Hearing Research 335 (2016) 207–219

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Categorization of common sounds by cochlear implanted and normal hearing adults

E. Collett [a, b, e], M. Marx [a, b, c], P. Gaillard [d], B. Roby [a, b, c], B. Fraysse [c], O. Deguine [a, b, c], P. Barone [a, b, *]

[a] Université de Toulouse, CerCo UMR 5549 CNRS, Université Paul Sabatier, Toulouse, France
[b] Université de Toulouse, CerCo UMR 5549 CNRS, Faculté de Médecine de Purpan, Toulouse, France
[c] Service d'Oto–Rhino–Laryngologie et Oto-Neurologie, Hopital Purpan, Toulouse, France
[d] Université de Toulouse, CLLE UMR 5263, CNRS, UT2J, Université de Toulouse Jean-Jaurès, Toulouse, France
[e] Advanced Bionics SARL, France

A R T I C L E I N F O

Article history:
Received 1 June 2015
Received in revised form
3 March 2016
Accepted 14 March 2016
Available online 2 April 2016

Keywords:
Cochlear implant
Auditory categorization
Free-sorting
Hierarchical clustering
Multiple Correspondence Analysis

A B S T R A C T

Auditory categorization involves grouping of acoustic events along one or more shared perceptual dimensions which can relate to both semantic and physical attributes. This process involves both high level cognitive processes (categorization) and low-level perceptual encoding of the acoustic signal, both of which are affected by the use of a cochlear implant (CI) device. The goal of this study was twofold: I) compare the categorization strategies of CI users and normal hearing listeners (NHL) II) investigate if any characteristics of the raw acoustic signal could explain the results. 16 experienced CI users and 20 NHL were tested using a Free-Sorting Task of 16 common sounds divided into 3 predefined categories of environmental, musical and vocal sounds. Multiple Correspondence Analysis (MCA) and Hierarchical Clustering based on Principal Components (HCPC) show that CI users followed a similar categorization strategy to that of NHL and were able to discriminate between the three different types of sounds. However results for CI users were more varied and showed less inter-participant agreement. Acoustic analysis also highlighted the average pitch salience and average autocorrelation peak as being important for the perception and categorization of the sounds. The results therefore show that on a broad level of categorization CI users may not have as many difficulties as previously thought in discriminating certain kinds of sound; however the perception of individual sounds remains challenging.

1. Introduction

For more than 25 years cochlear implants (CI) have been an efficient method for restoring partial hearing in patients with profound bilateral sensorineural hearing loss (Copeland and Pillsbury, 2004). The CI has led to considerable success in the functional rehabilitation of deafness in terms of the restitution of speech recognition ability (Moller, 2006). However the auditory information sent to the brain is spectrally degraded (Shannon et al., 1995) and lacks the fine temporal structure information which is crucial for certain aspects of speech comprehension, most notably the perception of prosodic information (Friesen et al., 2001; Zeng et al., 2005; Lorenzi et al., 2006; Marx et al., 2014). Temporal modulations are however more optimally retained by the implant and as a result CI users rely on these cues rather than spectral modulations for speech comprehension (Doucet et al., 2006; Loebach and Pisoni, 2007; Cabrera et al., 2014). These limitations contribute to an initial period of adaptation of about six months after implantation before CI users reach an optimal level of speech performance (Barone and Deguine, 2011). During this initial period of adaptation CI users develop adaptive strategies and rely strongly on visual cues to complement the impoverished auditory signal delivered by the implant (Summerfield, 1992; Tyler et al., 1997; Grant et al., 1998; Kaiser et al., 2003; Rouger et al., 2007).

Other aspects of auditory perception remain problematic for many CI users especially concerning the recognition of environmental sounds and music. The reduction in fine spectral information negatively affects the musical abilities of CI users, for example the perception of melody and timbre within instrumental passages

* Corresponding author. Cerveau & Cognition, CNRS UMR 5549, Pavillon Baudot, CHU Purpan, BP 25202, 31052 Toulouse Cedex, France.
E-mail address: Pascal.barone@cerco.ups-tlse.fr (P. Barone).

http://dx.doi.org/10.1016/j.heares.2016.03.007
0378-5955/

relies on fine spectral details that cannot be efficiently transmitted through the implant (Gfeller et al., 2002; Looi et al., 2008; Cousineau et al., 2010). Similarly, whilst a CI can improve the awareness of environmental sounds (Looi and Arnephy, 2010) environmental sound identification in CI users is commonly described as poor with correct identification varying from only 45% (Shafiro et al., 2011) to nearly 80% in experienced CI users (Reed and Delhorne, 2005). Studies of environmental sound perception have shown the importance of spectral and temporal dynamics for environmental sound recognition such that more complicated sounds (i.e. with faster changing temporal and spectral information) are more difficult for CI to identify (Reed and Delhorne, 2005; Reddy et al., 2009) whilst those with a more simple spectral or temporal content are more easily identified. This may result in sounds that have a strong sense of harmony (more than two harmonics present in the signal) or are repetitive (Inverso and Limb, 2010). Gygi et al. (2004) also found that information within a frequency band of between 1200 and 2400 Hz was linked to the recognition of environmental sounds. Finally Shafiro (2008) identified two groups of sounds that required high or low levels of spectral resolution for correct identification and was able to discriminate between them based only on the number of bursts in the envelope and the standard deviation of the centroid velocity.

Previous studies in CI users or in protocols using CI simulation (vocoding) in normal-hearing listeners (NHL) have shown that when reducing the spectro-temporal information both NHL and CI users present a strong impairment in discriminating vocal from environmental sounds (Loebach and Pisoni, 2008; Shafiro, 2008; Leech et al., 2009; Massida et al., 2011). However, it has not been possible to find an exact set of acoustic cues that can account precisely for environmental sound recognition in either NHL or CI users. In light of these observations on the limitations of sound processing imposed by the implant it is of importance to better understand how CI users process and discriminate vocal, environmental and musical sounds. A better grasp of the cues used to build-up a representation of these sounds could provide crucial information to develop coding strategies that would allow processing to be better adapted for specific sound categories.

In day to day life, human beings must contend with a multitude of sounds almost continuously and to attend to every single one would require a lot of effort. Part of the process that makes this easier is categorization, which enables knowledge about the sound to be inferred from the auditory category and is less demanding than specifically identifying the sound. This has been found in the categorization of phonemes, voice gender and even the material and action of a sound source (Liberman et al., 1967; Dubois, 2000; Belin et al., 2004; Lemaitre and Heller, 2012).

Two modes of listening are involved in categorization (Gaver, 1993). The first, “everyday listening”, refers to a listeners perception of the properties of a sound source and the semantic information that is perceived, for example the type of material (wood, metal, plastic), the action producing the sound (hitting, scraping, rubbing) or the age and gender of a speaker. The second mode, “musical listening”, refers to the perception of qualitative aspects of the sound such as the pitch, loudness, timbre and also the emotional content (pleasant or unpleasant). These qualities are related to physical aspects of the acoustic signal rather than the sound source itself. Studies have shown that in general categories are based on everyday listening and the semantic properties of the sound source rather than musical listening and the acoustical properties (Gygi et al., 2007; Inverso and Limb, 2010). However it has also been shown when a target sound is preceded by non-living sounds categorization is more strongly based on acoustic information (Giordano et al., 2010).

Common categories that have been described include human and animal vocalizations, impact sounds, water sounds (Gygi et al., 2007), human and traffic noise (Guastavino, 2007), living and non-living sounds (De Lucia et al., 2012). The process of categorizing is also influenced by the listener's experience which defines their pre-existing knowledge of categories and the potential identification of sound sources. Importantly if a sound source cannot be identified the relevant semantic information cannot be retrieved from categorical knowledge. Further, the type of experimental task has been shown to have an influence on the strategies developed by participants to categorize sounds. For example, testing similarity somewhat forces participants to use the musical listening mode and thus to rely more on acoustical information (Goldstone, 1994; Gygi et al., 2007). In trying to understand how participants categorize different sounds, some studies have analyzed the categorization of many stimuli with similarity matrices and Multidimensional Scaling (MDS) (Bonebright, 1996; Guastavino, 2007; Gygi et al., 2007). This is a process whereby a large array of similarity judgments can be reduced to only a few dimensions representing the entire perceptual space. Work can then be carried out to interpret the various perceptual and acoustic features on which these perceptual dimensions are based (Bonebright, 1996).

To date there has been no study that has directly addressed how CI users may categorize common sounds using a Free-Sorting Task protocol (Berland et al., 2015). Studies have either used simulations of CI processing with NHL (Gygi et al., 2007) or measured categorization accuracy with a number of predefined categories that participants were informed of prior to testing (Inverso and Limb, 2010). In the present study we therefore seek to add to the existing work on auditory categorization and to better understand how CI users perceive three predefined groups of sounds (environmental, musical and vocal) by comparing performance with NHL. Our strategy was to select three broad classes of sounds based on their respective ecological values. As previously mentioned the perception of vocal sounds is highly important for social interaction and thus the primary aim of CI coding strategies has been to focus on the processing of these sounds. Musical sounds also constitute a specific class of stimuli on which efforts have been directed in CI processing due to the strong social and personal value of music to peoples everyday lives (Looi et al., 2012) and the relative difficulty in the transmission of the richness of such sounds by the implant. A third class of sounds, common environmental sounds, was also chosen as these sounds are reported by CI users as being important for quality of life and have been studied less in comparison to vocal and musical sounds.

The first objective was to use Multiple Correspondence Analysis (MCA) and a Hierarchical Clustering with Principal Components (HCPC) to see if I) CI users are able to distinguish the three groups of sounds from one another and II) how exactly they may do this based on semantic or acoustical perceptions of the sounds. With poorer sound identification CI users may use categorization strategies more strongly based on the similarity of acoustical cues (via musical listening) rather than the semantic information of the sound or sound producing action (everyday listening). In addition to this, the link between hearing a sound and retrieving the associated semantic information is a process that can be affected by familiarity with the sound (Ballas, 1993) and which is likely affected by periods of hearing impairment experienced by all CI users. Finally, because auditory abilities of CI users can improve with training (Shafiro et al., 2012, 2015), we recruited CI users who had at least twelve months of experience with their implant. However, it is not known exactly how the adaptation to the implant and the period of hearing impairment may affect the link between the perception of a sound and its semantic representation and therefore the categorization strategies of CI users. The second objective is to ascertain whether any acoustical properties of the sounds can account for the categorization strategies used by NHL and CI users by correlating results of acoustical analysis to those of the MCA. The results of this study will therefore explore how CI users categorize and discriminate different types of sounds by analyzing participants' perceptions and acoustical cues.

