

Table 4. Summary of the Conditions Used in the Melodic Contour Identification (MCI) Test.

MCI test condition	Semitone spacing	Background instruments	Timbre
1	4	No	Single word
2	4	Yes	Single word
3	2	No	Single word
4	2	Yes	Single word
5	2	No	Multiple words
6	2	Yes	Multiple words

Each melodic contour was presented once and the subjects were asked to choose one of the five possible options. In the conditions where background instruments were presented, the intensity level of the background instruments and the notes were at the same level resulting in a VIRac level of zero. The intensity level of the background instruments was for two CI subjects (P1 and P3) 10 dB less than the notes resulting in a VIRac level of 10 dB. Prior to the test, the subjects conducted three training rounds where they received feedback on whether they could identify the correct melodic pattern or not. The MCI test was performed using a Samsung (Samsung Digital City, Maetan-dong, Yeongtong District, Suwon, South Korea) Tab S6 lite tablet. The tablet was connected to the loudspeaker through a sound card with 1-m distance between the listener and the loudspeaker. The tablet application used in the MCI test will be available on the Google Play store (Google Inc., Mountain View, California, USA) and can be used for training/rehabilitation purposes (Table 4).

Music Perception and Appreciation Questionnaire

All participants were asked to complete a questionnaire to assess the clarity of the singing voice in the music piece and their music enjoyment with each sound coding program/parameterization. A 10-s excerpt of three popular music tracks kindly shared by Wim Buyens (Buyens et al., 2014) shown in Table 5 was used in this test. Each music piece was played once prior to each question which was then ranked by the participant on a 15 steps Likert scale adapted from Nogueira et al. (2015) where 1 and 15 were the least and the highest value in the ranking, respectively.

Statistical Analysis

The normality of the distribution of the results gathered from the perceptual evaluation was tested with the Shapiro-Wilk test. A follow-up repeated measures analysis of variance (ANOVA) was performed for normally distributed data and a non-parametric test, Friedman rank sum test, for non-normally distributed data to assess whether there is a significant difference among the mean results obtained from the five parameterizations (Baseline, 6of20, NRA Medium, music VG, Combination). An ANOVA post-hoc test for normally distributed data and a post-hoc test through multiple Wilcoxon tests for non-normally distributed data with Bonferroni correction was conducted to assess the significance of differences between pairs of group means. In all tests, we rejected the null hypothesis and considered significant results when $p < .05$ (5%). Moreover, all p-values reported in this study have been adjusted using the Bonferroni method.

Results

Objective Measures

Electrical Vocals-to-Instruments Ratio. Figure 5A presents the effect of reducing the number of selected bands N on VIR_{el} [dB] while keeping the NRA fixed to soft. The effect is reported as a function of the VIR_{ac}, which is the vocals-to-instruments ratio at the input of the speech processor, that is, the VIR_{ac} of the original music mix. Reducing N from 10 to six increased the VIR_{el} [dB]. Figure 5B presents the effect of strengthening the NRA. Strengthening NRA from soft to medium, improved the VIR_{el} by around 0.25 dB and strengthening NRA from medium to strong improved the VIR_{el} by around 0.15 dB. Moreover, setting NRA to strong provided a higher VIR_{el} [dB] than setting NRA to medium for negative VIR_{ac} [dB]. However, both of these NRA settings provided the same VIR_{el} [dB] at 4 dB VIR_{ac} [dB]. It is likely that both of these settings applied the same amount of suppression to the singing voice and the background instruments at 4 dB VIR_{ac} [dB].

Table 5. Popular Music Tracks Used in Music Appreciation Test.

ID	Song name	Vocals	Piano	Guitar	Bass	Drum
Hal	Hallelujah (Leonard Cohen)	+	+	−	−	−
Bef	Before I Go (Papermouth)	+	+	+	−	−
Mic2	Michel (Anouk)	+	−	+	+	−

Measure of Enhancement. Figure 6 presents the VIR_{enh} [dB] results, taking the baseline parameterization as reference (see Equation 7), the red line and the hexagram in the boxes represent the median and the mean, respectively. The third parameterization where the N in NofM was reduced from 8 to 6 and the NRA setting was strengthened from soft to medium provided the maximum amount of enhancement with respect to the baseline parameterization with a mean VIR_{enh} of around 0.2 dB. On a related note, the compression applied by the Music VG to the signing voice and to the instruments was the same as that by the Auto VG resulting in the same results in VIR_{el} hence to VIR_{enh} difference.

Percentile Analysis. Figure 7 shows the percentile analysis of VG Auto and Music VG using the electrodograms of 100 music pieces of the MUSDB dataset. The black, blue, and red lines indicate the 33th, 66th, and 99th percentiles, respectively. The upper boundary of the DR (99th percentile) across

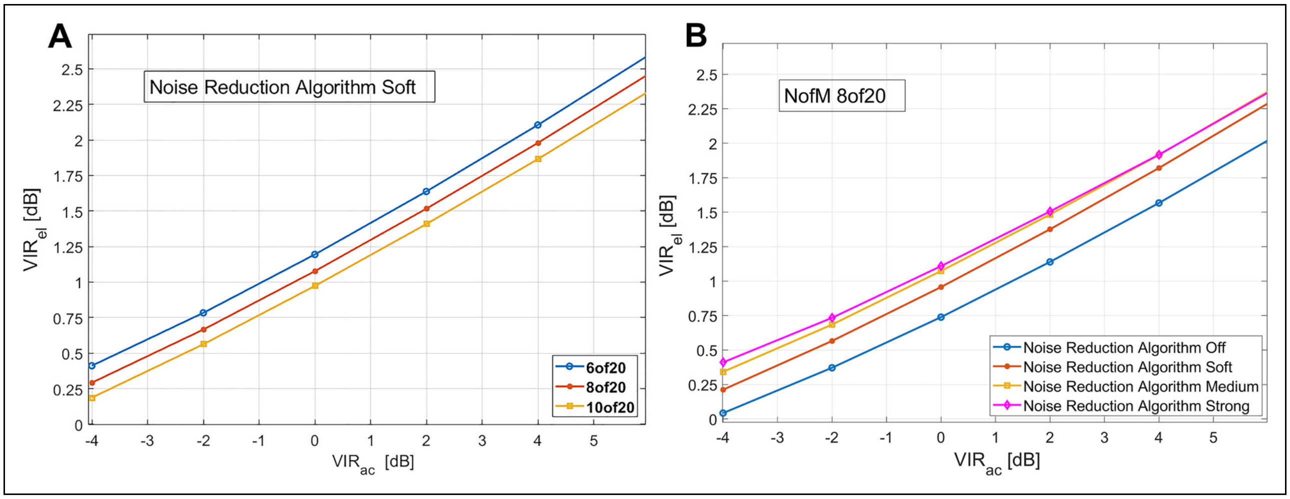

Figure 5. Results of the objective measure electrical vocals-to-instruments ratio VIR_{el} [db] as a function of the vocals-to-instruments ratio acoustic (VIR_{ac}) of the music material entering the cochlear implant sound processor for: (A) NofM band selection algorithm with noise reduction algorithm (NRA) set to soft. (B) NRA with NofM set 8of20.

all electrodes and the lower boundary of the DR (33th percentile) for electrodes 10 to 20 were improved using the new Music VG.

