



HAL
open science

Categorisation of natural sounds at different stages of auditory recovery in cochlear implant adult deaf patients

Kuzma Strelnikov, Edward Collett, Pascal Gaillard, Eric Truy, Olivier Deguine, Mathieu Marx, Pascal Barone

► **To cite this version:**

Kuzma Strelnikov, Edward Collett, Pascal Gaillard, Eric Truy, Olivier Deguine, et al.. Categorisation of natural sounds at different stages of auditory recovery in cochlear implant adult deaf patients. *Hearing Research*, 2018, 367, pp.182–194. <10.1016/j.heares.2018.06.006>. <hal-02341980>

HAL Id: hal-02341980

<https://hal.science/hal-02341980v1>

Submitted on 15 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Categorisation of natural sounds at different stages of auditory recovery in cochlear implant adult deaf patients

Kuzma Strelnikov^{a, b, c, *, 1}, Edward Collett^{a, b, d, 1}, Pascal Gaillard^e, Eric Truy^{f, g, h}, Olivier Déguine^{a, b, c}, Mathieu Marx^{a, b, c, 2}, Pascal Barone^{a, b, 2}

^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, Hôpital Purpan Toulouse, France

^d Advanced Bionics SARL, France

^e Université de Toulouse, CLLE UMR 5263, CNRS, UT2J, Université de Toulouse Jean-Jaurès, Toulouse, France

^f INSERM, U1028, CNRS, UMR5292, Lyon Neuroscience Research Center, IMPACT Team, Lyon, France

^g University of Lyon, Lyon, France

^h Service d'ORL, de Chirurgie cervico faciale et d'Audiophonologie, Hôpital Edouard Herriot, Centre Hospitalier et Universitaire de Lyon, Lyon, France

A B S T R A C T

Previous studies have demonstrated that cochlear implant (CI) patients are more efficient at performing sound categorisation than sound identification. However, it remains unclear how this categorisation capacity develops with time during the rehabilitation period after implantation. To investigate the role of the post-implantation auditory experience in the broad sound categorisation in CI patients, we recruited CI patients with different durations of CI experience: Newly implanted CI patients (less than six months), Intermediate CI patients (6–14 months) and Experienced CI patients with a duration of implantation greater than 14 months. The patients completed a Free Sorting Task (FST), which allowed them to categorise 16 natural sounds based on their own criteria. We found an early deficit in categorisation, especially for vocal sounds; the categorisation started to improve after approximately six months post-implantation with a change of categorisation strategy which relied on different acoustic cues as a function of time after CI. The separation of the category of vocal sounds from other sounds significantly increased between the Newly implanted and Intermediate groups, i.e. as experience with the cochlear implant was acquired. The categorisation accuracy of vocal sounds was significantly correlated with the post-implantation period only in the group of newly implanted CI patients. This is the first study to show that the categorisation of vocal sounds with respect to non-vocal sounds improves during the rehabilitation period post-implantation. The first six-month post-implantation period appears to be crucial in this process. Our results demonstrate that patients in different rehabilitation periods use different acoustic cues, which increase their complexity with the CI experience.

1. Introduction

Cochlear implants (CI) are used to restore hearing in patients with profound bilateral sensorineural hearing loss (Copeland and Pillsbury, 2004), and they have led to considerable success in the

functional rehabilitation of deafness in terms of the restitution of speech comprehension ability (Møller, 2006). Given the degraded nature of the sound transmitted from the implant to the brain, there is a necessity for patients to develop a specific strategy for speech analysis, which leads to an initial period of adaptation of about six months after implant activation, before CI users start understanding speech at a socially significant level (Barone and Deguine, 2011). During this adaptive period, CI users rely strongly on visual cues as compensatory cues to decipher the impoverished auditory signal delivered by the implant (Grant et al., 1998; Kaiser et al., 2003; Rouger et al., 2007; Summerfield, 1992; Tyler et al., 1997).

* Corresponding author. Cerveau & Cognition, CNRS UMR 5549, Pavillon Baudot, CHU Purpan, BP 25202 31052, Toulouse Cedex, France.

E-mail address: Kuzma.Strelnikov@cnrs.fr (K. Strelnikov).

¹ Equal contributions.

² Equal senior co-authorship.

Since the auditory information transmitted to the brain is spectrally degraded and lacks fine temporal structure (Shannon et al., 1995), some aspects of speech comprehension also suffer, most notably the perception of prosodic information (Friesen et al., 2001; Lorenzi et al., 2006; Marx et al., 2015; Zeng et al., 2005). As a result, CI users rely more on envelope temporal cues than on spectral ones in their sound analysis strategies (Cabrera et al., 2014; Doucet et al., 2006). Emphasis has been given to the restoration of speech processing in CI patients pertaining to their social functioning; however, these patients often struggle with the perception of other auditory entities; and the recognition of environmental sounds and music may remain problematic.

Thus, given the constraints and impairments of sound perception through cochlear implants, it is important to investigate how CI users discriminate vocal, environmental and musical sounds. The acoustic cues underlying these discrimination abilities should be better identified and understood to facilitate the adaptation of coding strategies, which could potentially improve the categorisation of sounds by CI patients.

In everyday situations, we regularly encounter a large number of coexistent sounds. It would thus require high computational resources to analyse each of them separately. The ability to categorise sounds was demonstrated by several studies of phonemes, voice gender, the material and action of a sound source etc. (Belin et al., 2004; Guastavino, 2007; Lemaitre and Heller, 2012; Liberman et al., 1967). Using this ability, one would estimate the probability of a sound belonging to a certain meaningful category without going into a deeper and more demanding process of identification.

According to Gaver (1993), categorisation can involve two different strategies. The first strategy is based on the source of the sound and the event occurring with this source. For example, the source may be a material such as wood or metal, which acts to produce the sound of hitting, rubbing etc. This strategy can also relate to living or human beings and their actions. It is mostly used for “everyday listening” and corresponds to the semantic load of the sound.

The second strategy of sound categorisation resembles acoustic listening (or musical listening) and refers to the perception of qualitative aspects of the sound such as the pitch, loudness, timbre and also the emotional content (pleasant or unpleasant). Thus, these “acoustic listening” qualities are related to physical aspects of the acoustic signal rather than to the sources of the sound and related actions. Since a “everyday listening” strategy extracts information about mundane objects and their actions, it is essential in adaptation to common real life situations, and as a consequence is predominantly used (Gygi et al., 2007; Inverso and Limb, 2010). However, strategies can be influenced by the acoustic context, e.g. when sounds are preceded by non-living sounds their categorisation is more strongly based on acoustic information corresponding to the “musical listening” strategy (Giordano et al., 2010). To avoid the confusion with music perception, we will refer to this listening mode as “acoustic listening”.