2. Method and materials

2.1. Participants

Two groups of participants were involved in the study. The first group consisted of 20 normal hearing listeners (NHL) who were native French speaking adults (9 males age 22 ± 2 years). Auditory functions for NHL were self-reported as normal with no recent exposure to loud noises or history of auditory or neurological disorders. The second group of participants consisted of 16 adult cochlear implant (CI) users, postlingually deaf and native French speaking adults (7 males, age 56 ± 15 years, see Table 1). Noticeably at only 15 years of age participant LERMAR is significantly younger than the other participants (one tailed Crawford & Howell test $t = -8.86$, $p < 0.001$). CI users were selected from a list of "experienced" listeners with the requirement being duration of implantation greater than twelve months. Only two patients, (see Table 1), had durations close to this (13 and 15 months) with the group average being greater (mean 57 ± 32 months). Most of the participants presented a high level of speech recovery with a mean speech comprehension score of 80% ± 12.4% (see Table 1). When compared to the group average only one patient (RAUCLA) has significantly lower scores of for disyllabic word recognition (57%, $t = -6.022$, $p < 0.001$) and for speech in noise comprehension (44%, $t = -4.48$, $p < 0.001$). The hearing threshold of the non-implanted ear shows that all patients have a poor level of residual hearing (mean 89 dB ± 18 dB). Overall the auditory performance scores show the experience of CI listeners is homogeneous with similar levels of auditory performance. Word recognition scores were collected from patients during assessments with a speech therapist prior to testing. All subjects were tested on open-set recognition for French disyllabic words using the Fournier list (see Rouger et al., 2007) and also for speech in noise using the standard MBAA list. Finally all participants gave written informed consent prior to their inclusion in the study.

2.2. Stimuli & procedure

The two groups of participants completed a Free Sorting Task (FST) of sixteen short sounds (2–3 s duration) which are described

Table 1
Summary of CI user data detailing participants age at testing and duration of deafness prior to receiving a CI in years, duration of implantation in months and which ear was implanted (L-left; R-right). The model of implant and coding strategy used by each participant is also listed with the stimulation speed and maximum stimulation available for participants with a Cochlear brand implant. Scores for auditory performance only using the implanted ear are also given for disyllabic word recognition – the number of words correctly identified out of 20 given as a percentage; sentences in noise where the SNR is +10 dB – given as a percentage score is also stated along with importantly the threshold of hearing in the participants' non-implanted ear which gives a measure of residual hearing for each participant. Unfortunately certain data could not be retrieved for participant LERMAR. Finally values for the mean and standard deviation are also stated.

Participant ID	Age at testing	Duration of deafness (years)	Duration of implantation (months)	Implanted ear	CI model	Coding strategy	Stimulation speed	Maximum stimulation	Hearing threshold (non-implanted ear (dB)	Disyllabic word recognition – CI only (% correct)	Sentence in noise – SNR +10 dB (% correct)
ALAREN	73	5	62	R	Cochlear CI24RE	ACE 900 11	900	11	72.5	80	96
ANDJOS	53	9	53	R	Cochlear CI24RECA	ACE 1200 10	1200	10	105	80	95
ANSCRI	48	5	35	R	Cochlear CI24CA	ACE 720 9	720	9	80	85	99
AUZAND	54	2	34	R	Cochlear Nucleus CI512	ACE 900 10	900	10	70	95	73
BERCAT	51	2	62	R	Cochlear Hybrid	ACE 900 8	900	8	50	60	64
CHAJEA	72	1	83	L	AB Hires 90 K CI	Hi-Res S Fidelity 120	NA	NA	93	80	97
DAMCRI	51	2	36	R	MED EL SONATA	FSP	NA	NA	Deaf	70	83
ESCPAT	48	3	114	R	Cochlear Nucleus CI24	ACE 720 9	720	9	100	80	95
GERJEA	52	8	81	L	Cochlear Nucleus Freedom CI24RECA	ACE 720 9	720	9	91	75	85
GIDGEN	63	1	23	R	CochelarCI24RECA	ACE 720 10	720	10	96	90	89
LAMCEC	63	0.5	17	L	Cochlear C512	ACE 900 10	900	10	90	95	95
LERMAR	15	14	126	R	Cochlear CI24RCS	ACE 900 16	900	16	–	–	–
MAMSOL	66	5	37	L	AB Hires 90 K	Hi-Res S	NA	NA	110	75	75
MATARL	76	2	14	R	Cochlear Nucleus C512	ACE 900 9	900	9	110	80	81
RAUCLA	70	2	82	L	AB Hires 90 k	Hi-Res S	NA	NA	80	55	13
SEBMAR	47	8	83	R	Cochlear Nucleus CI24	ACE 900 11	900	11	100	100	100
Mean	56	5	59	–			900	10	90	80	83
Std. dev.	15	4	34	–			–	–	17	12	22

in Table 2 alongside abbreviated ID tags and category labels. Sounds were taken from a database owned by the PETRA group at the University of Toulouse Le Mirail (http://petra.univ-tlse2.fr) and were chosen to cover a broad range of semantic and acoustic information as examples of the three predefined categories – environmental, vocal and musical sounds. All stimuli were monophonic and recorded in.wav format with a sampling frequency of 44,100 Hz. A limit of sixteen stimuli was chosen following advice from the ORL department at the Purpan Hospital to ensure that the test would not be too fatiguing for CI users and therefore limit this as a source of bias. The two participant groups were tested in quiet listening rooms, NHL at the CerCo laboratory and CI users at the Purpan Hospital. Both groups were seated in front of a PC monitor positioned at eye-level with two Roline Digital loudspeakers located on each side at a distance of 1 m. Stimuli were presented in stereo at a level of 65 dB SPL (measured at head height with a Sound Level Meter at a distance of 1 m) via the loudspeakers in free-field listening conditions. Testing was carried out using the open-source TCL-LabX software (http://petra.univ-tlse2.fr/tcl-labx/) which acted as the interface for the FST. The sixteen sounds were represented on the computer by sixteen numbered and colored squares which were positioned in the same order for all participants.

The task for participants was to listen to the sixteen sounds and place them into groups i.e. create categories by any means they chose. Only minimal feedback was given by the experimenter in order to facilitate the completion of the experiment. Sounds were played by using the PC mouse to double click on each square and categories were created by dragging and positioning squares together on screen. This was always done by the participants themselves, including the CI users. Once participants had finished positioning the squares into categories they were asked to listen to each sound a final time to verify their choices before ending the experiment. Participants were then asked to enter a brief description for each category into the computer using the keyboard. There was no limit on the amount of time given to complete the test or to the number of times a specific sound could be listened to (referred to as the number of playbacks). Participants were also allowed to create as many or as few categories as they wished such that a single category could contain only a single stimulus or all sixteen. The TCL-LabX software also recorded performance data and statistics for all participants including the number of categories created, the number of playbacks and the duration of the experiment.

Table 2
List of abbreviated labels with their predefined category (as determined by the experimenters) and full descriptions of the 16 sound stimuli used in the free-categorization task.

Sound ID	Description	Predefined category
ALRM	Alarm clock ringing	Environmental
CAR	Car engine starting	
DR	Door opening	
FSTP	Footsteps	
GLS	Glass breaking	
HELI	Helicopter flying overhead	
WTR	Running water	
BEL	Church bells	
GTR	Arpeggiated notes on an acoustic guitar	Musical
OBOE	Single note from an oboe	
VLN	Short 7 note melody on a violin	
XLY	Single note from a xylophone	
CGH	Male voice coughing	Vocal
FEM	Female voice speaking	
LGH	Female voice laughing	
MALE	Male voice speaking	

2.3. Categorical analysis

To analyze the categories that participants created two different functions were used in R′. Firstly Multiple Correspondence Analysis (MCA) was applied to the indicator matrix outputted by the TCL LabX software. The indicator matrix itself represents the results as an array of categorical variables (participants) as columns and categorical items (sound stimuli) as rows, with each cell containing a number defining the category membership of each sound for each participant. MCA uses Correspondence Analysis (CA) in order to represent each sound as a data point in an n-dimensional Euclidean space based on the categorical values i.e. the categories made by participants. Each dimension is chosen to account for the largest amount of variance possible within the data-set and dimensions are outputted in descending order of variance covered. MCA also performs analysis on the participants in order to find how strongly individual results coincide with the dimensions and as a consequence allows the similarity of participants categorization strategies to be analyzed (Cadoret et al., 2009). A total of fifteen dimensions were used in the analysis with those that covered 8% or more of the total variance being retained. This lead to the first 5 dimensions being used for Hierarchical Clustering based on Principal Components (HCPC). The two most significant dimensions (Dim 1 & Dim 2) were also focused on as they account for the most amount of variance in the data and also show the most significant correlations to acoustic variables measured for the sounds (see Table 6). Importantly dimensions are calculated only to account for variability within the data and are not directly related to any perceptual or physical characteristic of the sounds or ecological data of the participants. There is no a-priori knowledge that can be used to automatically make such a relation and so a certain amount of interpretation is used when commenting on the dimensions (Cadoret et al., 2009).

RV coefficient (RVc) values were also calculated in R′. The RVc is a variation on the squared Pearson correlation coefficient that calculates the correlation between two sets of coordinates represented in a matrix (Robert and Escoufier, 1976; Qannari et al., 2014). In the case of this study the RVc was used to find the correlation between the coordinate matrices for the first five MCA dimensions of the two participant groups.

Secondly a Hierarchical Clustering based on Principal Components (HCPC) was performed on the results of the MCA analysis in order to view a simplified version of the categories of sounds in the form of dendrograms. When using this analysis it is not possible to account for all of the variance (inertia) within the data, i.e. the variability of participant responses, and so a certain amount remains unaccounted for. By increasing the number of desired categories the inertia can however be reduced and it is using this process that the we can choose a final number of categories: if the number of categories is Q then the optimal number of categories is found when the change in inertia is greater when moving from Q_{-1} to Q than from Q to Q_{+1} (Husson et al., 2010). This can also be defined as the value for Q which minimizes Equation (1).

$$\frac{Q - Q_{+1}}{Q_{-1} - Q} \tag{1}$$

Categorical analysis was concluded by finding the Cophenetic Correlation Coefficient (CpCC) (Saraçli et al., 2013) which gives a measure of how accurately the distances between items in the raw data are preserved in the dendrogram (Carr et al., 1999).