Perceptual Experiments

Speech Understanding Test. Figure 8 shows the results obtained from the speech understating test for different parameterizations. The Music VG parameterization shows the best mean SRT of around 1.5 dB. A repeated measures ANOVA revealed a significant effect of parameterization on the SRTs ($F(4, 36) = 4.90$, $p = .003$). Therefore, follow-up post hoc Bonferroni-corrected pairwise comparisons were conducted and showed that NofM, Music VG were significantly different than the baseline parameterization ($p =$ .022, $p = .001$, respectively). There were no significant differences between other pairwise comparisons.

Melodic Contour Identification. Figure 9 shows the results for the six MCI test conditions (see Table 4) performed with the different parameterizations in CI users. Here, the effect of reducing the number of selected bands N in NofM, the effect of strengthening the NRA and the effect of the new music VG on the perception of melody contours of the singing voice was assessed using an adapted version of the MCI test. In Figure 9, the left panel (subfigures 9 A, C, and E) presents the results for the MCI tests without background instruments and the right panel (subfigures 9 B, D, and F) show the results for the MCI tests with background instruments. In the MCI tests without background instruments (subfigures 9 A and C), the 6of20 and NRA medium parameterization did not significantly improve nor worsen the results with respect to the baseline (p > .05 after Bonferroni correction). In the single word conditions (subfigures 9 A, B, C, and D), no significant difference was detected between mean values obtained from the tested parameterizations and the bassline parameterization after applying Bonferroni correction to the p values. The mean value of the results obtained from the MCI tests with multiple words (subfigures 9 E and F) were close to the chance level

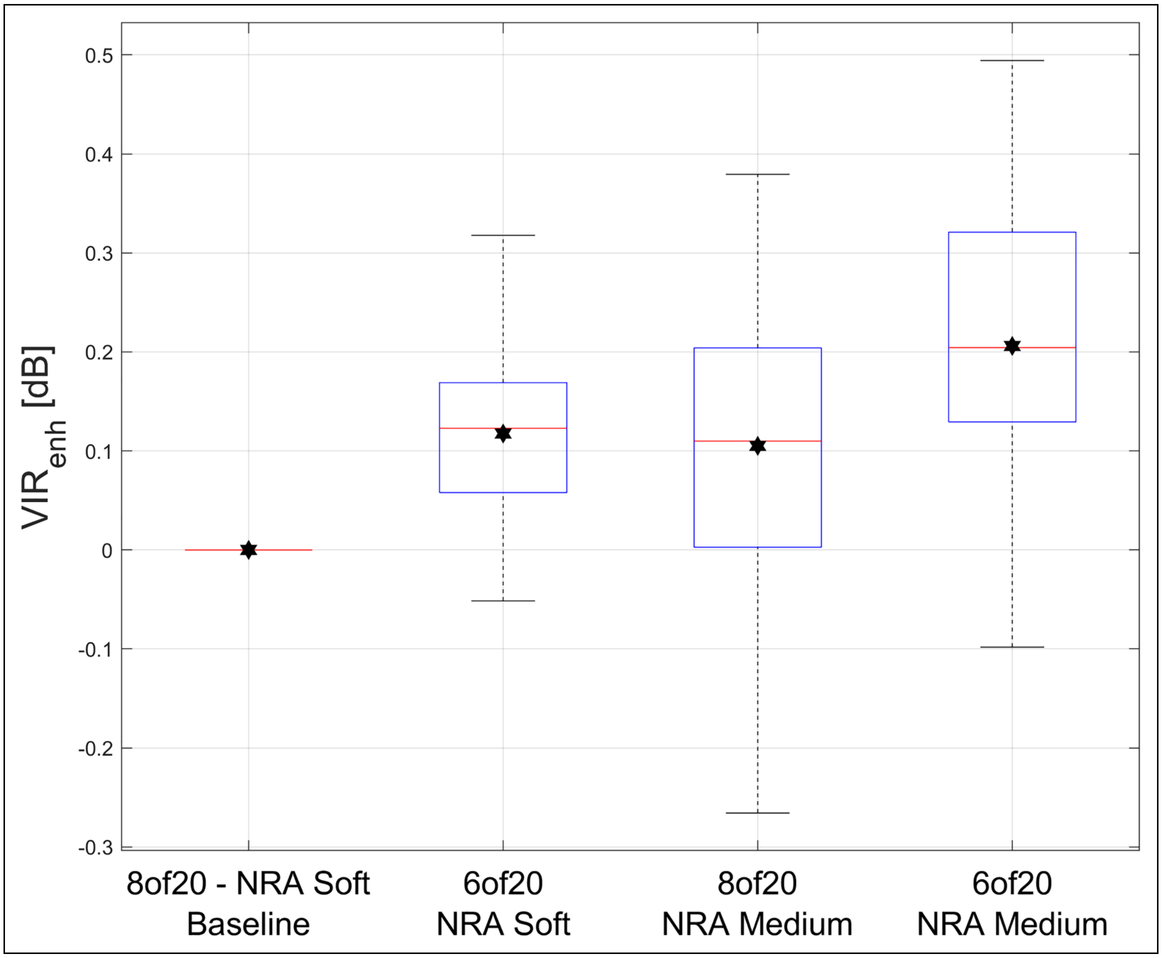

Figure 6. Measure of enhancement VIR$_{enh}$ [db] defined in Equation 7 as the difference in VIR$_{el}$ [db] obtained with the baseline parameterization (8of20 NRA soft) and a test parameterization (6of20 NRA soft, 8of20 NRA medium, and 6of20 NRA medium). The boxes present 25th and 75th percentiles. The red line and hexagram indicate median and mean values, respectively.