As examples of commonly used auditory categories, one can cite living and non-living sounds (De Lucia et al., 2012), human and animal vocalisations, impact sounds, water sounds (Gygi et al., 2007), as well as human and traffic noise (Guastavino, 2007). Evidently, the process of categorising using the “everyday listening” strategy requires the discernment of different sound sources and a repertoire of categories of these sources and/or their actions. It follows that if a sound source cannot be recognised or attributed to a certain category of sources and actions, it would be difficult to categorise sounds using the “everyday listening” strategy.

One of the ways to evaluate how participants categorise different sounds is to analyse the categorisation of many stimuli with similarity matrices and Multidimensional Scaling (MDS)

(Bonebright, 1996; Guastavino, 2007; Gygi et al., 2007). In these studies, similarity judgments are reduced to only a few dimensions representing the entire perceptual space. Further analysis is required to interpret the various perceptual and acoustic features on which these perceptual dimensions are based (Bonebright, 1996).

Another way of assessing the ability of patients to categorise sounds is the free sorting task (FST). In the previous study of our group (Collett et al., 2016), we used FST to investigate how CI users cope with auditory categorisation while being presented with three pre-defined groups of sounds (environmental, musical and vocal), and we compared these users with normal-hearing subjects (NHS). Using Multiple Correspondence Analysis (MCA) and Hierarchical Clustering based on Principal Components (HCPC), we showed that experienced CI users followed a similar categorisation strategy to that of NHS and were able to categorise the three different types of sounds despite being impaired in their exact recognition. Of interest, in spite of the deficit regarding voice discrimination (Massida et al., 2011, 2013), experienced CI users in an FST are able to discriminate vocal sounds from the broadly defined environmental and musical sounds (Collett et al., 2016).

Our results for CI users, however, were limited to those with more than one year post-implantation experience and these patients presented a high level of recovery for speech comprehension. In addition, they showed high variability, i.e. less inter-participant agreement than NHS. Though categorisation by CI users is likely to be based more on acoustic listening than NHS, acoustic analysis highlighted the usage of such complex acoustic cues as mean pitch salience and the average autocorrelation peak for the categorisation of the sounds. The results therefore indicated that on a broad level of categorisation, 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.

The aim of this study is to investigate how CI patients progressively develop their strategies to categorise sounds with experience of CI usage, i.e. as a function of the rehabilitation period post-implantation. To investigate the role of the rehabilitation process in the broad sound categorisation in CI patients, we recruited CI patients with different durations of CI experience: Intermediate CI patients (Int CI) from 6 to 14 months and Newly implanted CI patients (New CI) less than six months. They were compared with Experienced CI patients (Exp CI) with duration of implantation greater than 14 months. The data on Exp CI originated from Collett et al. (2016) and was further analysed for comparative purposes. The patients completed a Free Sorting Task (FST) of 16 short sounds exactly reproducing the stimuli and experimental paradigm of Collett et al. (2016).

2. Method and materials

2.1. Patients

A total of 48 CI patients took part in the study (including 16 Exp CI patients from our previous study [Collett et al., 2016]). Their performance in FST was collected during regular visits to the ENT department following a standard rehabilitation program, one assessment per patient. The standard rehabilitation program in our hospital consists of a monthly visit during the first 3 months post-implantation, then every 3 months during the first year and afterwards once in a year. The patients were distributed into three groups of equivalent size based on the duration of experience with the cochlear implant. Duration of implantation in Experienced CI patients (Exp CI, N = 16) was greater than 14 months; Intermediate CI patients (Int CI, N = 15) from 6 to 14 months and Newly implanted CI patients (New CI, N = 17) had less than six months of

CI experience. We arbitrarily divided the CI subjects according to the duration of CI experience based on previous results on recovery of auditory functions through a CI (Massida et al., 2011; Rouger et al., 2007). A subset of 6 Exp CI patients were also included for a supplementary comparison of the performance of sound recognition that was not collected during the original study; these patients are described as Supp Exp CI in Table 1. For control, 20 normal hearing subjects (NHS) were taken from our previous study (Collett et al., 2016). All subjects (CI and NHS) were native French speakers with no self-reported history of neurological or psychiatric disorders. Table 1 gives a summary of the demographic and clinical properties of the groups. The duration of implantation was taken as a rounded number from the date of the implant activation to the date when the FST was performed.

Participants with a post-implantation duration of 6.5 months or more were included in the Int CI group. The duration of deafness was taken from medical records based on the date that participants were first diagnosed with a severe hearing loss. The threshold of the non-implanted ear was obtained from the participant's audiogram recorded prior to implantation and calculated as the average threshold (using headphones) across frequencies 0.25, 0.5, 1, 2, 4 and 8 kHz. Word recognition and Sentence in noise scores were recorded by the ENT audiologists as part of regular appointments with participants, and were recorded either on the same date as the FST or within a few weeks, but never exceeded the limits of implantation duration used to define the three CI groups.

The three CI groups were quite homogeneous as there was no significant difference between the groups in terms of age distribution (Kruskal-Wallis, $p = 0.0815$), nor duration of deafness (Kruskal-Wallis, $p = 0.386$), nor for the non-implanted ear residual hearing (comparisons were made after exclusion of patients with complete deafness of the non-implanted ear) (Kruskal-Wallis, $p = 0.2822$) (see Table 1 for the mean and SD values). There was a significant difference between the Exp CI and New CI in disyllabic word recognition in quiet (rank sum, $p < 0.05$); however, no significant difference was observed between the Int CI and Exp CI and between the three groups in the Kruskal-Wallis test ($p > 0.1$), probably because the Int CI group had already achieved a good recovery level for speech comprehension (see Table 1) (Rouger et al., 2007). The study was approved by the local ethics committee, and all participants gave written informed consent prior to their inclusion in the study.

2.2. Experimental procedure

We applied the same experimental procedure as described in Collett et al. (2016). The participants completed a Free Sorting Task (FST) using 16 short sounds (2–3 s duration). The sounds were taken from a database owned by the PETRA group at the University of Toulouse UT2J (<http://petra.univ-tlse2.fr>) and were chosen to cover a broad range of semantic and acoustic information. We

established three pre-defined categories that contained environmental, musical and vocal sounds (Environmental: alarm clock ringing, car engine starting, door opening, footsteps, glass breaking, helicopter, running water; Musical: bells, guitar, oboe, violin, xylophone; Vocal: male voice coughing, female voice speaking, female voice laughing, male voice speaking). All stimuli were monophonic and recorded in .wav format with a sampling frequency of 44,100 Hz.