Participants' category descriptions were also evaluated as to whether or not the categorization was based on semantic information concerning the sound source or on qualitative information linked with the acoustic signal. For each participant group the number of descriptions that made reference to either semantic or qualitative acoustic aspects of the sound were totaled across all sounds and all participants and then turned into an overall percentage. Examples of descriptions relating to semantic information were *"domestic noise"; "voice, laughter, someone talking"; "musical instruments"*. Whilst examples of descriptions referring to qualitative acoustic aspects included *"melodies"; "noise"; "ringing tones"*; *"disagreeable sounds"*. In order to further understand the categorization performance of participants an estimation of "Category Identification" was inferred from the participants' category descriptions. For each sound the associated category description was evaluated to see if it corresponded to the designated pre-defined category (see Table 2) with percent correct scores then calculated across both sounds and participants In some cases specific descriptions were not given and this is reflected in Table 3 by the "% No comment". In addition a value for "Categorization Accuracy" was calculated as the percentage of stimuli-pairs present in each participant's category choices compared to the stimuli-pairs contained in the predefined categories, of which there are forty possible pairs (Table 2).

2.4. Acoustic analysis

Alongside categorical analysis of participant's responses sounds were analyzed using MatLab for a range of acoustic variables similar to Gygi et al., 2004. To evaluate the functional significance of the MCA dimensions the acoustical values obtained for each sound were then correlated using a Pearson correlation to the coordinates of each MCA dimension of the two participant groups. Six different kinds of acoustical measurements were explored and are detailed below.

1. *Pitch measures* – Including the mean and median frequency, standard deviation of frequency, max frequency, mean pitch salience, max pitch salience. Pitch values were calculated using Slaneys Correlogram model of pitch perception (Slaney and Lyon, 1990) using a temporal windows of 16 ms, frame-rate of 12 frames per second and sampling rate of 16 kHz having firstly downsampled the signal to 16 kHz. Pitch salience, which can also be described as the perceptual strength of the pitch, is calculated by dividing the maximum value of the correlogram (which occurs at zero time lag) by the estimated pitch – taken as the time lag with largest correlation energy. Values range from 0 to 1 where 1 indicates a periodic sound with easily perceivable pitch.
2. *Spectral measures* – Analysis was made on individual sounds to compute the centroid, skew, kurtosis, mean centroid, spectral centroid velocity, spectral centroid uniformity and spectral centroid standard deviation. Centroid, skew and kurtosis are measures of the moments of the spectrum calculated from the Power Spectral Density. The centroid is the central mass of the spectrum linked to the brightness of sound. The skew corresponds to the asymmetry of the probability distribution of frequencies while the kurtosis is a descriptor of the shape of a probability distribution. The remaining variables related to the centroid values are measures of spectral movements within the sounds.
3. *Envelope measures* – Including the number of peaks, mean peak, number of bursts, mean burst, total burst duration, duration ratio computed from the wave envelope of each sound. Peaks, also described as fast transient changes, are defined as points in the envelope where the amplitude is greater than the preceding point by at least 80% of the total amplitude range. Bursts are defined as continuous increases in amplitude of 4 dB held for at least 20 ms (Ballas, 1993), whilst the duration ratio provides an evaluation of the "roughness" of the envelope from the ratio of bursts to overall duration. Envelopes were extracted using 5th order low-pass Butterworth filters with cut-off frequency at 20 Hz, with peak measures calculated having first downsampled the signal to 10 kHz.
4. *Periodicity measures* – Values were obtained on the number of autocorrelation peaks, the maximum autocorrelation peak, mean and standard deviation of autocorrelation peaks and range of data. Periodicity is evaluated by firstly computing the autocorrelation of each sound then calculating the previously stated variables to give an indication of the frequency, strength and uniformity of periodicities. The stated variables then give an indication towards the frequency, strength and uniformity of these periodicities.
5. *Cross-Channel correlation* – an average value of the correlations between envelopes across different frequency channels. Envelopes were extracted using the same process described above and frequency channels were created using band-pass filters centered at 212, 424, 848, 1697, 3394 and 6788 Hz. The six frequency channels generated a total of 15 correlations from which an average was calculated to give an indication of signal uniformity.
6. *RMS values* – The RMS power was measured across different frequency channels centered at frequencies of the Bark scale (Zwicker, 1961).

2.5. Statistical analysis

Statistical testing (ANOVA and *t*-test) was performed on participants performance data (recorded by the TCL-LabX software) to reveal if any significant effects of the participant group or type of sound could explain possible differences of categorization between the two groups. Similar to correlations made with acoustical information, a correlation analysis has also been performed between the participant coordinates of the MCA dimensions and the audiological and historical information that distinguish the CI users – age at testing, duration of deafness, duration of implantation and audiological performances (disyllabic word recognition, sentence in noise comprehension).

Table 3
Category Identification – calculated by comparing the participant's category descriptions to the predefined categories for each sound and inferring whether or not they accurately match. Percentages are calculated across participants for each sound and then averaged for each of the three predefined categories – environmental (E), musical (M) or vocal (V). Some participants were unable to give a comment for certain categories and this is reflected by the % of "No comment".

Category identification (%)	NHL				CI			
	E	M	V	Average	E	M	V	Average
Accurate	56.9	82.5	77.5	68.4	58.6	57.4	70.3	60.5
Inaccurate	40.6	17.5	22.5	30.3	37.5	45.3	28.1	37.1
No comment	2.5	0	0	1.3	3.9	0	1.6	2.3

3. Results

Our results show that in CI users and NHL categorization is predominantly based on semantic information as a result of identifying a sound source. CI users perform the FST in a similar manner to NHL and also appear to categorize the sounds in a very similar manner with outputs of MCA and HCPC analysis in close resemblance for both participant groups. However, certain differences do exist, most notably the greater variability amongst CI users in comparison to NHL. Differences are most likely explained by the poorer perception of spectral and temporal information by CI users as well as differences in listening experience and familiarity with the auditory environment.

3.1. Global performance

The duration for completing the task was on average longer for CI users compared to NHL (mean 330 s vs. 539 s, $p < 0.01$). Regarding the number of playbacks, ANOVA and post-hoc analysis (Turkey-Kramer test) showed no differences between NHL and CI ($p = 0.272$). However, when grouping sounds as either musical, vocal or environmental there was a significant effect on the number of playbacks (ANOVA, $p < 0.01$) such that NHL listened to environmental sounds more times on average (7.5 $\pm$ 0.9) than vocal sounds (5.0 $\pm$ 0.46) or musical sounds (5.4 $\pm$ 0.9). A similar pattern is seen for CI with environmental sounds being listened to more (5.6 $\pm$ 0.44) than musical (5.3 $\pm$ 0.5) and vocal sounds (4.97 $\pm$ 0.92) however this only observed as a trend (ANOVA, $p = 0.5$). There is also an effect of specific sounds ($p = 0.026$) such that across both NHL and CI users the sound of a door closing (*DR*) was listened to significantly more times (6.8 $\pm$ 0.55) than the sounds of oboe (*OBO*) (4.1 $\pm$ 0.55) and male voice (*MALE*) (3.7 $\pm$ 0.56).

Category descriptions given by participants were used to evaluate category identification. Table 3 shows that performance was similar for both participant groups with mean values of 68% for NHL and 61% for CI users, whilst ANOVA analysis showed no significant differences across the participant groups or category type. Fig. 1 provides further evidence for the similarity of category identification by demonstrating a positive correlation between the two participant groups (spearman rank correlation r = 0.55, p = 0.0273). Category identification of individual sounds is similar for both participant groups apart from the sounds of coughing, xylophone and oboe (*CGH*, *XLY* and *OBO*) which are less accurately categorized by CI users.

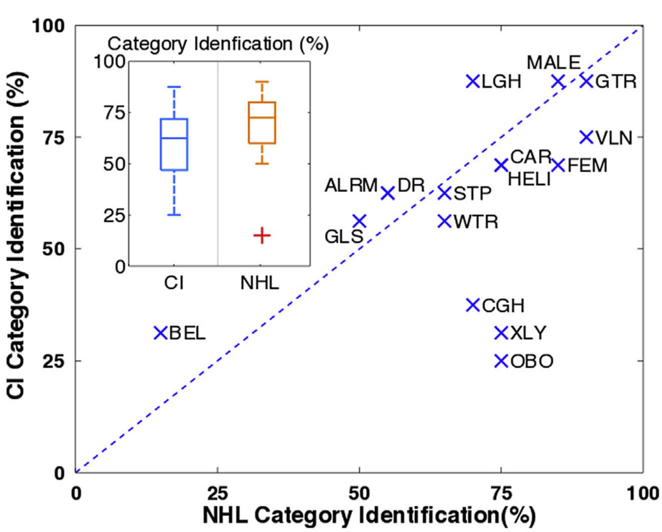

Fig. 1. Line graph of category identification – Category identification of NHL (x-axis) and CI users (y-axis) is plotted for all sixteen stimuli as labeled. The box plot embedded within the figure shows the overall data for both participant groups, with the median value, 25% and 75% percentiles, whiskers for most extreme non-outliers and finally outliers shown as red crosses (+). The line of equal categorization accuracy between CI users and NHL is plotted with the dashed blue line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Categorization accuracy was significantly greater for NHL compared to CI users, 39% vs. 28% (Kruskall-Wallis, $p < 0.05$). Values are relatively low likely due to inaccurate judgment of the predefined categories, for example the misallocation of the church-bell sound (*BEL*). This sound was considered by the experimenters as environmental because it corresponds to a typical sound heard in a traditional countryside. However it was paired with environmental sounds by only 10% of NHL and CI users, instead most of the time being paired with musical sounds. Category descriptions for *BEL* also referred to musical characteristics in 70% of NHL and 56% of CI users. Fig. 1 shows that category identification of *BEL* as an environmental sound is only 15% for NHL and 31% for CI users. When removing *BEL* from the calculations of categorization accuracy results are raised for both NHL and CI users 45% and 31% respectively, which are again significantly different (Kruskall-Wallis, p = 0.038). Overall these results show that NHL perform the task more accurately i.e. in stronger agreement to the predefined categories of environmental, musical and vocal sounds. However a certain number of CI users also follow this strategy of categorization and show clear similarities to NHL as evidenced by the example of the church-bell sound.