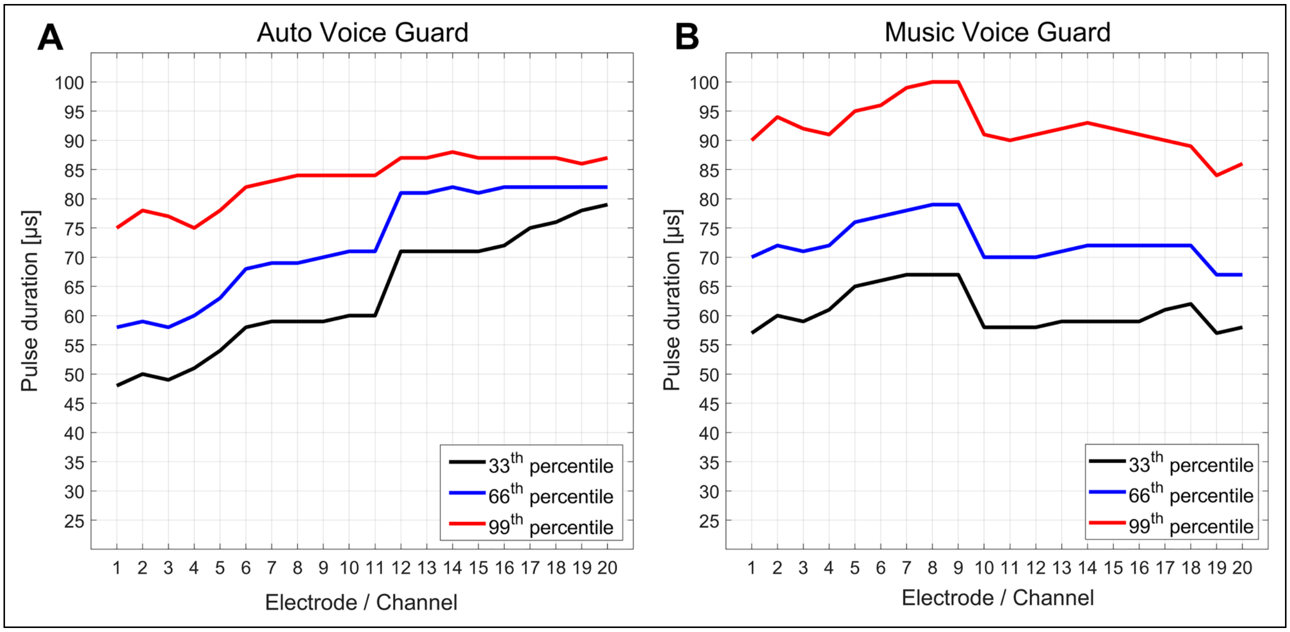

Figure 7. Percentile analysis for the 33th (black), 66th (blue), and 99th (red) percentiles for Auto Voice Guard (A) and Music Voice Guard (B).

and there was no significant difference between tested parameterizations and the baseline parameterization.

A Wilcoxon test revealed that adding background instruments and decreasing the number of semitone spacing from four to two did significantly worsen the MCI results with $p < .001$ for both Wilcoxon test comparisons. Moreover, the results obtained with MCIs based on multiple words (subfigures 9 E and F) were also significantly worse than the ones obtained with single word (subfigures 9 A, B, C, and D) MCI test, as revealed by another Wilcoxon test ($p < .001$).

Music Perception and Appreciation Questionnaire. Figure 10 (a) and (b) present the results from the music appreciation and perception questionnaire filled out by CI users. In the first question, the subjects were asked about the intelligibility of the lyrics, and in the second question about the overall impression of the music piece. The participants were asked to fill out a questionnaire to assess the clarity of the singing voice in the music piece and their music enjoyment with each parameterization. All tested parameterizations provided a better score with respect to the baseline parameterization in both "clarity of the singing voice" and "overall music enjoyment." Both the NRA and Music VG parameterization provided the highest scores with around 2 score points of improvement with respect to the baseline parameterization in either perceptual questionnaire. For the first question, we employed the non-parametric test of Friedman; there was no significant difference among the parameterizations. For the second question, we employed a repeated measures ANOVA F(4,36) = 3.18, $p = .024$, and a follow-up post hoc Bonferroni-corrected pairwise comparison between tested parameterizations. The Music VG parameterization was significantly better than the baseline parameterization ($p = .025$). There were no significant differences between other pairwise comparisons. Nevertheless, all parameterizations improved the scores in both questionnaires with respect to the baseline parameterization.

Discussion

In this work, we investigated the optimization of a sound coding strategy to improve music perception and appreciation for CI users. We hypothesized that reducing the number of selected bands for stimulation and strengthening the NRA can reduce the complexity of the music signal and attenuate the background instruments increasing the VIR_{el}. Furthermore, we investigated the optimization of a back-end compressor in the CI sound coding strategy. Specifically, a new back-end compressor for the Crystalis sound coding strategy has been designed and optimized to increase the EDR for singing music. The effects of the novel optimizations for music perception were evaluated through newly designed tests tailored to measure and assess two aspects of singing music, namely the speech understanding and the melody conveyed by the lyrics.

Ten CI users, of which six were bilateral and four bimodal, participated in the perceptual evaluations. In the first experiment, we used the OLSA speech understanding test in the presence of background instruments as a simple model to measure the understanding of lyrics in a singing voice. We assumed that if a new algorithm improves speech understanding in background music, it would

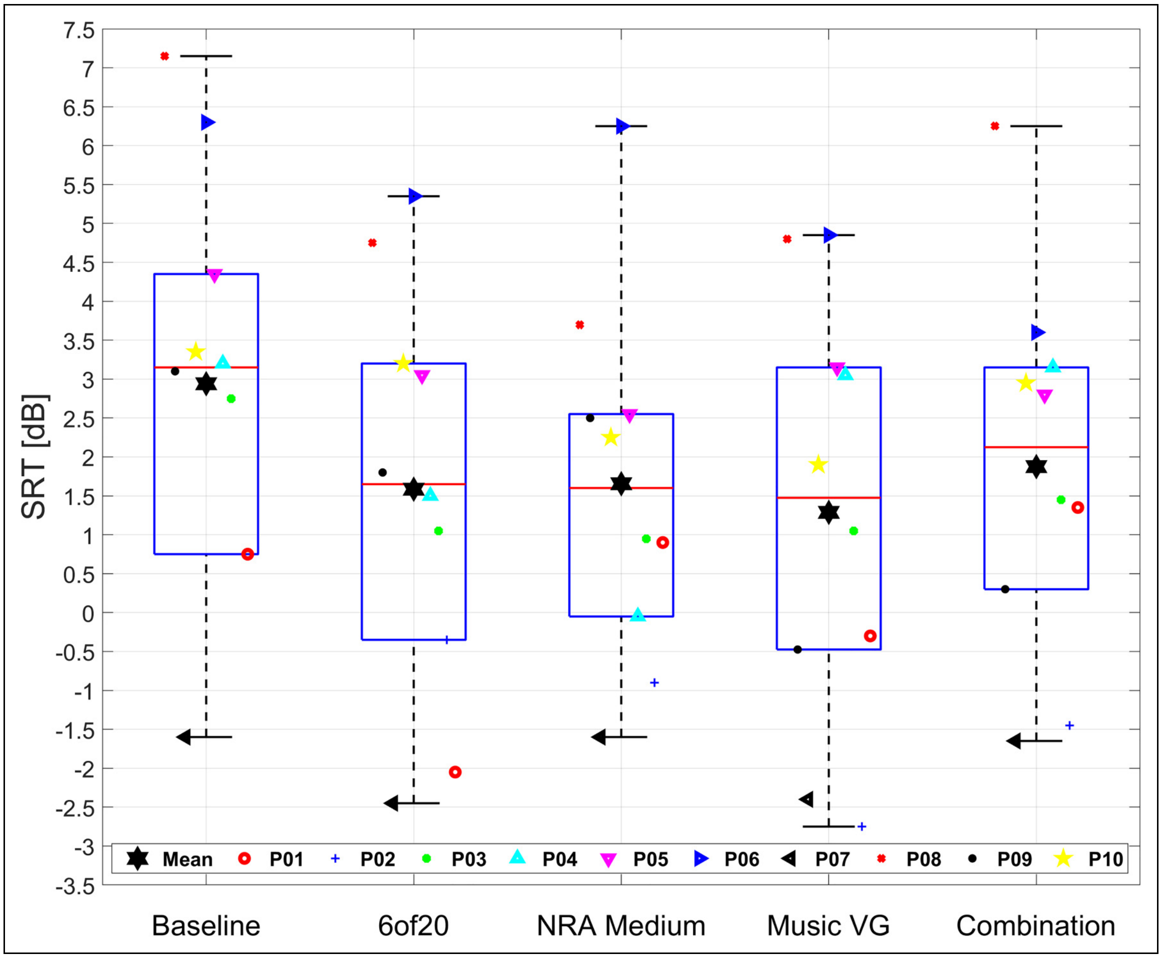

Figure 8. Individual and averaged speech reception thresholds (SRTs) across subjects. The boxes present 25th and 75th percentiles. The red line and the filled black hexagram indicate median and mean values, respectively. Individual results are presented with colored symbols.