The participant groups were tested in quiet listening rooms. 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. The stimuli were presented 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 16 sounds were represented on the computer by 16 numbered and coloured squares that were positioned in the same order for all participants.

The task for participants was to listen to the 16 sounds and place them into groups, i.e. to 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 using the PC mouse by a double click on each square and the participants created categories by dragging and positioning squares together on screen. Once participants had finished positioning the squares into categories they were asked to listen to each sound one last time to verify their choices before ending the experiment. They were then asked to enter a brief description for each category into the computer using the keyboard. The average duration of the test was 670 ± 330 (SD) sec. As a supplementary open-set identification test, after the FST the subjects were asked to listen to each sound and to provide their description of the sound.

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 16. 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.

2.3. Data analysis

To analyse the categories that the participants created, two approaches were used in R environment (Husson et al., 2014): Multiple Correspondence Analysis (MCA) and Hierarchical Clustering based on Principal Components (HCPC). These approaches permitted us to represent stimuli associations as hierarchical dendrograms (HCPC) and to obtain the group-level statistics of the

Table 1
Summary for the CI patients in three groups (means and SD).

	Defined by duration of implantation (months)	Duration of implantation (months)	Age (years)	Duration of deafness (years)	Threshold, dB	Word recognition, % correct	Phrase in silence, % correct
Exp CI	>14	56.53 ± 33.7 (SD)	56 ± 15	4.62 ± 3.74	90.17 ± 17.02	80.00 ± 12.39	83.44 ± 22.05
Int CI	juil-14	9.5 ± 2.2	45 ± 22	6.27 ± 7.55	85.63 ± 12.43	74.93 ± 22.5	86.31 ± 21.21
New CI	0–6	2.17 ± 1.21	60 ± 14	15.67 ± 17.45	67.06 ± 42.48	55.63 ± 27.62	62.87 ± 35.54
Supp Exp CI	>14	37.4 ± 18.1	34 ± 14	18 ± 13.69	99 ± 4.54	80.8 ± 27.2	90 ± 22.36

Concerning the differences in averages and SD for the duration of deafness, they are caused by the two patients in the New group who have deafness duration of 42 and 54 years while no such long durations exist in other groups. The threshold of the non-implanted ear was taken from the participants audiogram taken prior to implantation and calculated as the average threshold (using headphones) across frequencies 0.25, 0.5, 1, 2, 4 and 8 kHz.

preferred associations of stimuli (MCA).

More specifically, Hierarchical Clustering based on Principal Components was performed 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 by using this process that 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., 2014).

Multiple Correspondence Analysis was applied to the multi-subject categorisation table (raw data not included in the article) outputted by the TCL LabX software. This table 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 dataset, 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 (Cadoret et al., 2009). A total of fifteen dimensions were used in the analysis. The two most significant dimensions (Dim 1 & Dim 2) were 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. 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).

To characterise distances between the sounds in MCA maps, the corresponding Euclidean distances were calculated. 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 participant groups. Six different acoustical domains were explored:

- pitch measures (mean and median fundamental frequency (F0), standard deviation of fundamental frequency (F0), maximum of fundamental frequency (F0), mean fundamental pitch salience, max fundamental pitch salience, mean BPF (band pass frequency), which reflects the mean correlation between octave frequency bands of the sound envelope (a higher value therefore corresponds to sounds that are more uniform in their envelope response);
- spectral measures (centroid, skew, kurtosis, mean centroid, spectral centroid velocity, spectral centroid uniformity and spectral centroid standard deviation);
- temporal-envelope statistics (number of peaks, mean peak, number of bursts, mean burst, total burst duration, duration ratio computed from the broadband temporal envelope of each sound);
- periodicity measures (number of peaks in the envelope autocorrelation, maximum autocorrelation peak, mean and standard deviation of autocorrelation peaks, range of data); cross-channel correlation;
- raw root-mean-square (RMS) measures.

Fundamental frequency (F0) is a correlate of the 'periodicity

pitch' perceptual phenomenon, for brevity we will sometimes refer to these features as frequency, max frequency etc. Detailed descriptions of these acoustical properties are provided in Collett et al. (2016), the acoustic analyses were identical to Gygi et al. (2007) using their Matlab scripts. Along with the HCPC and MCA analyses, we also assessed the global performance of the subjects in the FST: test duration, number of playbacks for sounds, identification of sounds, the number of created categories and category identification with respect to predefined categories.

3. Results

3.1. Global performance

There was a difference in the duration of FST, such that New CI required more time to accomplish the test than Int CI, but the duration of the test was similar between Exp CI and Int CI ($p < 0.05$): 539 ± 275 s (SD) in Exp CI; 545 ± 213 s in Int CI; 922 ± 484 s in New CI. This difference was probably due to the New CI participants having the lowest amount of hearing experience with CI, and thus they found the task more difficult. Such time to perform the task is also much higher in each group of patients compared to NHS (mean 330 s, see Collett et al., 2016). In consequence, the between-group difference was significant when comparing the number of playbacks (Kruskal-Wallis, $p < 0.05$) with the highest number of replays observed in the New CI group. However, no between-category difference for the number of playbacks was detected in New, Int and Exp CI groups (Fig. 1A) (while in each group there was a tendency for voice stimuli to be listened to less frequently and environmental sounds more often).

3.2. Sound and category identification

In the present study, the open-set identification of our stimuli was tested in the New and Int CI groups but not in the Exp CI group due to organizational reasons (a high load of patients with other tests during the visit). We checked the identification data on 6 additional patients with similar CI experience (Supp Exp CI). The correct identification score in Supp Exp CI was $30 \pm 17\%$ (SD), in Int CI $35 \pm 29\%$, in New CI $33 \pm 30\%$, the Kruskal-Wallis test revealed no difference between the groups ($p > 0.9$).

In terms of categorisation strategies, no significant difference was found between the three groups in the number of categories that were created (6.4 ± 1.6 in Exp CI; 5.4 ± 1.5 in Int CI; 7.18 ± 2.5 in New CI). To facilitate the interpretation of the results of the HCPC and MCA analyses and thereby explore the categorisation strategies used by participants, we computed a measure of category identification. Categorisation accuracy was inferred from the participants' category descriptions. For each sound, the associated category description was evaluated as to whether they referenced musical, vocal or environmental sounds. For example, the sound described as "some instrument" was put into the musical category, "a person" into the vocal category and "something moving" into the environmental category. These were then compared to the predefined categories and the percentage of comments in agreement was calculated across all participants for each sound. As a result, a significant difference between the three CI groups concerning category identification was found, (Kruskal-Wallis, $p < 0.05$) where the mean value was lower for New CI (42.6% correct) compared to Int CI (55%) and Exp CI (60%). In (Collett et al., 2016) no difference was found between Exp CI and controls who had an average of 68% for categorisation identification.