Using the category descriptions given by participants', analysis was also carried out to determine the percentage of comments that referred to either semantic or acoustic perceptions. Table 4 shows that semantic descriptors were used by NHL in 83.9% of cases and by CI in 67.2%, whilst acoustic descriptions were used by NHL for 15.3% of cases and by CI in 25.4%. Whilst the pattern of results is similar for both groups, when comparing the data across individual cases there are significant differences between NHL and CI for both semantic (Kruskall-Wallis, $p = 0.002$) and acoustic descriptions (Kruskall-Wallis, $p = 0.034$). This shows that the perception of sounds is different between NHL and CI users. Finally, assuming that the use of acoustic descriptions is a result of an inability to identify sounds the difference in results may be a linked to the poorer identification ability of CI users.

3.2. Categorical analysis

3.2.1. Hierarchical clustering

The overall categories produced by both participant groups are displayed in the dendrogram output of HCPC analysis in Fig. 2.

Based on the change in inertia (as described in the Methods and Materials under Categorical Analysis) six final categories were found for both CI users and NHL. However, from individual subset data the average number of categories is slightly lower for NHL compared to CI users (5.4 $\pm$ 1.4 and 6.4 $\pm$ 1.6 respectively) although

Table 4
Percentage of participants comments that refer to either semantic or acoustic attributes. Values have been calculated by scoring the comments for semantic or acoustic descriptions, summating across participants and giving a percentage for each sound and then finally averaging across all sounds. For example 83.9% of comments given by NHL referenced concerned semantic descriptions.

Subject group	Category description		
	% Semantic	% Acoustic	% Other
NHL	83.9	15.3	0.8
CI	67.2	25.4	7.4

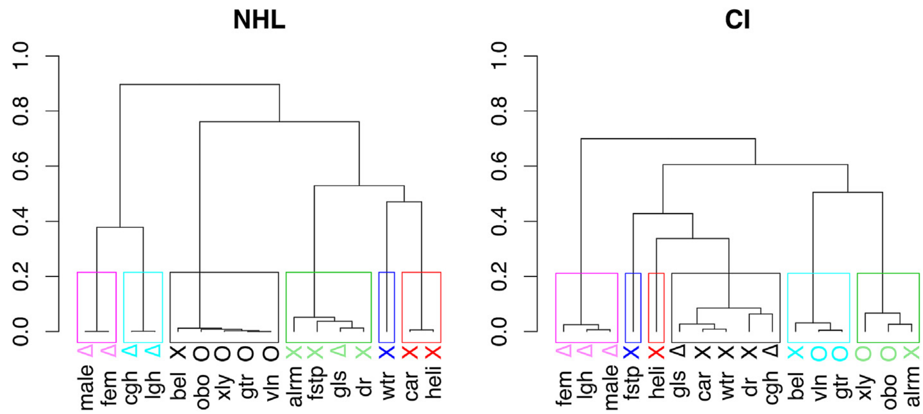

Fig. 2. Hierarchical clustering trees show the arrangement and categorization of stimuli for NHL (left panel) and CI users (right panel). Overall categories outputted by the hierarchical clustering (HCPC) are indicated by colored rectangles whilst the upper limit of these rectangles indicates the point at which each tree has been cut (as described in the main text). The height axis gives the perceptual distance between each stimulus whereby a large height indicates that participants deemed those two stimuli to be highly dissimilar and vice versa. Finally stimuli are labeled using the abbreviated sound ID's from Table 2 and symbols representing the predefined categories of environmental ('X'), musical ('Δ') and vocal sounds ('O').

this does not reach significance (two-tailed *t*-test $p = 0.068$). Further, Cophenetic Correlation Coefficients (CpCC) were calculated as 0.77 for NHL and 0.79 for CI users and demonstrate that the dendrograms represent well the initial data which is therefore similar.

Whilst the dendrograms do not directly show the three categories of environmental, musical and vocal sounds both participant groups produce categories pertaining to these descriptions. For example both groups produce categories of vocal sounds which are clearly separated from the others at heights of 0.897 for NHL and 0.7 for CI users. NHL also separate the linguistic (*MALE*, *FEM*) and non-linguistic vocal sounds (*LGH*, *CGH*), at a height of 0.379 on the dendrogram, where as CI users create only one group of vocal sounds (*MALE*, *FEM*, *LGH*) and place the sound of a male coughing (*CGH*) amongst environmental sounds.

Category descriptions associated with this sound were split evenly (33%) between vocal, environmental (described as "*noise*") and musical (described as "*rhythmic*") highlighting an unclear consensus of CI users when categorizing this sound.

Both participant groups create categories of musical sounds although CI users distinguish these further into two groups. The first contains the bell, guitar and violin sounds (*BEL*, *GTR and VLN*) which include multiple pitches whilst the second contains the oboe, xylophone and alarm (*OBOE*, *XLY* and *ALRM*) which are sounds containing only one sustained pitch. Environmental sounds are also grouped together for both participant groups although NHL produce a category of vehicle sounds containing the car and helicopter (*CAR* and *HELI*) and also place the water sound (*WTR*) on its own. In contrast CI users place all environmental sounds together except the footstep (*FSTP*) and the helicopter (*HELI*) sounds.

The ordinate of Fig. 2 shows that inter-category distance is greater for NHL (*mean* $= 0.76 \pm 0.19$) than for CI users (*mean* $= 0.59 \pm 0.12$) (two-tailed *t*-test $p = 0.006$). This reflects a more homogeneous strategy of categorization in NHL where categories are more strongly separated. When comparing the intra-category values these are smaller for NHL (*mean* $= 0.017 \pm 0.018$) compared to CI users (*mean* $= 0.034 \pm 0.024$) (two-tailed *t*-test, $p = 0.019$). Sounds within each category are therefore more similar for NHL than CI users. In addition co-occurrence (similarity) matrices of the sounds were used to find the intra-category similarity by averaging the values of all intra-category pairs in accordance with the dendrograms (Fig. 2). Values of similarity were significantly higher (Kruskall-Wallis, $p < 0.05$) for NHL (0.65 ± 0.2) than for CI users (0.4 ± 0.16) echoes the analysis of the dendrogram intra-category distances. A value of the RV coefficient was also calculated between the two co-occurrence matrices and found to be 0.89, the same as originally calculated using only the first two MCA dimensions. Co-occurrence matrices were also plotted (Supplementary Fig. 1) and show the greater agreement of NHL participants in their categorization strategy, 23% of category pairs having value either greater than 0.95 or less than 0.05 i.e. pairs that were created by very many or very few participants. However there are no values beyond these limits for CI users such that the pattern of results shows less agreement. Similar pairs can also be observed between the two groups – GTR-VLN, BEL-GTR, FEM-MALE whilst CI users also pair together CAR-WTR and CAR-GLS, suggesting a difference in categorization strategy compared to NHL.

3.2.2. Multiple Correspondence Analysis

MCA was applied to the categorization performed by the two groups of participants in order to assess their overall categorization strategies. Analysis was restricted to the dimensions able to explain the most variance within the original data, corresponding to values of equal to or greater than 8% of the total. When taking into account these criteria, five dimensions emerged in both groups (Table 5). The amount of variance accounted for by each dimension is shown in Table 5 and is much higher for NHL than for CI users. Indeed for NHL Dim 1 & 2 account for 39% of the total variance compared to only 27% of that for CI users whilst the variance unaccounted for is less for NHL (28%) compared to CI users (45.7%). All together these data constitute another indicator as to the greater variability of categorization strategies performed by the CI users. Based on the fact that Dim 1 & Dim 2 account for the most variance of any two dimensions together they were prioritized for further analysis especially with regard to the acoustic analysis (see below).

Fig. 3 displays the Factor Map of the two principal dimensions

Table 5
Variance accounted for by each dimension of the MCA analysis for both CI users and NHL. A total of 15 dimensions were used in the analysis with the five most important retained.

Participant group	Variance accounted for (%)					
	D1	D2	D3	D4	D5	Unaccounted
NHL	20.4	18.6	12.6	11.6	8.8	28.0
CI	14.5	13.0	9.7	9.4	7.7	46.8

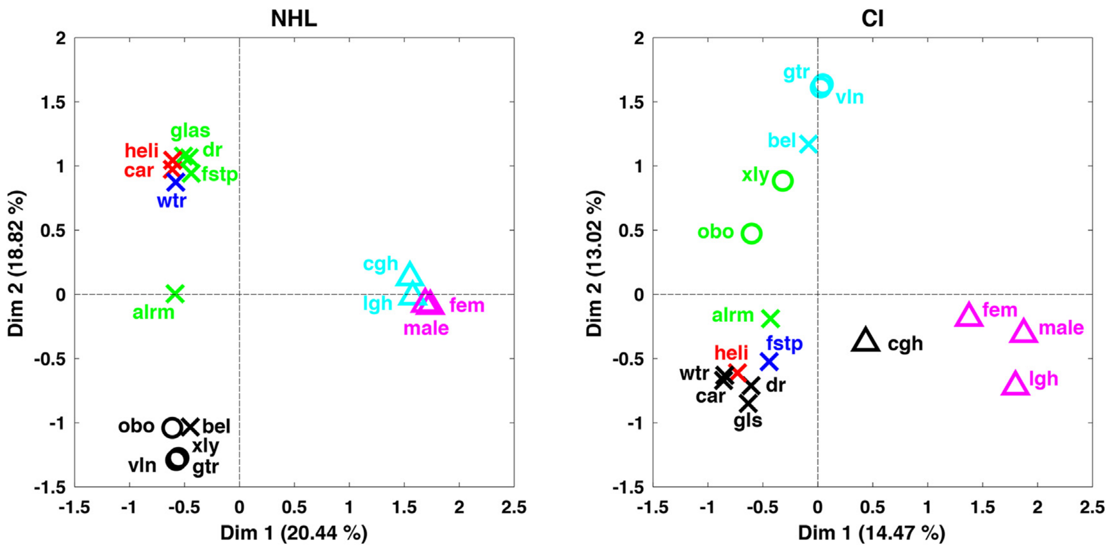

Fig. 3. Factor maps of category individuals (sound stimuli) – stimuli are plotted along the first two dimensions of the MCA analysis (Dim 1 and Dim 2) for NHL (left panel) and CI users (right panel). Analysis on each participant group was calculated separately meaning that dimensions are not directly equatable between the two figures. As noted in the differences between the x and y scales and the amount of variance covered by each dimension – given as a percentage. Stimuli that lie at zero (dotted line) are not considered to be part of the corresponding dimension e.g. the alarm sound (*ALRM*) on Dim 1. Finally the stimuli are colored according to the categories presented in Fig. 2 and are labeled by the abbreviated sound ID's given in Table 2 symbols representing the predefined categories of environmental ('X'), musical ('Δ') and vocal sounds ('O').