improve lyrics understanding in music. However, for future studies, we suggest further developing new evaluation methods using natural singing voice in music rather than regular speech in music to measure the effect of new algorithms on lyrics understanding specifically. By comparison, many studies have used questionnaires (Nogueira et al., 2015, 2019) to assess lyrics understanding. However, these tests are highly subjective causing inter and intrasubject variability and consequently leading to a lack of accuracy to compare different CI sound coding strategies. In the second and third experiments, the music perception of the participants was assessed by means of a modified version of the MCI test (Crew et al., 2016; Galvin et al., 2009) and by utilizing questionnaires. In the MCI test, we used a speech corpus sung by a trained singer, where the melody in the contours is conveyed by sung speech. Moreover, we expanded the sung version of the MCI test by Crew et al. (2016) in that we added background instruments to simulate singing music.

The results from the speech understanding test show positive mean SRTs. In this test, we used ~~a~~ spoken speech with background instruments to simulate lyrics understanding. The positive SRTs indicate that the lyrics must be louder than the background instruments to be understood by CI users emphasizing the importance of enhancing the lyrics for CI use. Since a large variability in the MCI performance of CI users was expected, we performed the experiment in six conditions with different degrees of difficulty. The six conditions differ from each other in the amount of semitone spacing, the inclusion of background instruments, and the use of different words instead of a single word within one melody contour. As predicted, individual results show huge variability in the performance in the MCI test. Consistent with previous studies, decreasing the number of semitone spacing between the notes in the contours worsened the MCI performance of CI users significantly ($p < .001$). Similarly, mean scores obtained from the multiple-word MCI condition were statistically significantly worse than the mean MCI score obtained from the single-word condition ($p < .001$). These outcomes are consistent with the results of Crew et al. (2016) study. Finally, adding background

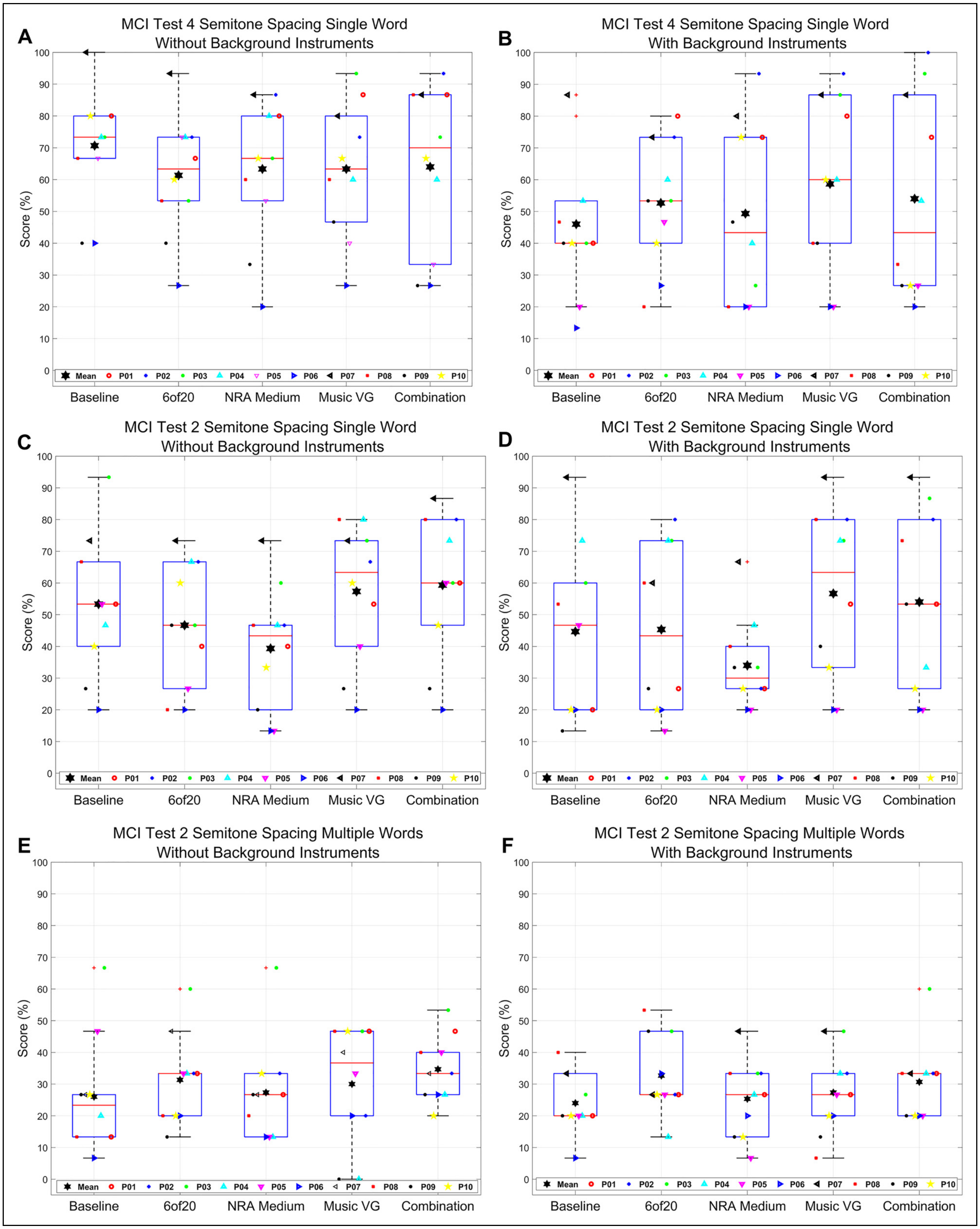

Figure 9. Box plot for MCI score. Each panel presents the results for different test conditions. The boxes present 25th and 75th percentiles. The red line and the filled black hexagram indicate median and mean values, respectively. Individual results are presented with colored symbols.