The pattern of results per category was somewhat different across the three participant groups. For Exp CI, the results between categories were significantly different (Chi-squared test, $p < 0.05$)

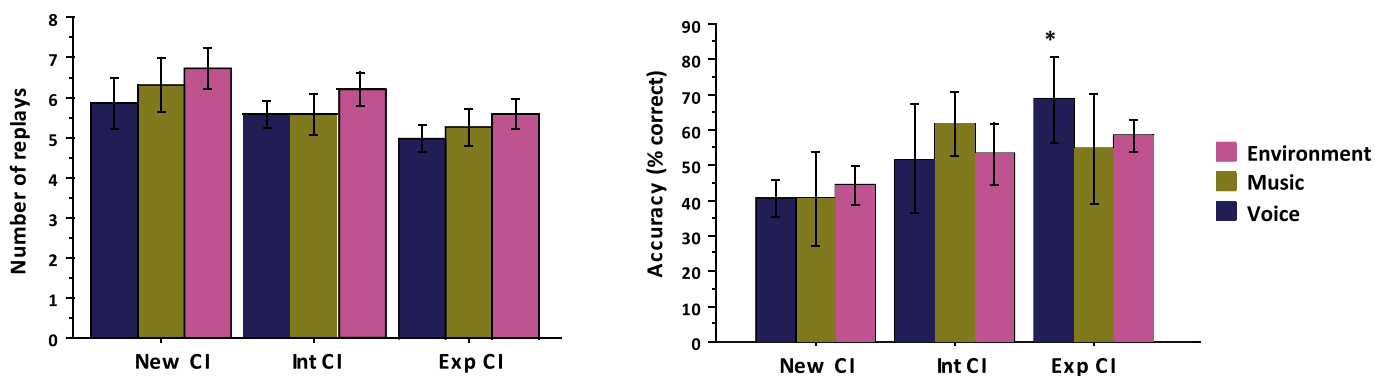


Fig. 1. Number of replays and accuracy of categorisation in CI patients.
 A. Numbers of replays between the groups of CI patients per sound category.
 B. Accuracy of categorisation between the groups of CI patients per sound category.
 Error bars indicate standard errors. Voices are more accurately categorised in Exp CI patients compared with Environment and Voice categories ($p < 0.05$).

the highest category identification being for vocal sounds (70%). For Int and New CI the difference between categories was not significant. A tendency in Int CI was observed for the musical sounds (62%) and New CI had a tendency to be slightly better for categorisation of environmental sounds (59%) than for other categories (Fig. 1B). The only significant result in Exp CI for voice stimuli would suggest better recognition of vocal sounds in relation to other categories; vocal recognition improves with longer duration of implantation.

3.3. Hierarchical clustering

The first representation of the categorisation strategies employed by the participant groups is given by the dendrograms in Fig. 2A, where the different coloured rectangles represent the final categories. Importantly, this is only an average representation based on the categories created by all participants in each group, such that some participants may have only weakly followed this representation and some not at all. To measure how well the trees reflect the original data, CCC (Cubic Clustering Criterion) values were calculated and show that the least representative dendrogram is that created for the New CI (CCC = 0.69). This constitutes a first indication of a higher variability of categorisation of the New CI patients. However, both Int and Exp CI groups presented high and similar CCC values, respectively of 0.83 and 0.80 (see Collett et al., 2016).

While the same number of categories (six) are seen on the dendrograms for all participant groups, this is somewhat dependent on the number of dimensions pre-defined in the analysis and does not suggest in itself strong evidence for similarities among the participant groups. The dendrograms show that HCPC analysis produces a category of vocal sounds containing male and female voices and laughter for all three participant groups (see Collett et al., 2016 for the Exp CI results). To compare the different groups, a similar level of distinction can arbitrarily be chosen (height of the dendrograms), the higher values corresponding to a better clustering compared with the lower values for the other categories. Comparing the Exp CI data (Collett et al., 2016) with Int CI data (Fig. 2A), one can remark that in the two groups, cutting the dendrograms at a low height (at about 0.2) results in only three large categories and the results have a tendency to correspond to the three pre-defined categories of musical, vocal and environmental sounds, though two different stimuli in each group are “abnormally” categorised. The results for New CI fail to exhibit this tendency. Here, following a cut at low height (but over 0.3) the

dendrogram creates four categories, such that the sounds of door and footsteps are distinct from other environmental sounds. In NHS (see Collett et al., 2016) the dendrograms clearly show the three categories of environmental, musical and vocal sounds but NHS also separate the linguistic and non-linguistic vocal sounds.

3.4. Multiple correspondence analysis

A more detailed breakdown of the categorisation performance can be given with MCA analysis (Fig. 2B). In this analysis, the original data is represented in a multidimensional Euclidean space. Table 2 provides information on the five first dimensions. As one can notice in Table 2, the higher the dimension, the lower the amount of variance it explains and the less it is used by the subjects. For clarity, we have retained the first two dimensions, which were used by more than 50% of patients and in which the separation of voice stimuli was the most evident. It is noteworthy that the five dimensions in sum cover only about 50% of the total variance in free sorting. This amount of the explained variance was not significant between the groups (chi-squared test, $p > 0.5$). It is important to mention that in NHS the first five dimensions represent over 70% of the explained variance (Collett et al., 2016), adding further indication of a larger variability in CI patients.

In New CI (Fig. 2B), Dimension 1 separates environmental sounds, to the right, from the musical and vocal sounds together. However, some sounds are rather poorly separated in this dimension: oboe, cough and laugh are the closest to environmental sounds. Dimension 2 separates the sounds of door and footsteps from the other environmental sounds. Thus, one can see that voices and musical instruments are put together into the same group, because New CI subjects fail to separate them.

The factor maps for the Int CI participant group (Fig. 2B) show similarities to that of Exp CI (Collett et al., 2016). Noticeably, the first two dimensions generate three clusters of environmental, musical and vocal sounds, in contrast to the New CI group who put vocal and musical sounds together. Dimension 1 in Int CI differentiates vocal sounds from non-voice stimuli, although the sound Cough appears further separated from vocal sounds than in Exp CI, being for some reason more grouped with musical sounds than with voice sounds in Int CI patients. In Exp CI patients, Cough approaches the group of voice sounds indicative of a difference in perception of the cough sound between the Int and Exp CI participants.