(Dim 1 & Dim 2) and shows three categories of sounds for both NHL and CI users. The pattern is clearer for NHL, with distinct categories corresponding to human-voice (far right), musical sounds (bottom-left) and environmental sounds (top-left). Of interest, and unexpectedly, the Factor Map for CI users presents a similar pattern although the distribution of sounds is more dispersed. For both participant groups Dim 1 separates the vocal sounds from all others. Linguistic and non-linguistic stimuli are indistinguishable although the sound of a male coughing (*CGH*) is located away from the vocal sounds by CI users in agreement with the dendrogram clustering (Fig. 2). With regards to Dim 2 there is a clear separation of musical and environmental sounds for NHL. This dimension also has little or no bearing on the categorization of the vocal and alarm sounds (*ALRM*, *MALE*, *FEM*, *LGH* and *CGH*) as they lie in close proximity to the zero-line. Even though Dim 2 also separates the musical and environmental sounds for CI users, it now plays a role in the categorization of all sounds suggesting that Dim 2 cannot be interpreted in exactly the same way for both participant groups.

To more precisely assess the similarities of categorization an RVc score was calculated using the five retained dimensions (Table 5) and was found to be 0.69 ($p < 0.001$) whilst for only the first two dimensions, 0.89 ($p < 0.001$). These results show similarities in the raw data sets and emphasize that the performance of CI users is very similar to that of NHL.

To assess the homogeneity of categorization, analysis was performed on the factor maps derived from the two populations of participants (Fig. 4). Similar to that obtained from Fig. 3, the results are more concise for NHL who present a strong uniformity with respect to the use of Dim 1 & 2. Indeed, 86% of NHL have a value of 0.8 or greater for Dim 1 and 66% for Dim 2. Two participants (ACE16 and JRE06) are quite atypical and appear opposed to this common strategy. Although it is not possible to know how each sound was identified an idea of the category perception can be garnered from the category descriptions. These include words such as "*happiness, destruction, leaving,, morning*" and suggest that the two participants tend to group the sounds based on emotional or ambiguous criteria, rather than semantic information related to the sound producing object.

CI users present a more dispersed pattern of results (Fig. 4) with only 50% of participants having a value of 0.8 or greater for Dim 1 and 31% for Dim 2. Unlike NHL this shows that fewer CI users are using the first two dimensions and are instead likely using varying categorization strategies. Some participants are again outside the main repartition, for example participant MATARL, who did not use any voice-related descriptions for any of the four vocal sounds. This is in contrast to other CI users and lends weight to the interoperation of Dim 1 being associated with the perception of vocal sounds. In order to have a stronger indication of the functional interpretation of each dimension the data was re-analyzed using only those participants who had a value greater than 0.8 (see Fig. 4). For Dim 2 this increased the amount of variance accounted for from 13.03% to 15.45% and had the effect of making the cluster containing musical sounds more distinct. When the same technique was applied to Dim 1 the variance accounted for rose from 14.47% to 15.17% and the separation between vocal sounds and other stimuli became clearer, thus confirming the initial interpretations of the first two dimensions.

3.3. Acoustic analysis of sounds

In order to understand the influence of specific acoustics characteristics on the categorization of the stimuli, multiple correlation analysis was performed between the results of the MCA and acoustical measurements as selected from Gygi et al., 2004. Table 6 shows the most significant correlations to Dim 1 & Dim 2.

For both participant groups Dim 1 & 2 show correlations to a selected set of nine acoustical variables. The only differences being the Cross-channel correlation, Mean Pitch Salience and the number of peaks (as defined in the methods section) in the waveform.

The most strongly correlated variables are illustrated as box plots for three broad categories of sounds (Fig. 5) which are obtained from the dendrograms (Fig. 3) and which correspond to environmental, musical and vocal sounds. The clearest distinction between these three groups comes in the analysis of the Mean Pitch Salience and Mean Autocorrelation Peak (Fig. 5 panels A & C). As would likely be predicted musical sounds have a higher value of

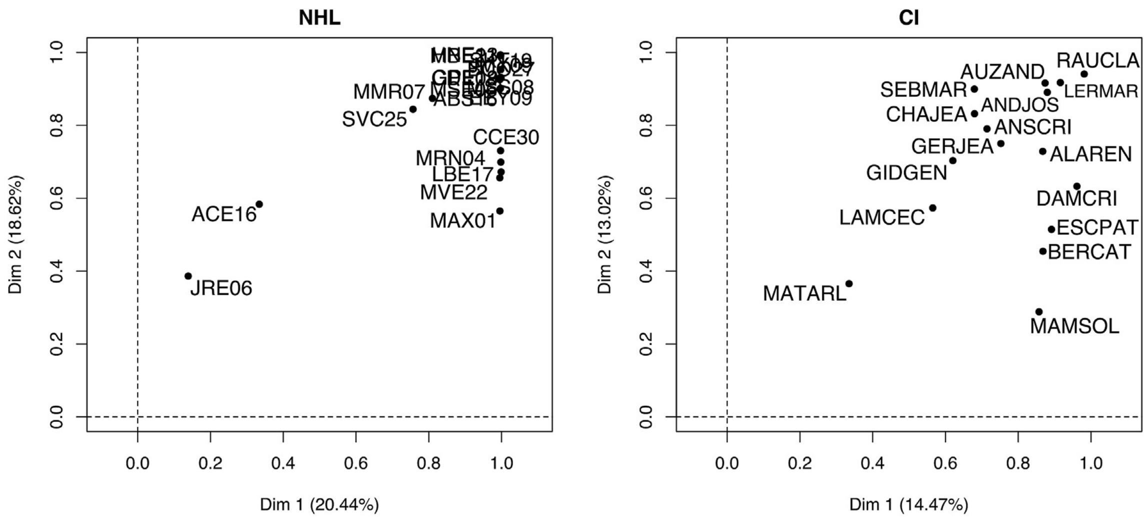

Fig. 4. Factor maps of category variables (participants) – similar to Fig. 3, however this time participants involved in the study are plotted along the first two dimensions of the MCA analysis. This is done to show how strongly each participant adheres to the use of a particular dimension with high values indicating a strong adherence. Percentages are again given to show much variance is covered by each dimension.

Table 6
Acoustic variables that show the strongest correlation to the coordinates of Dimension 1 & 2 with significant correlations in bold. It should be noted that some r-values were initially negative but have been written without sign due to the scale of the axis used for MCA analysis (see Fig. 2) which is in fact arbitrary and as such does not influence the relationship between the acoustic values and the MCA dimensions.

Acoustic Variable	Correlation (r & p values)			
	CI-D1	CI-D2	NHL-D1	NHL-D2
Autocorr. Peak (Mean)	0.13, 0.62	**0.64, 0.008**	0.26, 0.34	**0.72, 0.002**
Autocorr. Peak (std)	0.14, 0.61	**0.64, 0.008**	0.27, 0.31	**0.71, 0.002**
Waveform Peak (Mean)	0.45, 0.083	0.08, 0.77	0.49, 0.056	0.015, 0.95
Mean bpf	**0.63, 0.01**	0.35, 0.19	0.4, 0.13	0.44, 0.09
Pitch Salience (mean)	0.068, 0.8	**0.72, 0.002**	0.13, 0.62	**0.84, $p < 0.001$**
Pitch Salience (max)	0.27, 0.32	0.45, 0.08	0.055, 0.84	**0.55, 0.027**
RMS 150 Hz	**0.53, 0.036**	0.17, 0.53	**0.61, 0.012**	0.11, 0.69
RMS 250 Hz	0.064, 0.81	**0.66, 0.005**	0.21, 0.42	**0.64, 0.008**
RMS 350 Hz	**0.47, 0.064**	0.27, 0.31	0.42, 0.11	0.22, 0.42

mean pitch salience (median NHL = 0.94, CI = 0.84) than environmental sounds (NHL = 0.34, CI = 0.34). The value of the mean autocorrelation peak is also higher for musical sounds (NHL = 0.043, CI = 0.039) than both environmental (NHL = 0.0078, CI = 0.0078) and vocal sounds (NHL = 0.01, CI = 0.009). This pattern could indicate that both participant groups distinguish the sounds based on the pitch salience – the strength of pitch; and the value of the autocorrelation peaks – a measure of the uniformity of repetitions within the sound.

Dim 1 coordinates are also correlated with the RMS at lower frequency bands centered at 317 and 400 Hz in both NHL and CI patients and with 150 Hz in NHL (approaching significance for CI users, see Table 6). Importantly the fundamental frequency of the human voice is on average 85–180 Hz for males and 165–255 Hz for females (Baken, 1987). Linguistic sounds used in this study (MALE & FEM) have mean frequencies of 134 and 196 Hz and also RMS values in the 150 Hz band of 81 dB & 75 dB respectively. The correlation of the RMS at lower frequency bands therefore lends support to the idea that Dim 1 is associated with discriminating between vocal and non-vocal sounds. The correlation of Dim 2 to the RMS at 250 Hz is likely caused by the fact that within this frequency band the musical sounds *XLY*, *GTR*, *BEL* and *VLN* have high RMS values of 80, 90, 83 and 80 dB respectively. Average RMS values across the lower frequency bands (Fig. 5D) do not mirror the pattern of results along Dim 1 (Fig. 3) which clearly separates the vocal stimuli from the other sounds. Fig. 5D shows vocal sounds have a narrower range of RMS values that also overlap with the musical and environmental sounds such that the RMS in these frequency bands is likely not the only method of distinguishing the vocal sounds.