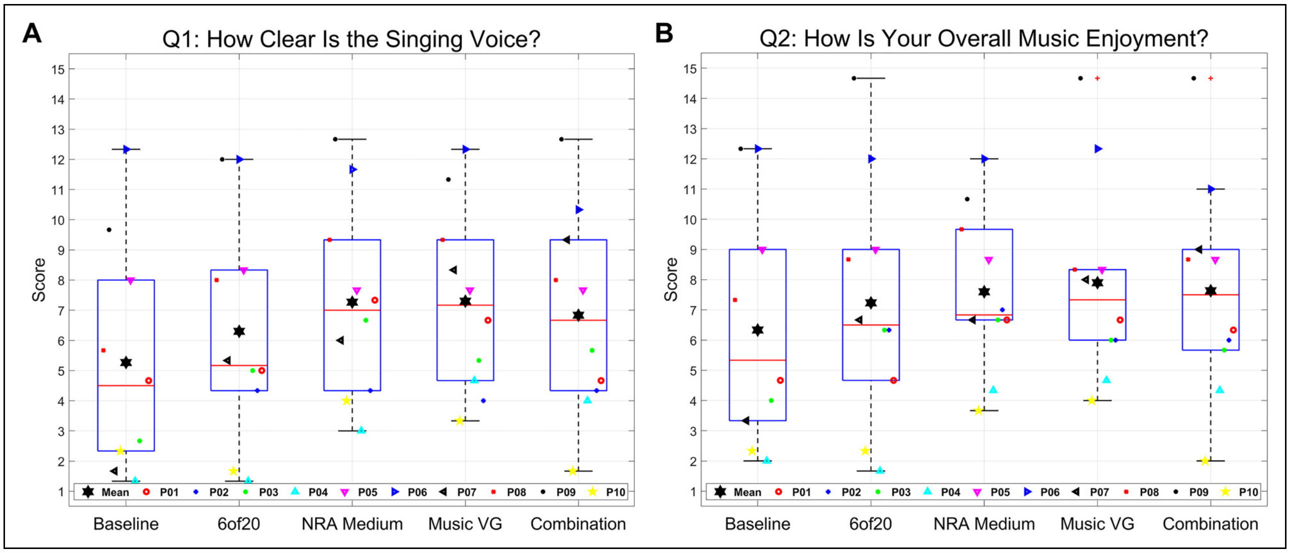

Figure 10. Results of the music perception and appreciation questionnaire. The left panel presents the results for the question "How clear is the signing voice?". The right panel presents the results for the question "How is your overall Music Enjoyment?". The red line and the filled black hexagram indicate median and mean values, respectively. Individual results are presented with colored symbols.

instruments to the melodic contours provided statistically significantly worse results in the MCI test ($p < .001$). The MCI performance of two bimodal CI subjects (P5 and P6) differed drastically from the mean results; Crew et al. (2016) showed that the MCI performance of bimodal CI users is largely driven by the HA. Since the MCI test was conducted only with the CI, the lack of HA cues might have led to poorer performance in the MCI test for two bimodal subjects (P5 and P6). Similar to the MCI results, there was large intersubject variability in the results gathered from both questionnaires. It is worth mentioning that the two CI users that obtained poor performance in the MCI test (P05 and P06), gave high ratings on both questioners above the mean values. These results show that even though these two subjects faced difficulties perceiving pitch direction, they found the singing voice in them clear and could enjoy the music pieces. These results are consistent with Looi et al. (2012) who showed that CI users can enjoy music despite the lack of ability to perceive melodic pitch. In addition, Limb et al. (2010) reported a near-normal rhythm discrimination of CI users and Innes-Brown et al. (2012) reported that CI users rely on vocals and rhythm cues to enjoy music.

Motivated by previous studies (Gajecki & Nogueira, 2018; Pons et al., 2016; Tahmasebi et al., 2020) that used front-end processing algorithms such as DNNs and showed that CI users prefer an acoustic VIR enhancement of 6 dB, this study showed that an NRA could enhance the singing voice. Results from objective measures show that adjusting the NRA setting from off to medium enhances the vocals by around 0.6 dB electrically. It is worth mentioning that due to the compression applied by the CI sound coding strategy, 6 dB of acoustic enhancement on the singing voice will correspond to a much smaller enhancement of around 1.25 dB on the electrical domain. Results obtained from the speech understanding test with background instruments show that strengthening the NRA in the Crystalis sound coding strategy from soft to medium improved the mean SRT in music by around 1.5 dB. These results are consistent with the results obtained from the objective measures, where strengthening the NRA from soft to medium improved $\mathrm{VIR}_{\mathrm{el}}$. In the MCI test, NRA did not significantly worsen nor improve the melody perception of signing voice in CI users with or without background instruments. Even though there were no statistically significant differences in the mean MCI scores obtained from the two NRA settings, the mean scores obtained from the NRA set to medium were slightly worse than the NRA set to soft. This difference is probably because of distortions introduced to the target singing voice melody by NRA. These results are consistent with the results of Kim et al. (2020), where a disabled NRA was suggested to improve pitch perception in HA users. Moreover, it is likely that setting the NRA to a higher strength level (greater attenuation) or conducting an MCI test with an instrument instead of sung speech corpus in the melodic contour might worsen the MCI scores. Lastly, results obtained from the questionnaires indicate that strengthening the NRA improves slightly the clarity of the singing voice and the music enjoyment. However, no significant difference was detected. These results are consistent with Kim et al. (2020), where the effect of NRA on HA users on the perceived music quality was investigated and no effect was shown. Results obtained from the questionnaires are in contrast to some previous studies on NRA used in HA users (Chasin & Rousso, 2004; Croghan et al., 2014) that suggested disabling the NRA for music listening. Some studies have investigated the effect of NRA on music perception and enjoyment by evaluating only two conditions, namely, NRA off and NRA set to the highest level. A related point to consider is the amount of distortion caused by the NRA on the target voice. It is likely that a too-aggressive configuration of the NRA, for example, NRA set to the highest attenuation level, introduces distortions that can be perceived by CI users (Benesty et al., 2006; Chen et al., 2006) to both the singing voice and the background instruments.

Another algorithm in the CI sound coding strategy that may reduce the complexity of the music is the NofM band selection. Results from objective measures show that a reduction in the number of selected bands, the *N* in NofM algorithm, from 10 to 6 will enhance the vocals by around 0.2 dB electrically. Results obtained from the perceptual evaluation show that reducing the number of selected bands for stimulation N from 8 to 6 provided significantly better SRTs in the speech understanding test with background instruments. Results from the MCI test and questionnaires show no significant benefit of reducing *N* from 8 to 6, however, the mean scores obtained from questionnaires show an improvement in overall music enjoyment and the clarity of the singing voice for CI users when the number of selected bands is reduced.

The third signal processing block in the CI sound coding strategy that may improve music perception in CI users is the compression system. Objective percentile analysis shows that measures based on Music VG provide a wider EDR for singing music to CI users. In the speech understanding with background instruments test, Music VG provided the best mean SRT with respect to other tested parameterizations. Moreover, in the MCI test, in the conditions containing background instruments, Music VG provided the highest mean score. Finally, in the questionnaires, the Music VG parameterization obtained the highest score in both the overall singing music enjoyment and the clarity of the singing voice. In the questionnaires, Music VG was the only parameterization that provided a significant improvement ($p = .020$) in comparison to the baseline parameterization. Halliwell et al. (2015) showed that CI users did not prefer a compressed version (with a 1.5:1 compression ratio) of a music piece with respect to the original music. Gilbert et al. (2019), however, showed that CI users are less sensitive to the amount of compression in music. This result might have been caused by the strong compression applied to the music by the sound coding strategy and the much narrower EDR available for CI users in applied to recorded comparison to the perceivable DR in NH listeners. Gilbert et al. (2022) showed that CI users prefer less compression (allocate more EDR) to the softer passages (below the knee point) and be more compressive for louder (above the knee point) parts of music with respect to the clinical compression system. Results of our study show the efficacy and importance of a back-end compression for music in the Crystalis sound coding strategy.