Similar to the Exp CI results, Dimension 2 separates the musical, vocal and environmental sounds. However, they are grouped together differently along this dimension. Int CI patients group

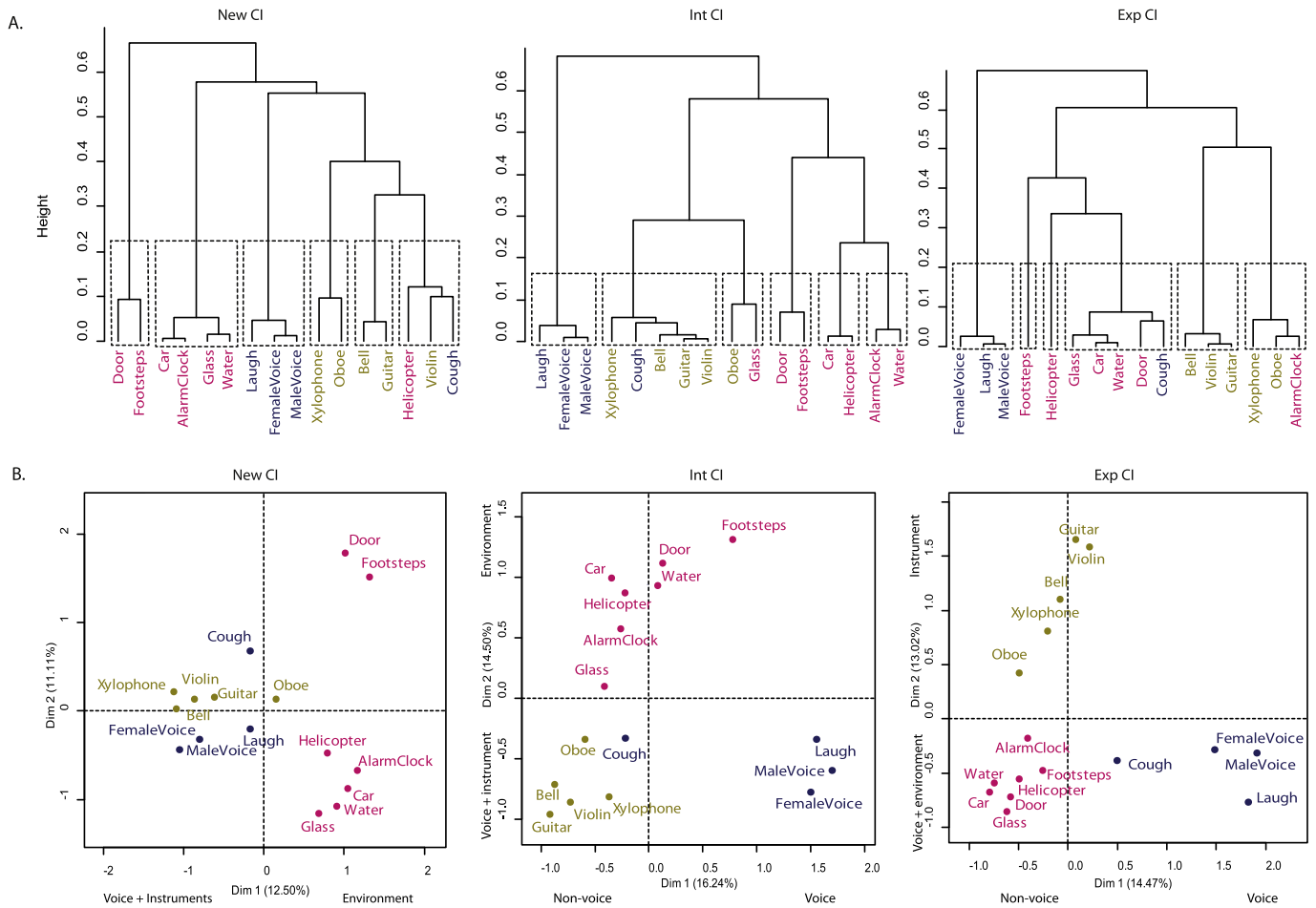


Fig. 2. Categorisation in the groups of CI patients.

A. Dendrograms of the HCPC analysis.

New CI patients failed to distinguish the pre-defined categories, they created four categories where only the environmental sound category could be identified. Int CI patients and Exp CI data (Collett et al., 2016) are close to the pre-defined categories of musical, vocal and environmental sounds. The colour code reflects the pre-defined categories of stimuli.

B. MCA maps.

Dimension 1 in New CI patients corresponds to the separation between environmental sounds and the rest of the sounds. Dimension 1 in Int CI patients reflects the difference between vocal sounds and non-vocal sounds. Dimension 2 in Int patients reflects the separation between environmental sounds and the rest of the sounds. In Exp CI users, Dimension 1 separates voice and non-voice sounds, and Dimension 2 separates musical instruments from the rest of the sounds (Collett et al., 2016). The colour code reflects the pre-defined categories of stimuli.

Table 2

Percent of variance explained by the dimensions of MCA and percent of subjects using each dimension.

	dim 1	dim 2	dim 3	dim 4	dim 5
Exp CI					
% variance	14,5	13,0	9,7	9,4	7,7
% subjects	62,5	50,0	37,5	12,5	12,5
Int CI					
% variance	16,2	14,5	11,6	8,8	8,1
% subjects	53,3	40,0	13,3	6,7	6,7
New CI					
% variance	12,5	11,1	10,5	8,3	7,9
% subjects	70,6	52,9	47,1	17,7	5,9

We consider the direction as used if the subject's coordinates are >0.8 on subject maps (see Fig. 3 as example).

together voices and instruments differentiating them from environmental sounds. Exp CI patients put together voice and environmental sounds differentiating them from musical instruments.

The subject maps illustrated in Fig. 3 indicate the usage of the first two dimensions by each patient and can attest to the

homogeneity of categorisation across patients. It is apparent in Fig. 3 that most of the patients in each group used a similar strategy, as in every CI group the upper right corners of the graphs are the most densely populated in every group. In the New CI and Int CI groups only two patients were below the 0.5 level of usage of both conditions. In the New CI group only one patient was below the 0.5 level of usage for the two conditions. The pattern in NHS was earlier reported to be similar to Exp CI (Collett et al., 2016): NHS had distinct categories corresponding to human-voice, musical sounds and environmental sounds, however, the dispersion of sounds in MCA maps was smaller than in Exp CI. Besides, in comparison to CI patients (see Collett et al., 2016), NHS presented a much less dispersed subject map: over 80% of subjects had a value greater than 0.8 for Dim1 that separates musical sounds from voice sounds.