With regards to Dim 3–5 there are few significant correlations for NHL – only Dim 5 is correlated to the RMS at 150 Hz, spectral kurtosis and number of bursts. The data for CI users on the other hand is correlated to multiple variables, for example Dim 3 to the number of autocorrelation peaks, the mean burst, mean peak, mean & median frequency, spectral skew & kurtosis and RMS of frequency bands centered at 450, 570 & 700 Hz. Also Dim 4 correlates to measures of the autocorrelation function. Although these correlations are not large it again shows the greater variability of results within the CI users and use of multiple categorization strategies.

3.4. Impact of CI users individual history

Additional correlation analyses were performed between the coordinates of the CI user map (Fig. 4) and the CI user data (Table 1), with the results shown in Table 7. Notably the level of speech comprehension recovery does not match the pattern of results seen in Fig. 4. There are also no significant correlations observed between MCA results and measures of speech performance (disyllabic word recognition and sentence in noise recognition). For example, participants GERJEA, BERCAT, MAMSOL and RAUCLA are the

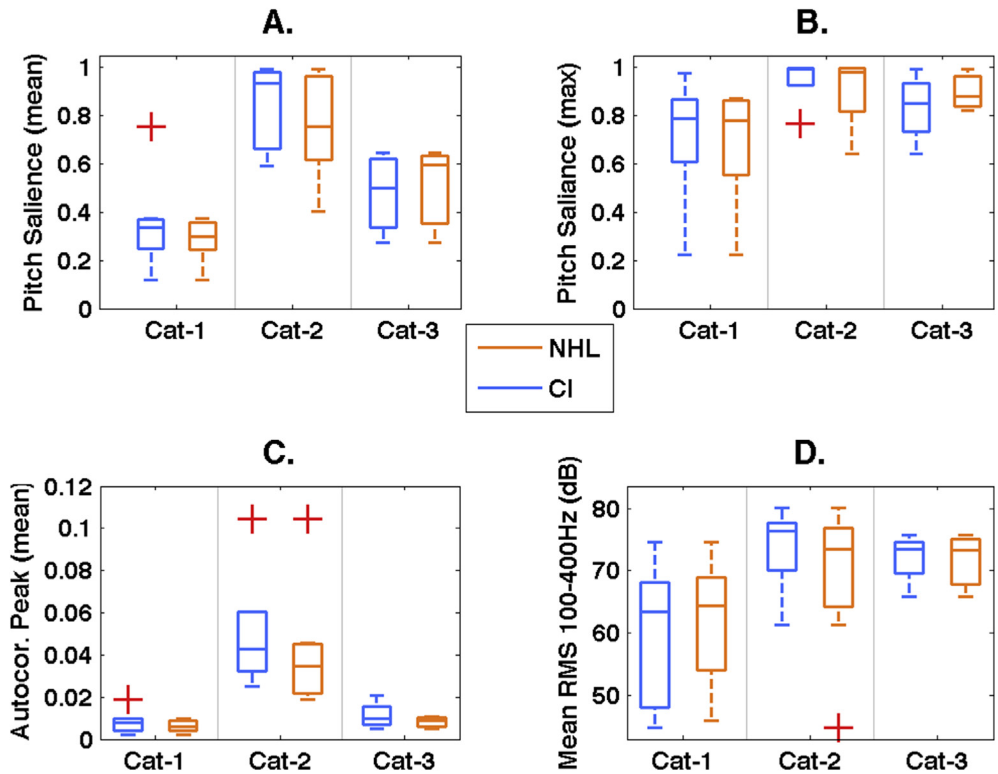

Fig. 5. Box plots for the acoustic measurements of stimuli which show significant correlation to both participant groups as highlighted in Table 6. Values are calculated for the three categories of sound demonstrated by NHL (orange) and CI users (blue) in Fig. 3, where Cat-1 corresponds to environmental sounds, Cat-2 to musical sounds and Cat-3 to vocal sounds. Box plots show the median value, 25% and 75% percentiles with whiskers the most extreme non-outliers and finally outliers shown as red crosses (+). A,B) Pitch salience (mean and max) – measured from 0 to 1 gives the strength of perceived pitch of a sound with the mean calculated as the average of all correlogram frames and max the maximum value; C) Autocorrelation Peak mean – the mean value of peaks found in the autocorrelation function of each sound; D) RMS (dB) of lower frequency bands centered at 150, 250 & 350 Hz covering a bandwidth of 100–400 Hz. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 7
Correlations of CI user data from Table 1, with other participant data as well as coordinates of each CI user from the MCA analysis, as seen in Fig. 3. Only the most significant correlations (i.e. lowest p values) are shown with those having a p value less than 0.05 shown in bold. Correlations in the table are made to the entire data set including subject LERMAR – see main text for further comments.

CI user data	Correlation (r & p values)	
	Dim 1	Dim 5
Age at testing		0.45, 0.084
Duration post implantation	0.49, 0.053	
Hearing threshold of the non-implanted ear	0.48, 0.07	
Disyllabic words score	0.49, 0.066	

poorest performers in sentence recognition in noise but show a similar use of Dim 2 to other participants. Similarly participant MATARL who shows good speech perception scores has a much lower coordinate value of both Dim 1 and Dim 2 (Fig. 4, 0.34 & 0.37 respectively). Further, the correlation analysis shows a trend for Dim 1 associated with the duration of implantation ($r = 0.456$, $p = 0.053$) and for Dim 5 with the age at testing ($r = 0.446$, $p = 0.084$). Dim 5 accounts for a small amount of the variance (7.4%) compared to the other dimensions and contrasts the sound helicopter (HELI) with other sounds. Thus the trend for an effect of age associated with Dim 5 is likely to be very weak in the present study.

4. Discussion

In the present study a Free-Sorting Task (FST) of common sounds was developed and tested with groups of normal hearing listeners (NHL) and cochlear implant (CI) users. Whilst previous studies have tested aspects of auditory categorization amongst populations of normal hearing and hearing impaired listeners this study is the first to directly test CI users and NHL within an open FST. Analysis of the categorization performed by both participant groups is carried out using MCA and HCPC and is supported by correlations with acoustic analysis and CI user data to provide more details of the categorization strategies used. Overall CI users show comparable levels to NHL in the discrimination of the three predefined categories of sounds (environmental, musical and vocal). Correlations with acoustic measurements also show that certain spectral and temporal information within the raw acoustic signal may be used by CI users to discriminate between the different categories. This includes information within lower frequency channels important for the perception of vocal sounds; spectral information (pitch saliency) used to distinguish musical sounds and temporal information (measures of periodicity) used to distinguish repetitive sounds.

4.1. Similarities of categorization processes between NHL and CI users

There are obvious similarities when comparing the results of the two participant groups, most notably in the separation of the vocal, environmental and musical sounds. Interpretations of Dim 1 & 2 strongly suggest that NHL and CI users categorize the sixteen sounds in a similar manner. However, the strategies used by the two participant groups to arrive at these results may be different. Firstly the implant delivers a poorer quality stimulus with less spectral information that CI users can use in their perception of sounds. And secondly, due to periods of deafness and adaptation to hearing with an implant, cortical reorganization experienced by CI users could affect their subjective experience of sounds, for example their degree of familiarity with previously encountered sounds and the development of familiarity with new sounds (Loebach and Pisoni, 2007).

Previous studies of auditory categorization (Gaver, 1993; Gygi et al., 2007; Lewis, 2012) suggest that categories are primarily based on the semantic information of a sound producing object/action, which follows from identification of said object/action. Less important are the acoustical characteristics of the sound which determine the qualitative perceptions of a listener, for example the pitch, the roughness or the emotional content of a sound. Indeed, the descriptions provided by participants mostly refer to source properties (e.g. "*vehicle noise*", "*noise in the house*") rather than qualitative aspects of the sounds (e.g. "*treble tones that crescendo*" or "*complementing rhythm*"). However the "acoustical" bias observed by Giordano et al. (2010), whereby sounds preceded by non-living sounds were categorized more strongly by acoustical means, suggests that in certain conditions the preference for using semantic information is altered. This could be due to the identifiably of sounds, as non-living sounds were less well identified, and could therefore have led to stronger use of the "musical listening mode" and the concentration on acoustical rather than semantic similarities. The current study would support the initial conjecture that listeners focus predominantly on the semantic source information and then secondly the qualitative acoustic information. In addition the similarity in results for categorization accuracy and category identification suggests that at a category level CI users perceive the sounds in a similar way to NHL. This means that the link between hearing the sound and the semantic representation is still strong for CI users in spite of the reduced auditory input. However, the differences in category descriptions (Table 4) would suggest that NHL have a greater ability to identify the sounds. CI users more often describe the qualitative acoustic characteristics, a process which happens when sound-source identification fails.

4.2. Dimension 1 — the importance of vocal sounds

MCA analysis has shown for both participant groups that Dim 1 is related to a distinction of vocal vs. non-vocal sounds. The correlation of Dim 1 values to the RMS of voice related frequency bands and the distinction of vocal sounds in Figs. 1 and 2 supports this idea and is therefore most likely the first and most important distinction that participants are making.

A similar free-sorting approach in NHL (Gygi et al., 2007) found a separation of vocal vs. non-vocal sounds as well as clear categories of animal vocalizations and transport/mechanical sounds. Alongside the present study this appears to contradict previous studies that showed difficulties in vocal vs. non-vocal sound discrimination for CI users (Fu et al., 2005; Molin et al., 2005; Gonzalez and Oliver, 2005; Kovacic and Balaban, 2009; Luo et al., 2007). Massida et al., 2011 also showed that when spectral information is reduced both NHL and CI users present a strong impairment in discriminating vocal from environmental sounds. Although our results show CI users perform similarly to NHL, previous protocols used a two-alternative forced choice paradigm involving much shorter stimuli. In the present study sounds were 2–3 s long and participants could listen multiple times such that identifying the sounds was likely easier.