All perceptual experiments were conducted without any acclimatization time. Adaptation to the parameterizations might improve the results obtained from the perceptual evaluation. Moreover, the clinical CI sound coding parameterization of eight out of ten participants was identical to the baseline parameterization used in this study. The familiarity of these subjects with the baseline parameterization may have had a beneficial impact on the results obtained with this parameterization. The outcomes of this study can have a clinical implication for enhancing music appreciation for CI users. This can be in the form of an additional program based on a specific parameterization of the sound coding strategy stored on the speech processor.

Conclusions

In this study, we aimed at reducing the complexity of music and enhancing the singing voice by optimizing the sound coding strategy. We introduced new objective measures and performed perceptual evaluations by utilizing speech understanding and music perception tests in CI users. For this purpose, we created a speech understanding test with background instruments to measure lyrics understanding and created a sung speech corpus that was used in an MCI task with and without background instruments. Finally, we used questionnaires to assess the subjective experience of CI users. The results of the study show that:

- Reducing the number of selected bands for stimulation and strengthening the NRA improved the $\mathrm{VIR}_{\mathrm{el}}$ using objective measures introduced in this study.
- MCI performance was significantly poorer with multiple-word contours than with single-word contours. Moreover, adding background instruments and decreasing the number of semitone spacing did significantly worsen the MCI performance of CI users.
- Reducing the number of selected bands for stimulation significantly improved speech understanding with background instruments.
- The new music VG back-end compressor significantly improved speech understanding in music with respect to the clinical back-end compressor and provided significantly better scores in questionnaires.

Based on these results, we conclude that a novel sound coding strategy for music can be created to improve music appreciation in CI users. The results from the study indicate that reducing the number of selected bands N in NofM and using a back-end compressor specifically for music listening can improve lyrics understanding and music appreciation for CI users. A music sound coding strategy could be configured based on these specific signal processing parameterizations and algorithms to improve the perception of lyrics in music and music appreciation in general.

Acknowledgments

The authors thank all study participants for their time and commitment. We are also thankful to Marianna Vatti, Tim Rausch, and Anika Morgenstern for constructive feedback and support during the study.

Declaration of Conflicting Interests

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

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Oticon Medical, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; grant number Project number 19541470, Project number 446611346).

ORCID iD

Sina Tahmasebi https://orcid.org/0000-0003-1687-9227

References

Benesty, J., Chen, J., Huang, Y., Doclo, S., & Makino, S. (2006). Study of the Wiener Filter for Noise Reduction. https://doi.org/10.1007/3–540-27489-8_2

Bierer, J. A. (2007). Threshold and channel interaction in cochlear implant users: Evaluation of the tripolar electrode configuration. *The Journal of the Acoustical Society of America*, *121*(3), 1642–1653. https://doi.org/10.1121/1.2436712

Buyens, W., van Dijk, B., Moonen, M., & Wouters, J. (2014). Music mixing preferences of cochlear implant recipients: A pilot study. *International Journal of Audiology*, *53*(5), 294–301. https://doi.org/10.3109/14992027.2013.873955

Carlson, M. L., Patel, N. S., Tombers, N. M., DeJong, M. D., Breneman, A. I., Neff, B. A., & Driscoll, C. (2017). Hearing preservation in pediatric cochlear implantation. *Otology & Neurotology : Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, *38*(6), e128–e133. https://doi.org/10.1097/MAO.0000000000001444

Carlyon, R. P., Mahendran, S., Deeks, J. M., Long, C. J., Axon, P., Baguley, D., & Winter, I. M. (2008). Behavioral and physiological correlates of temporal pitch perception in electric and acoustic hearing. *The Journal of the Acoustical Society of America*, *123*(2), 973–985. https://doi.org/10.1121/1.2821986

Chan, T-S., Yeh, T-C., Fan, Z-C., Chen, H-W., Su, L., Yang, Y-H., & Jang, R. (2015). Vocal activity informed singing voice separation with the iKala dataset. 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 718–722. https://doi.org/10.1109/ICASSP.2015.7178063

Chasin, M. (2006). Sound levels for musical instruments. *Hearing Review*, *13*(3), 34–41.

Chasin, M., & Russo, F. A. (2004). Hearing aids and music. *Trends in Amplification*, *8*, 35–47. https://doi.org/10.1177/108471380400800202

Chatterjee, M., Galvin, J. J3rd, & Fu, Q. J., & R. V Shannon. (2006). Effects of stimulation mode, level and location on forward-masked excitation patterns in cochlear implant patients. *Journal of the Association for Research in Otolaryngology : JARO*, *7*(1), 15–25. https://doi.org/10.1007/s10162-005-0019-2

Chen, J., Benesty, J., Huang, Y., & Doclo, S. (2006). New insights into the noise reduction Wiener filter. *IEEE Transactions on Audio, Speech, and Language Processing*, *14*, 1218–1234. https://doi.org/10.1109/TSA.2005.860851

Crew, J. D., Galvin, J. J., & Fu, Q-J. (2016). Perception of sung speech in bimodal cochlear implant users. *Trends in Hearing*, *20*, 2331216516669329. https://doi.org/10.1177/2331216516669329

Croghan, N. B., Arehart, K. H., & Kates, J. M. (2014). Music preferences with hearing aids: Effects of signal properties, compression settings, and listener characteristics. *Ear and Hearing*, *35*, e170–e184. https://doi.org/10.1097/AUD.0000000000000056

Drennan, W. R., & Rubinstein, J. T. (2008). Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of Rehabilitation Research and Development*, *45*(5), 779–789. PMID: 18816426; PMCID: PMC2628814. https://doi.org/10.1682/JRRD.2007.08.0118

Eargle, J. (2005). *Handbook of recording engineering*. Springer Science & Business Media. ISBN 9780387284705.

Fujita, S., & Ito, J. (1999). Ability of nucleus cochlear implantees to recognize music. *Annals of Otology, Rhinology & Laryngology*, *108*(7), 634–640. https://doi.org/10.1177/000348949910800702

Gajecki, T., & Nogueira, W. (2018). Deep learning models to remix music for cochlear implant users. *The Journal of the Acoustical Society of America*, *143*, 3602–3615. https://doi.org/10.1121/1.5042056

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

Galvin, J. J.III, Fu, Q.-J., & Shannon, R. V. (2009). Melodic contour identification and music perception by cochlear implant users. *Annals of the New York Academy of Sciences*, *1169*, 518–533. https://doi.org/10.1111/j.1749-6632.2009.04551.x

Gauer, J., Nagathil, A., Martin, R., Thomas, J. P., & Völter, C. (2019). Interactive evaluation of a music preprocessing scheme for cochlear implants based on spectral complexity reduction. *Frontiers in Neuroscience*, *13*, 904. https://doi.org/10.3389/fnins.2019.01206

Gfeller, K. (2009). Music and cochlear implants: Not in perfect harmony. *ASHA Leader*, *2009*, 090616g.