3.5. Distances between categorised stimuli

The previous analysis of MCA maps demonstrated that the sounds of the different categories tended to be progressively more clustered together with increasing CI experience, a result that is supported by a better categorisation accuracy with CI duration. We

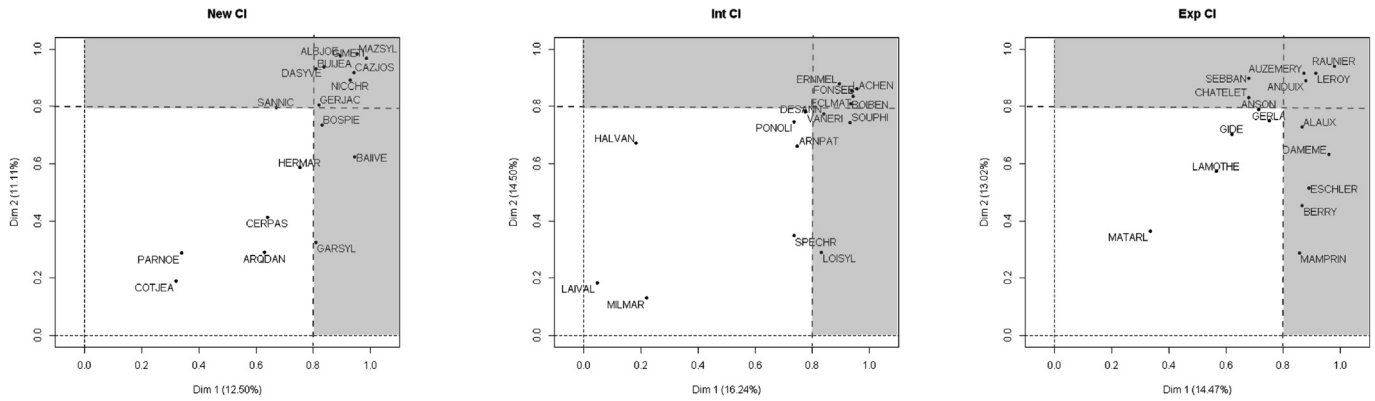


Fig. 3. Usage of the dimensions by the patients in each group. Subject maps indicate the usage of the first two dimensions in MCA maps (see Fig. 2B) by each patient and can attest to the homogeneity of categorisation across patients. Subjects located above 0.8 represent the highest usage of the given dimension.

attempted to quantify this clustering by computing the distances between each stimulus and all other stimuli in the MCA maps. Considering Euclidian distances between stimuli (Fig. 4), one can remark that in the New CI group the distances are short between voice and musical stimuli. However, in Exp CI distances become smaller between voice stimuli (left upper corner) and between environmental stimuli (right lower corner). Fig. 4B demonstrates that comparing Exp CI and New CI there is an increase from New to Exp CI of the Euclidian distance between the voice stimuli (male and female voices, laugh and cough) and most of the other stimuli, though this increase is less pronounced for the cough stimulus.

Furthermore, to quantify the separation of voice stimuli following the free sorting, we compared the Euclidian distances between voice stimuli and the rest of the stimuli (Fig. 5 A). This was done for the New, Int and Exp CI and in NHS. The analysis indicated that the distances between voice and non-voice stimuli were the lowest in the New CI group, revealing a deficit in the discrimination of these categories. They became significantly higher in the Int CI group. The difference between Int CI and Exp CI was not significant, though there is a tendency of improvement with CI experience. The highest distance between voice and non-voice stimuli is observed in controls, and it is significantly higher than in Exp CI. This means

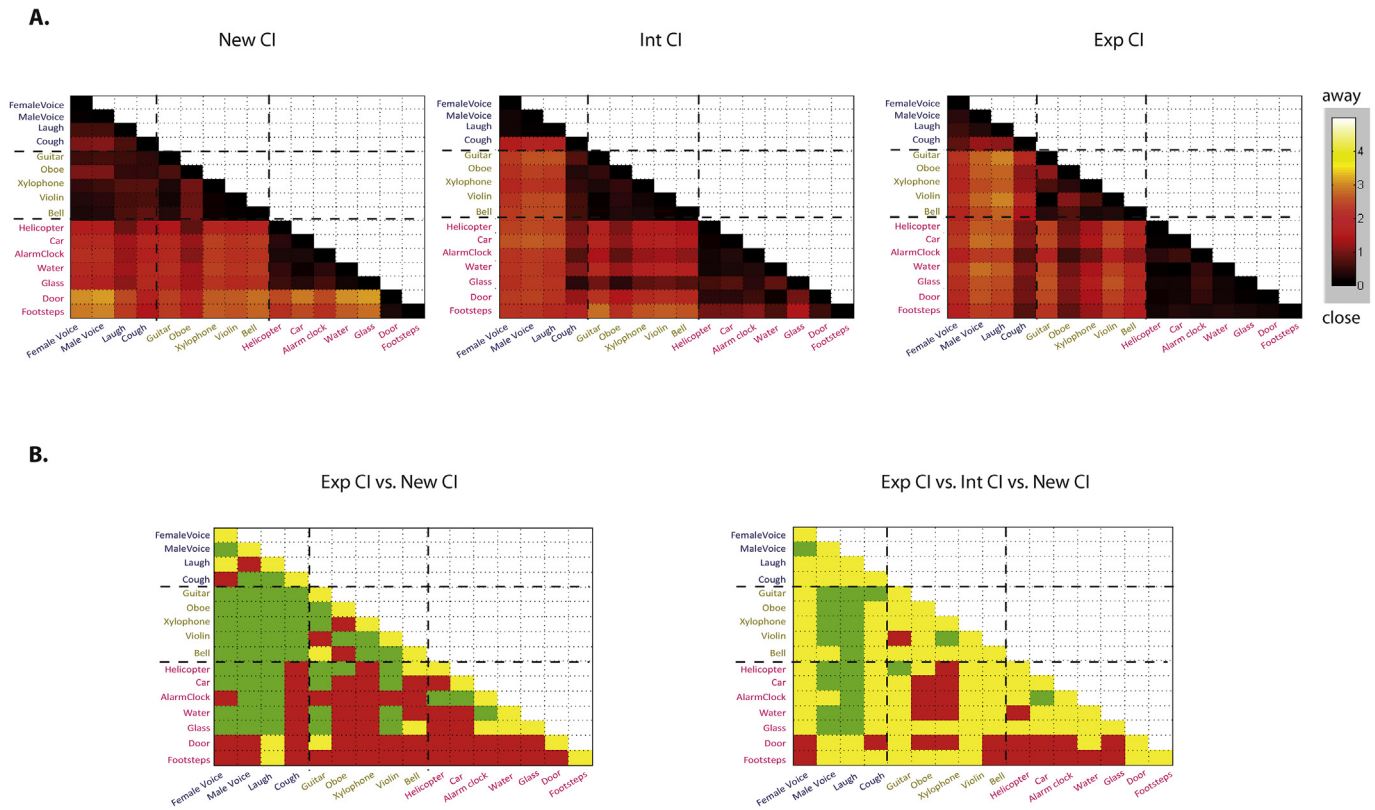


Fig. 4. Euclidian distances between stimuli in MCA maps. **A.** Euclidian distances between stimuli in MCA maps per group. The scale indicates the ranges of the distances. **B.** Changes in Euclidian distances between experienced and newly implanted CI patients per stimulus. Green means a distance increase, red means a distance decrease. Yellow means no detectable change. The criterion for the difference is beyond the standard error in the Exp CI vs. New CI comparison.