Dim 1 coordinates are also correlated to the RMS within the frequency band centered at 150 Hz, which overlaps the range of human voice fundamental frequencies. The highest values for RMS at 150 Hz are seen for FEM and MALE and therefore give further evidence that Dim 1 is related to the perception of vocal sounds. Of course the human voice is also special due to its importance for social communication (Belin et al., 2004) such that most humans hear speech frequently in their everyday lives and are highly familiar with it. In addition speech and other vocal sounds are produced by a single unique source, the human vocal tract, which reduces the variability of spectra compared to environmental and musical sounds. Such particularity might confer to vocal sounds a specific familiarity feature set used to build-up perceptual strategies and leading to more efficient recognition (Kidd et al., 2007). Finally with regards to CI users most research is conducted with the aim of improving speech perception such that implants are designed to deal better with vocal rather than musical or environmental sounds.

4.3. Dimension 2 — pitch and periodicity cues

Dimension 2 contrasts the musical and environmental sounds for both NHL and CI users. Values are also correlated with the Max Pitch Salience and Autocorrelation peak, suggesting an importance of specific spectral-temporal information in discriminating these sounds. Similar results have also been found by Gygi et al. (2007) and Inverso and Limb (2010), whilst Reddy et al. (2009) concluded that the categorization of environmental sounds was linked to the variation in rate of spectral dynamics. High values of pitch salience correspond to slowly changing spectral cues (Terhardt et al., 1982) and may therefore be more easily perceived by CI users who normally perform poorly in pitch related tasks (Pressnitzer, 2005; Geurts and Wouters, 2001; Marx et al., 2014). CI users also retain a certain amount of musical perception when presented with simple rhythms of notes (Lassaletta et al., 2008; Looi et al., 2008). Musical sounds with temporal movement and a sense of melody may therefore have a stronger sense of musicality than singular tones. This may explain why the violin, guitar and bell sounds are grouped together separately from the oboe and xylophone. In addition difficulties in timbre perception (Inverso and Limb, 2010) and increased causal uncertainty may contribute to the categorization of the oboe and xylophone as environmental sounds. Finally Reed and Delhorne (2005) reported a link between temporal and spectral cues used for both speech and environmental sound perception. Although Dim 2 is linked with certain spectro-temporal variables there is no correlation to the speech perception scores of the CI users (Table 1) suggesting that different spectro-temporal information may be important for speech and environmental sound perception. This would suggest that although the perceptual differences originate from low level processing abilities, the effect of higher order cognitive processes cannot be dismissed (Shafiro et al., 2011).

4.4. Variability of categorization in CI users

One of the main differences between participant groups is the larger variability and looser segregation of categories produced by CI users. This suggests a greater diversity of categorization strategies are being used, as revealed by Fig. 4. A likely contributor to this is the variability in CI users listening experience, which plays an important role in the identification of sounds (Ballas, 1993). Although only experienced post-lingually deaf CI users were included in the testing, their familiarity with different sounds may vary greatly due to differences in the duration of deafness, the level of rehabilitation and everyday listening. Further, none of the speech performance scores were correlated to Dim 1 or Dim 2 suggesting that additional cognitive factors such as a decline in the cognitive abilities of participants may also contribute to the variability of results (see Gygi and Shafiro, 2013).

Our previous work aimed at speech perception in CI users highlighted the impact of brain plasticity before and after implantation on the outcomes of a cochlear implantation (Strelnikov et al., 2010; Lazard et al., 2014). For example, the temporal brain area which is specifically involved in human voice processing (see Belin et al., 2000, 2002) is weakly activated in CI users (Coez et al., 2007) probably because this area is involved in processing visual speech information during deafness (Rouger et al., 2012), a cross modal reorganization that negatively impacts on recovery through the cochlear implant (Strelnikov et al., 2013). Concerning the brain network specifically involved in environmental sound recognition (see Lewis et al., 2011; Lewis, 2005; Giraud and Price, 2001), there is no indication how a prolonged period of deafness may alter the functional integrity and recovery via the cochlear implant. However the possibility of a deleterious functional reorganization of this network should be envisaged as an important factor influencing sound categorization.

4.5. Methodological considerations

Using a Free-Sorting Task and a combination of MCA and HCPC analysis, our results provide evidence that CI users present categorization strategies that are much more similar to NHL than would be expected. However, results may not entirely reflect the abilities of CI users to discriminate natural sounds in real-world environments. Firstly our conclusion is based on a restricted choice of 16 sounds that belong to 3 predefined categories (environmental, musical and vocal). The limited number of stimuli was chosen such that the FST could be completed in a short time frame and not be fatiguing for the elderly CI users. It was therefore difficult to represent all environmental, musical and vocal sounds that exist in everyday life. Using only three categories may also have made it easier for CI users to produce results that resembled those of NHL. Testing more sounds in a FST would also be possible but would however take a longer amount of time and constitute a higher cognitive load, especially for CI participants. Apart from testing more sounds another goal of future testing would be to increase the diversity of the auditory categories to be studied. For instance no examples of animal vocalizations have been used and results also show that categories of linguistic, non-linguistic and transport/mechanical sounds likely exist. Secondly the FST was chosen over other methods of data collection in order to minimize the complexity of the test protocol. However in a recent comparative article on methods of categorical data collection, Giordano et al. (2010) shows that in comparison to hierarchical sorting and similarity ratings the results of FST have low reliability (repeatability) when applied to different groups of participants. It is also claimed that FST is inaccurate in representing the *raw* performance of individuals. Further the use of MDS models may not cover all of the variance present in the data and may hide some of the similarities and differences between NHL and CI users. In order to strengthen our conclusions we present converging results using complementary methods HCPC and the more in-depth MCA. Whilst HCPC is able to display the overall categories in a simple manner MCA allows us to more precisely interpret categorization strategies as well as the agreement of participants to said strategies. Raw co-occurrence (similarity) matrices were also used to perform a model-free analysis. Again, this revealed performance values comparable between NHL and CI users and supports our conclusions. Finally the results of categorization accuracy and identification have highlighted that the sound *BEL* was categorized as a music sound. In contrast this sound was originally considered by the researchers to belong to the category of environmental sounds as church bells are often heard outside as part of a complex sound environment. This is an important example to remember when considering the possible sounds and predefined categories to be used in future testing. Finally when measuring the identification of sounds there are many different paradigms that can be used and which may not entirely reflect the identification abilities of listeners in a FST. In the case where descriptions of sounds have been judged, the results also depend on the interpretation of the experimenter(s) and the consistent use of the same evaluation criteria.

5. Conclusions

This study has demonstrated that in a Free Sorting Task, CI users and NHL built-up three categories of sounds which matched our predefined choice of musical, environmental and vocal sounds. Performance of NHL and CI users showed similarities in the accuracy and strategy of categorization as well as subtle differences concerning CI users' ability to categorize certain sounds. It is most likely that these differences are a result of CI users' limitations in auditory perception and a poorer ability to identify the sound source. Our results are in line with the general theory that sounds are categorized primarily based on semantic information associated with the sound source and that qualitative information related to the acoustic signal become important only when source identification fails. In analyzing the categorization strategies used by participants the importance of vocal sounds has been identified. Acoustic analysis has also highlighted specific spectral and temporal variables, notably the pitch salience and the mean autocorrelation peak value, which are important for the perception and categorization of the sixteen common sounds used in this study.

Author's contribution

EC and MM contributed equally to this work.

Acknowledgments

We would like to thank all participants for their time as well as the staff at the ORL Hôpital Purpan for helping to accommodate and arrange testing, especially ML Laborde, M Tartarye and G Iversenc. We also thank Pr Qannari of ONIRIS (Nantes) for his advice concerning the MCA analysis, Julien Tardieu as well as Olivier Macherey and Grace Edwards for their comments on the manuscript and Didier Caron & Julien Labourse (Advanced Bionics) for their help and support. This work was supported by a doctoral subvention to E Collett (Advanced Bionics SARL, France) and recurrent funding of the CNRS.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.heares.2016.03.007.

References

Ballas, J.A., 1993. Common factors in the identification of an assortment of brief everyday sounds. J. Exp. Psychol. Hum. Percept. Perform. 19, 250–267.

Baken, R.J., 1987. Clinical Measurements of Speech and Voice. Taylor & Francis, London.

Barone, P., Deguine, O., 2011. Multisensory processing in cochlear implant listeners. In: Zeng, F.G., et al. (Eds.), Auditory Prostheses: New Horizons. Springer, New York, NY, pp. 365–381.

Belin, P., Zatorre, R.J., Lafaille, P., et al., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312.

Belin, P., Zatorre, R.J., Lafaille, P., 2002. Human temporal-lobe response to vocal sounds. Cogn. Brain Res. 13, 17–26.

Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135.

Berland, A., Gaillard, P., Guidetti, M., et al., 2015. Perception of everyday sounds: a