Gfeller, K., Knutson, J. F., Woodworth, G., Witt, S., & DeBus, B. (1998). Timbral recognition and appraisal by adult cochlear implant users and normal-hearing adults. *Journal of the American Academy of Audiology*, *9*(1), 1 – 19. PMID: 9493937.

Gfeller, K., Olszewski, C., Rychener, M., Sena, K., Knutson, J. F., Witt, S., & Macpherson, B. (2005). Recognition of "real-world" musical excerpts by cochlear implant recipients and normal-hearing adults. *Ear and Hearing*, *26*(3), 237 –250. https://doi.org/10.1097/00003446-200506000-00001

Gfeller, K., Witt, S., Mehr, M. A., Woodworth, G., & Knutson, J. (2002). Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Annals of Otology, Rhinology & Laryngology*, *111*(4), 349–356. https://doi.org/10.1177/000348940211100412

Gfeller, K., Woodworth, G., Robin, D. A., Witt, S., & Knutson, J. F. (1997). Perception of rhythmic and sequential pitch patterns by normally hearing adults and adult cochlear implant users. *Ear and Hearing*, *18*(3), 252–260. https://doi.org/10.1097/00003446-199706000-00008

Gilbert, M., Jiradejvong, P., & Limb, C. (2019). Effect of compression on musical sound quality in cochlear implant users. *Ear & Hearing*, *40*(6), 1368–1375. https://doi.org/10.1097/AUD.0000000000000715

Gilbert, M. L., Deroche, M. L. D., Jiradejvong, P., Chan, B., & Limb, C. J. (2022). Cochlear implant compression optimization for musical sound quality in MED-EL users. *Ear and hearing*, *43*(3), 862–873. https://doi.org/10.1097/AUD.0000000000001145

Goldsworthy, R. L., & Shannon, R. V. (2014). Training improves cochlear implant rate discrimination on a psychophysical task. *The Journal of the Acoustical Society of America*, *135*(1), 334–341. https://doi.org/10.1121/1.4835735

Greenwood, D. D. (1991). Critical bandwidth and consonance in relation to cochlear frequency-position coordinates. *Hearing Research*, *54*(2), 164–208. https://doi.org/10.1016/0378-5955(91)90117-R

Guevara, N., Bozorg-Grayeli, A., Bebear, J. P., Ardoint, M., Saaï, S., Gnansia, D., Hoen, M., Romanet, P., & Lavieille, J. P. (2016). The voice track multiband single-channel modified Wiener-filter noise reduction system for cochlear implants: Patients' outcomes and subjective appraisal. *International Journal of Audiology*, *55*(8), 431–438. https://doi.org/10.3109/14992027.2016.1172267

Hahlbrock, K. H. (1953). Über sprachaudiometrie und neue wörterteste. *Archiv Fur Ohren-, Nasen- Und Kehlkopfheilkunde*, *162*, 394–431. https://doi.org/10.1007/BF02105664

Halliwell, E. R., Jones, L. L., Fraser, M., Lockley, M., Hill-Feltham, P., & McKay, C. M. (2015). Effect of input compression and input frequency response on music perception in cochlear implant users. *International Journal of Audiology*, *54*(6), 401–407. https://doi.org/10.3109/14992027.2014.986689

Hsiao, F. (2008). Mandarin melody recognition by pediatric cochlear implant recipients. *Journal of Music Therapy*, *45*(4), 390–404. https://doi.org/10.1093/jmt/45.4.390

Hsiao, F., & Gfeller, K. (2012). *Music Perception of Cochlear Implant Recipients with Implications for Music Instruction: A Review of Literature*.

Innes-Brown, H., Au, A., Stevens, C., Schubert, E., & Marozeau, J. (2012). *New music for the Bionic Ear: An assessment of the enjoyment of six new works composed for cochlear implant recipients*.

Kam, A. C., Ng, I. H., Cheng, M. M., Wong, T. K., & Tong, M. C. (2012). Evaluation of the ClearVoice strategy in adults using HiResolution fidelity 120 sound processing. *Clinical And Experimental Otorhinolaryngology*, *5*(Suppl 1), S89–S92. https://doi.org/10.3342/ceo.2012.5.S1.S89

Kasturi, K., & Loizou, P. C. (2007). Effect of filter spacing on melody recognition: Acoustic and electric hearing. *The Journal of the Acoustical Society of America*, *122*(2), EL29–EL34. https://doi.org/10.1121/1.2749078

Kim, H. J., Lee, J. H., & Shim, H. J. (2020). Effect of digital noise reduction of hearing aids on music and speech perception. *Journal of Audiology & Otology*, *24*(4), 180–190. https://doi.org/10.7874/jao.2020.00031

Kirchberger, M., & Russo, F. A. (2016). Dynamic range across music genres and the perception of dynamic compression in hearing-impaired listeners. *Trends in Hearing*, *20*, 2331216516630549. https://doi.org/10.1177/2331216516630549

Kludt, E., Nogueira, W., Lenarz, T., & Buechner, A. (2021). A sound coding strategy based on a temporal masking model for cochlear implants. *PLoS One*, *16*(1):e0244433. https://doi.org/10.1371/journal.pone.0244433

Kohlberg, G. D., Mancuso, D. M., Griffin, B. M., Spitzer, J. B., & Lalwani, A. K. (2016). Impact of noise reduction algorithm in cochlear implant processing on music enjoyment. *Otology & Neurotology*, *37*(5), 492–498. https://doi.org/10.1097/MAO.0000000000001041

Landsberger, D. M., Vermeire, K., Stupak, N., Lavender, A., Neuka, J., Van de Heyning, P., & Svirsky, M. A. (2020). Music is more enjoyable with two ears, even if one of them receives a degraded signal provided by a cochlear implant. *Ear and Hearing*, *41*(3), 476–490. https://doi.org/10.1097/AUD.0000000000000771

Langner, F., Büchner, A., & Nogueira, W. (2020). Evaluation of an adaptive dynamic compensation system in cochlear implant listeners. *Trends in Hearing*, *24*, 2331216520970349. https://doi.org/10.1177/2331216520970349

Limb, C. J., Molloy, A. T., Jiradejvong, P., & Braun, A. R. (2010). Auditory cortical activity during cochlear implant-mediated perception of spoken language, melody, and rhythm. *Journal of the Association for Research in Otolaryngology*, *11*(1), 133–143. https://doi.org/10.1007/s10162-009-0184-9

Limb, C. J., & Roy, A. T. (2013). Technological, biological, and acoustical constraints to music perception in cochlear implant users. *Hearing Research*, *308*, 13–26. https://doi.org/10.1016/j.heares.2013.04.009

Looi, V., Gfeller, K., & Driscoll, V. (2012). Music appreciation and training for cochlear implant recipients: A review. *Seminars in Hearing*, *33*(4), 307–334. https://doi.org/10.1055/s-0032-1329221

Looi, V., McDermott, H., McKay, C., & Hickson, L. (2007). Comparisons of quality ratings for music by cochlear implant and hearing aid users. *Ear and Hearing*, *28*(2), 59S – 61S. https://doi.org/10.1097/AUD.0b013e31803150cb

Looi, V., McDermott, H., McKay, C., & Hickson, L. (2008). Music perception of cochlear implant users compared with that of hearing aid users. *Ear and Hearing*, *29*(3), 421–434. https://doi.org/10.1097/AUD.0b013e31816a0d0b

Ma, Z., Man B, De, Pestana, PE. L., Black, DA. A., & Reiss, JO. D. (2015). Intelligent multitrack dynamic range compression. *J. Audio Eng. Soc*, *63*(6), 412–426. https://doi.org/10.17743/jaes.2015.0053

Macherey, O. (2010). Temporal pitch percepts elicited by dual-channel stimulation of a cochlear implant". *The Journal of the Acoustical Society of America*, *127*, 339–349. https://doi.org/10.1121/1.3269042

Man, B. D., Leonard, B., King, R., & Reiss, J. D., (2014). An analysis and evaluation of audio features for multitrack music mixtures. 15th International Society for Music Information Retrieval Conference.