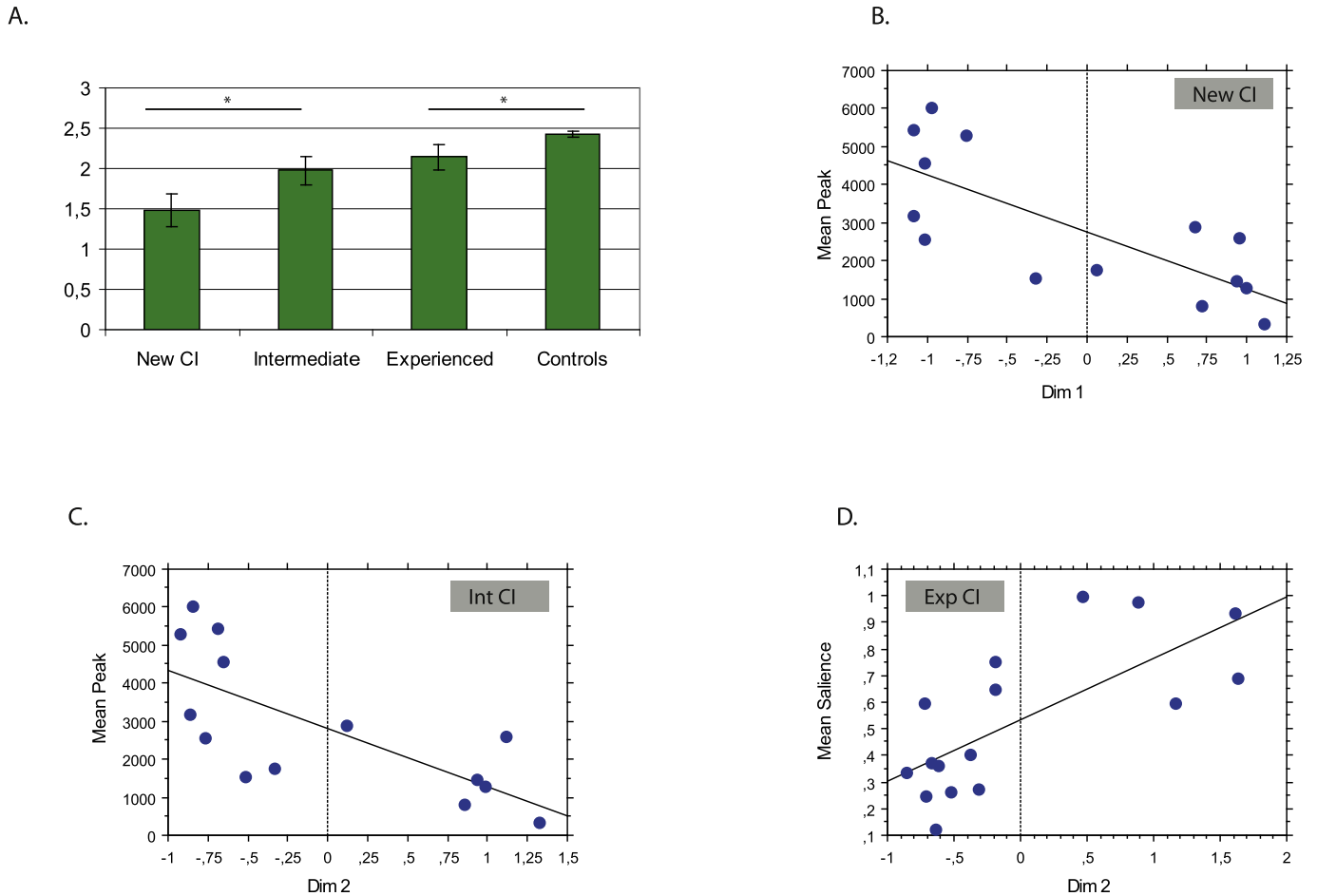


Fig. 5. Voice separation and acoustic property usage.

A. Comparison of the distances between voices and other stimuli between the patient groups. Error bars represent bootstrap confidence intervals.

B. Relation between Dimension 1 in New CI and the Mean Peak measurement of the sound stimuli.

C. Relation between Dimension 2 in Int CI and the Mean Peak measurement of the sound stimuli.

D. Relation between Dimension 2 in Exp CI and the Mean Saliency measurement of the sound stimuli.

that there is an increasing distinction between voice and non-voice sounds with experience in CI patients and the most dramatic increase happens in the period after six months post-implantation. However, even Exp CI patients encompassed poorer distinction abilities than normal-hearing control subjects (tested in Collett et al., 2016).

3.6. Correlations with acoustic properties and clinical parameters

A final analysis was conducted to determine the acoustical cues used by the patients to build up their sound categorisation. Multiple correlation analyses were performed between the results of the MCA and acoustical measurements, and Table 3 presents correlations for each group which passed the level of $p < 0.05$ (uncorrected for multiple comparisons). The sign of the correlation coefficient is meaningless because of the arbitrary direction of the axes in the MCA maps. In New CI, we obtained ten significant correlations with spectral and temporal sound properties for the MCA dimension 1 (Table 3). This dimension corresponds to the distinction of environmental sounds from voice and music sounds. Thus, the correlations reflect a complex combination of spectral and temporal properties, on which CI patients distinguish environmental sounds from music and voice. For example, Max frequency values are correlated with dimension 1 because there are higher

frequency values in musical and vocal sounds compared with mostly low-frequency sounds of the environment. This can indicate that in New CI patients the discrimination of high and low frequencies is a strategy used to differentiate (or combine) sounds. In Int CI, there were eight significant correlations but only for dimension 2 (Table 3). These correlations were also both for spectral and temporal acoustic properties. Dimension 2 in Int CI reflects the same opposition of environmental sounds to voice and music as dimension 1 in New CI. Thus, one can observe that New and Int CI use similar acoustic properties to distinguish between these categories. Int CI use Med frequency, which is a more difficult acoustic feature to extract and analyse; this reflects the increase of the complexity of frequential analysis in Int CI patients. As for Exp CI, it was reported in (Collett et al., 2016) that they had six significant correlations distributed between dimensions 1 and 2. Interestingly, one can also observe an increase of the complexity of the used acoustic features. While Max frequency is used in New CI and Med Frequency in Int CI, these simply frequential types of analysis are not significantly used by Exp CI but are replaced by the much more complex analysis of pitch (mean pitch saliency). Remarkably, mean pitch saliency demonstrates the highest correlation in controls as well (Collett et al., 2016) meaning that Exp CI become closer to normal-hearing subjects in their strategies of acoustic analysis.