developmental study of a free sorting task. PLoS One 10 (2), e0115557.
Bonebright, T.L., 1996. Perceptual structure of everyday sounds: a multidimensional scaling approach. Behav. Res. Methods Instrum. Comput. 28, 275–278.
Cabrera, L., Feng-Ming, T., Gnansia, D., et al., 2014. The role of spectro-temporal fine structure cues in lexical-tone discrimination for French and Mandarin listeners. J. Acoust. Soc. Am. 136, 877–882.
Cadoret, M., Lê, S., Pagès, J., 2009. A factorial approach for sorting task data (FAST). Food Qual. Prefer 20, 410–417.
Carr, D.B., Young, C.J., Aster, R.C., et al., 1999. Cluster analysis for CTBT seismic event monitoring. Technologies for Monitoring the CTBT. In: Proceedings of the Twenty-first Seismic Research Symposium, Las Vegas.
Coez, A., Zilbovicius, M., Ferrary, E., et al., 2007. Cochlear implant benefits in deafness rehabilitation: PET study of temporal voice activations. J. Nucl. Med. 40, 60–67.
Copeland, B.J., Pillsbury, H.C., 2004. Cochlear implantation for the treatment of deafness. Annu. Rev. Med. 55, 157–167.
Cousineau, M., Demany, L., Meyer, B., et al., 2010. What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant. Hear Res. 269, 34–41.
De Lucia, M., Tzovara, A., Bernasconi, F., et al., 2012. Auditory perceptual decision-making based on semantic categorization of environmental sounds. Neuroimage 60, 1704–1715.
Doucet, M.E., Bergeron, F., Lassonde, M., et al., 2006. Cross-modal reorganization and speech perception in cochlear implant users. Brain 129, 3376–3383.
Dubois, D., 2000. Categories as acts of meaning: the case of categories in olfaction and audition. Cogn. Sci. Q. 1, 35–68.
Friesen, L.M., Shannon, R.V., Baskent, D., et al., 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am. 110, 1150–1214.
Fu, Q., Chinchilla, S., Galvin, J.J., 2005. Voice gender discrimination and vowel recognition. J. Assoc. Res. Otolaryngol. 5, 253–260.
Gaver, W.W., 1993. What in the world do we hear?: An ecological approach to auditory event perception. Ecol. Psychol. 5, 1–29.
Geurts, L., Wouters, J., 2001. Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants. J. Acoust. Soc. Am. 109, 713–726.
Gfeller, K., Witt, S., Woodworth, G., 2002. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. Ann. Otol. Rhinol. Laryngol. 11, 349–356.
Giordano, B., McDonnell, J., McAdams, S., 2010. Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environ-mental sounds. Brain Cogn. 73, 7–19.
Giraud, A.L., Price, C.J., 2001. The constraints functional neuroimaging places on classical models of auditory word processing. J. Cogn. Neurosci. 13, 754–765.
Goldstone, R.L., 1994. The role of similarity in categorization: providing a groundwork. Cognition 52, 125–157.
Gonzalez, J., Oliver, J.C., 2005. Gender and speaker identification as a function of the number of channels in spectrally reduced speech. J. Acoust. Soc. Am. 118, 461–510.
Grant, K.W., Walden, B.E., Seitz, P.F., 1998. Auditory-visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory-visual integration. J. Acoust. Soc. Am. 103, 2677–2690.
Guastavino, C., 2007. Categorization of environmental sounds. Can. J. Exp. Psychol. 61, 54–63.
Gygi, B., Kidd, G.R., Watson, C.S., 2004. Spectral-temporal factors in the identification of environmental sounds. J. Acoust. Soc. Am. 115, 1252–1265.
Gygi, B., Kidd, G.R., Watson, C.S., 2007. Similarity and categorization of environmental sounds. Percept. Psychophys. 69, 839–855.
Gygi, B., Shafiro, V., 2013. Auditory and cognitive effects of aging on perception of environ-mental sounds in natural auditory scenes. J. Speech Lang. Hear Res. 56, 1373–1388.
Husson, F., Josse, J., Pagès, J., 2010. Principal Component Methods – Hierarchical Clustering – Partitional Clustering: Why Would We Need to Choose for Visualising Data? Technical Report from Applied Mathematics Department Agrocampus Ouest. Retrieved from. http://www.factominer.free.fr/docs/HCPC_husson_josse.pdf.
Inverso, D., Limb, C., 2010. Cochlear implant-mediated perception of nonlinguistic sounds. Ear Hear 31, 505–514.
Kaiser, A.R., Kirk, K.I., Lachs, L., et al., 2003. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. J. Speech Lang. Hear Res. 46, 390–404.
Kidd, G.R., Watson, C.S., Gygi, B., 2007. Individual differences in auditory abilities. J. Acoust. Soc. Am. 122, 418–435.
Kovacic, D., Balaban, E., 2009. Voice gender perception by cochlear implantees. J. Acoust. Soc. Am. 126, 762–814.
Lassaletta, L., Castro, A., Bastarrica, M., 2008. Musical perception and enjoyment in post-lingual patients with cochlear implants. Acta Otorrinolaringol. Esp. 59, 228–234.
Lazard, D.S., Innes-Brown, H., Barone, P., 2014. Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging. Hear Res. 307, 136–143.
Leech, R., Gygi, B., Aydelott, J., et al., 2009. Informational factors in identifying environmental sounds in natural auditory scenes. J. Acoust. Soc. Am. 126, 3147–3155.
Lemaitre, G., Heller, L.M., 2012. Auditory perception of material is fragile while action is strikingly robust. J. Acoust. Soc. Am. 131, 1337–1348.
Lewis, J.W., 2005. Distinct cortical pathways for processing tool versus animal sounds. J. Neurosci. 25, 5148–5158.
Lewis, J.W., 2012. Auditory object salience: human cortical processing of non-biological action sounds and their acoustic signal attributes. Front. Syst. Neurosci. 6, 1–15.
Lewis, J.W., Talkington, W.J., Puce, A., et al., 2011. Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. J. Cogn. Neurosci. 23, 2079–2101.
Liberman, A.M., Cooper, F.S., Shankweiler, D.P., et al., 1967. Perception of the speech code. Psychol. Rev. 74, 431–461.
Loebach, J.L., Pisoni, D.B., 2007. Perceptual Learning under a Cochlear Implant Simulation Spoken Language Processing Progress Report No. 28. Retrieved from. http://www.iu.edu/~srlweb/pr/PR28_p47_Perceptual_Learning_Under_a_Cochlear_Implant_Simulation.pdf.
Loebach, J.L., Pisoni, D.B., 2008. Perceptual learning of spectrally degraded speech and environmental sounds. J. Acoust. Soc. Am. 123, 1126.
Looi, V., McDermott, H., McKay, C., et al., 2008. Music perception of cochlear implant users compared with that of hearing aid users. Ear Hear 29, 421–434.
Looi, V., Arnephy, J., 2010. Environmental sound perception of cochlear implant users. Cochlear Implants Int. 11, 203–227.
Looi, V., Gfeller, K., Driscoll, V., 2012. Musical appreciation and training for cochlear implant recipients: a review. Semin. Hear 33, 307–334.
Lorenzi, C., Gilbert, G., Carn, H., et al., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc. Natl. Acad. Sci. U. S. A. 103, 18866–18869.
Luo, X., Fu, Q., Galvin, J.J., 2007. Vocal emotion recognition by normal-hearing listeners and cochlear implant users. Trends Amplif. 11, 301–315.
Marx, M., James, C., Foxton, J., et al., 2014. Speech prosody perception in cochlear implant users with and without residual hearing. Ear Hear 36, 239–248.
Massida, Z., Belin, P., James, C., et al., 2011. Voice discrimination in cochlear-implanted deaf subjects. Hear Res. 275, 120–129.
Molin, E., Leijon, A., Wallsten, H., 2005. Spectro-temporal discrimination in cochlear implant users. In: Proceedings of the International Conference of Acoustics, Speech, and Signal Processing, 2005. Stockholm, Sweden. http://dx.doi.org/10.1109/ICQSSP.2005.1415637.
Moller, A.R., 2006. History of cochlear implants and auditory brainstem implants. Adv. Otorhinolaryngol. 64, 1–10.
Pressnitzer, D., 2005. Music to electric ears: pitch and timbre perception by cochlear implant patients. Ann. N. Y. Acad. Sci. 1060, 343–345.
Qannari, E.M., Courcoux, P., Faye, P., 2014. Significance test of the adjusted Rand index. Application to the free sorting task. Food Qual. Prefer 32, 93–97.
Reddy, R.K., Ramachandra, V., Kumar, N., et al., 2009. Categorization of environmental sounds. Biol. Cybern. 100, 299–306.
Reed, C.M., Delhorne, L.A., 2005. Reception of environmental sounds through cochlear implants. Ear Hear 26, 48–61.
Robert, P., Escoufier, Y., 1976. A unifying tool for linear multivariate statistical methods: the RV-coefficient. Appl. Stat. 25, 257–265.
Rouger, J., Lagleyre, S., Fraysse, B., et al., 2007. Evidence that cochlear- implanted deaf patients are better multisensory integrators. Proc. Natl. Acad. Sci. U. S. A. 104, 7295–7300.
Rouger, J., Lagleyre, S., Demonet, J.F., Fraysse, B., et al., 2012. Evolution of crossmodal reorganization of the voice area in cochlear-implanted deaf patients. Hum. Brain Mapp. 33, 1929–1940.
Saraçli, S., Dog an, N., Dog an, I., 2013. Comparison of hierarchical cluster analysis methods by cophenetic correlation. J. Inequal. Appl. 203, 1–8.
Shafiro, V., 2008. Identification of environmental sounds with varying spectral resolution. Ear Hear 29, 401–420.
Shafiro, V., Gygi, B., Cheng, M.Y., et al., 2011. Perception of environmental sounds by experienced cochlear implant patients. Ear Hear 32, 511–523.
Shafiro, V., Sheft, S., Gygi, B., Ho, K., 2012. The influence of environmental sound training on the perceptin of spectrally degraded speech and environmental sounds. Trends Amplif. 16 (2), 83–101.
Shafiro, V., Sheft, S., Kuvadia, S., Gygi, B., 2015. Environmental sound training in cochlear implant users. J. Speech Lang. Hear Res. 58, 509–519.
Shannon, R.V., Zeng, F.G., Kamath, V., et al., 1995. Speech recognition with primarily temporal cues. Science 270, 303–304.
Slaney, M., Lyon, R.F., 1990. A perceptual pitch detector. In: Proceedings of the International Conference of Acoustics, Speech, and Signal Processing, pp. 357–360.
Strelnikov, K., Rouger, J., Demonet, J.F., et al., 2010. Does brain activity at rest reflect adaptive strategies? Evidence from speech processing after cochlear implantation. Cereb. Cortex 20, 1217–1222.
Strelnikov, K., Rouger, J., Demonet, J.F., et al., 2013. Visual activity predicts auditory recovery from deafness after adult cochlear implantation. Brain 136, 3682–3695.
Summerfield, Q., 1992. Lipreading and audio-visual speech perception. Philos. Trans. R. Soc. Lond. B Biol. Sci. 335, 71–78.
Terhardt, E., Stoll, G., Seewann, M., 1982. Algorithm for extraction of pitch and pitch salience from complex tonal signals. J. Acoust. Soc. Am. 71, 679–688.
Tyler, R.S., Parkinson, A.J., Woodworth, G.G., 1997. Performance over time of adult patients using the Ineraid or Nucleus cochlear implant. J. Acoust. Soc. Am. 102, 508.
Zeng, F.G., Nie, K., Stickney, G.S., 2005. Speech recognition with amplitude and frequency modulations. Proc. Natl. Acad. Sci. U. S. A. 102, 2293–2298.
Zwicker, E., 1961. Subdivision of the audible frequency range into critical bands (frequenzgruppen). J. Acoust. Soc. Am. 33, 248–248.