McDermott, H. J. (2004). Music perception with cochlear implants: A review. *Trends in Amplification*, *8*(2), 49–82. https://doi.org/10.1177/108471380400800203

McKay, C. M., McDermott, H. J., & Clark, G. M. (1996). The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees. *The Journal of the Acoustical Society of America*, *99*(2), 1079–1090. https://doi.org/10.1121/1.414594

Moore, B. C. (2008). The choice of compression speed in hearing AIDS: Theoretical and practical considerations and the role of individual differences. *Trends In Amplification*, *12*(2), 103–112. https://doi.org/10.1177/1084713808317819

Moore, B. C., & Sęk, A. (2016). Preferred compression speed for speech and music and its relationship to sensitivity to temporal fine structure. *Trends In Hearing*, *20*, 2331216516640486. https://doi.org/10.1177/2331216516640486

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing*, *2*(5–6), 183–197. https://doi.org/10.1016/0925-2312(91)90023-5

Nagathil, A., Weihs, C., & Martin, R. (2016). Spectral complexity reduction of music signals for mitigating effects of cochlear hearing loss. *IEEE ACM Trans. Audio Speech Lang. Process*, *24*, 445–458. https://doi.org/10.1109/TASLP.2015.2511623

Nagathil, A., Weihs, C., Neumann, K., & Martin, R. (2017). Spectral complexity reduction of music signals based on frequency-domain reduced-rank approximations: An evaluation with cochlear implant listeners. *Journal of the Acoustical Society of America*, *142*(3), 1219. https://doi.org/10.1121/1.5000484

Nogueira, W., Büchner, A., Lenarz, T., & Edler, B. (2005). A psychoacoustic 'NofM'-type speech coding strategy for cochlear implants. *EURASIP J. Appl. Signal Process*, 101672. https://doi.org/10.1155/ASP.2005.3044

Nogueira, W., Litvak, L. M., Saoji, A. A., & Büchner, A. (2015). Design and evaluation of a cochlear implant strategy based on a "phantom" channel. *PLOS ONE*, *10*(3), e0120148. https://doi.org/10.1371/journal.pone.0120148

Nogueira, W., Nagathil, A., & Martin, R. (2019). Making music more accessible for cochlear implant listeners: Recent developments. *IEEE Signal Processing Magazine*, *36*(1), 115–127. https://doi.org/10.1109/MSP.2018.2874059

Nogueira, W., Rode, T., & Büchner, A. (2016). Spectral contrast enhancement improves speech intelligibility in noise for cochlear implants. *Journal of the Acoustical Society of America*, *139*(2), 728–739. PMID: 26936556. https://doi.org/10.1121/1.4939896

Omran, S. A., Lai, W., & Dillier, N. (2010). Pitch ranking. Melody contour and instrument recognition tests using two semitone frequency maps for nucleus cochlear implants. *EURASIP J. Audio Speech Music Process*, 948565 (2011). https://doi.org/10.1155/2010/948565

Pons, J., Janer, J., Rode, T., & Nogueira, W. (2016). Remixing music using source separation algorithms to improve the musical experience of cochlear implant users. *The Journal of the Acoustical Society of America*, *140*(6), 4338–4349. https://doi.org/10.1121/1.4971424

Rafii, Z., Liutkus, A., Stöter, F.-R., Mimilakis, S. I., & Bittner, R. (2017). The MUSDB18 corpus for music separation. https://doi.org/10.5281/zenodo.1117372

Schulz, E., & Kerber, M., (1994). Music perception with the MED-EL implants. In I. J. Hochmair-Desoyer & E. S. Hochmair (Eds.), *Advances in cochlear implants* (pp. 326 – 332). Datenkonvertierung, Reproduktion und Druck.

Segovia-Martinez, M., Gnansia, D., & Hoen, M. (2016). Coordinated adaptive processing in the neuro cochlear implant system [Oticon Medical White Paper.

Skinner, M. W., Holden, L. K., Holden, T. A., Demorest, M. E., & Fourakis, M. S. (1997). Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants. *Journal of the Acoustical Society of America*, *101*(6), 3766 – 37682. https://doi.org/10.1121/1.418383

Sorkin, D., & Zombek, L. (2021). Optimizing outcomes with a cochlear implant: Tips for adults. *The Hearing Journal*, *74*(7), 28. https://doi.org/10.1097/01.HJ.0000766260.80033.cf

Stakhovskaya, O., Sridhar, D., Bonham, B. H., & Leake, P. A. (2007). Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants. *Journal of the Association for Research in Otolaryngology*, *8*(2), 220–233. https://doi.org/10.1007/s10162-007-0076-9

Stickney, G. S., Loizou, P. C., Mishra, L. N., Assmann, P. F., Shannon, R. V., & Opie, J. M. (2006). Effects of electrode design and configuration on channel interactions. *Hearing Research*, *211*(1-2), 33–45. https://doi.org/10.1016/j.heares.2005.08.008

Tabibi, S., Kegel, A., Lai, W. K., & Dillier, N. (2016). Investigating the use of a gammatone filterbank for a cochlear implant coding strategy. *Journal of Neuroscience Methods*, *277*, 10.

Tahmasebi, S., Gajęcki, T., & Nogueira, W. (2020). Design and evaluation of a real-time audio source separation algorithm to remix music for cochlear implant users. *Frontiers in Neuroscience*, *14*, 294. https://doi.org/10.3389/fnins.2020.00434

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C.J., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. *Hearing Research*, *377*, 109 –121, https://doi.org/10.1016/j.heares.2019.03.011, ISSN 0378-5955

Wagener, K. C., Kühnel, V., & Kollmeier, B. (1999). Entwicklung und evaluation eines satztests für die deutsche sprache I: Optimierung des oldenburger satztests [development and evaluation of a sentence test for the German language II: Optimization of the oldenburg sentence test]. *Zeitschrift Audiologie/Audiological Acoustics*, *38*(2), 44–56.

Zeng, F. G. (2004). Trends in cochlear implants. *Trends in Amplification*, *8*(1), 1 – 34. https://doi.org/10.1177/108471380400800102

Zeng, F. G., & Galvin, J. J. (1999). 3rd. Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear and Hearing*, *20*(1), 60 –74. PMID: 10037066. https://doi.org/10.1097/00003446-199902000-00006