There is also a similarity in the usage of acoustic keys between CI

Table 3
Statistical analysis of correlations between MCA dimensions and acoustic properties.

	New CI, <i>Dim1 (env.sounds)</i>		Int CI, <i>Dim2 (env.sounds)</i>		Exp CI, <i>Dim1 (vocal)</i>		Exp CI, <i>Dim2 (music vs. env)</i>		Controls, <i>Dim1 (vocal)</i>		Controls, <i>Dim2 (music vs. env)</i>	
	r	p	r	p	r	p	r	p	r	p	r	p
<i>Spectrally related</i>												
Max Frq	-0,53	0,044										
Med Frq			-0,57	0,026								
Mean Centroid	0,58	0,017										
Pitch salient, mean							0,72	0,002			0,84	<0,001
Spectral STD	0,51	0,043										
Mean bpf	0,67	0,004	0,59	0,016	0,63	0,01						
RMS 150 Hz	-0,6	0,01	0,51	0,04	0,53	0,04			0,61	0,012		
RMS 250 Hz							0,66	0,005			0,64	0,008
RMS 350 Hz												
<i>Temporally related</i>												
Mean Peak	-0,75	0,0012	-0,68	0,005			0,64	0,008			0,72	0,02
Max Peak	-0,65	0,008	-0,61	0,016								
Wav Range	0,55	0,026	0,62	0,011								
Wav No Peaks												
Wav Mean Peak	-0,53	0,036	-0,54	0,031								
Wav std Peak	-0,52	0,040	-0,55	0,028			0,64	0,008			0,71	0,02

groups: all of them used the mean BPF (band path frequency measured with 6 Eric Healy's FIR filters (Healy et al., 2013)), which reflects the mean correlation between octave frequency bands of the sound envelope.

The highest correlations in New and Int CI were for the Mean Peak measurement (Fig. 5B and C), which is a temporal acoustic property. It is noteworthy that this and other correlations with temporal acoustic properties (Table 2) exist for dimension 1 in New CI and for dimension 2 in Int CI. This confirms that Dim1 and Dim 2 correspond to a similar sorting strategy in New CI and Int CI. However, little similarity with other CI groups for temporal correlations can be observed for the Exp CI where the highest correlation is for the spectral property - mean pitch salience (Fig. 5D) with dimension 2, in which musical instruments were categorised. Thus, Exp CI patients used a different strategy to categorise sounds; the strategy is closer to the controls, as shown in our previous study (Collett et al., 2016).

Finally, we checked for the correlations of the obtained MCA dimensions with such clinical characteristics of CI patients as speech perception scores (disyllabic words and sentence in noise), deafness duration, and post-implantation duration. The only significant correlation was in New CI for post-implantation duration and MCA dimension 1 (Spearman $\rho = 0.6$, $p < 0.05$). As mentioned earlier, this dimension reflects the discrimination of environmental sounds from voice and music. This correlation, therefore, indicates a progressive improvement of the distinction of environmental sounds during the first months post-implantation.

There is numerous evidence that aging, independently of perceptual deficit, impact on auditory processing, especially on speech comprehension (Füllgrabe et al., 2014). To test if aging impact on the sound categorisation in CI patients, we conducted Spearman correlations of MCA dimensions with age of the patients within each group (the MCA analysis provides the value of the usage of each dimension per subject). This analysis did not reveal any correlations ($p > 0.13$) suggesting that aging might not be the primary parameter that explains sound categorisation strategies.

4. Discussion

4.1. General

In this study, using the Free Sorting Task we showed that

recently implanted patients can distinguish environmental sounds from the musical and vocal sounds but make no categories for the latter. In both the Int CI and Exp CI participant groups, a difference between vocal and other stimuli was observed. The distances between voice and non-voice stimuli were the lowest in the New CI group. They became significantly higher in the Int CI group. The difference between Int CI and Exp CI was not significant though there was a tendency of increase in these distances. Nevertheless, the greatest distance between voice and non-voice stimuli was observed in controls being significantly higher than in Exp CI. Our study being cross-sectional, the effect of the categorisation strategy memorisation is excluded.

Concerning the usage of acoustic features for categorisation, while Max frequency was used in New CI and Med Frequency in Int CI, these simply frequential types of analysis were not significantly used by Exp CI but were replaced by the much more complex analysis of pitch (mean pitch salience). The observed correlations indicated that there is a gradual increase in the capacity for the distinction of sounds during the first months post-implantation in parallel with a change in cognitive sound processing strategies used by the patients.

The Free Sorting Task (FST) is an inductive experimental procedure that has commonly been used in cognitive and social sciences since the 1950s but only recently developed in sensory analysis for consumer product evaluations (Abdi et al., 2007). Advantages of FSTs include the possibility of testing a large set of stimuli and not structuring the set of stimuli into *a priori* dimensions; therefore, allowing subjects to decide on their own criteria/principles for categorisation. In this way, we have shown that an FST is well-suited to evaluate auditory perception in patients and also developing children (Berland et al., 2015; Collett et al., 2016). However, specific Hierarchical Clustering based on Principal Components (HCPC) and Multiple Correspondence Analysis (MCA) needs to be performed on a large set of assessors, a condition that also requires the subject group to be homogeneous in terms of expertise level (Chollet et al., 2011). Such a constraint justifies the arbitrary decision of splitting subjects into three groups according to the level of CI experience. A global analysis (not presented) that includes all the participants without any distinction of CI duration tends to bias the final results toward the result observed in the most experienced subgroup. Furthermore, our criteria to define CI experience groups are supported by a large body of

studies that point to six months being the critical time before which the auditory skills show the largest improvement and after which the auditory efficiency either reaches a plateau or improves much more slowly (Massida et al., 2011; Rouger et al., 2007).

We pre-defined three broad classes of sounds which have high ecological values. A limit of 16 stimuli was chosen to ensure that the test would not be too fatiguing for the CI users. This is different from a similar study performed on NHS (Gygi et al. (2007) using 50 sounds and may be a certain limitation of our study. Evidently, vocal sounds are the most prominent for social interactions. The category of musical sounds refers to quality of life but also helps with the distinction of different signals (like mobile phones) and was chosen as a second auditory category. A third class of sounds was pre-defined as general environmental sounds, such as running water, door opening, footsteps etc.

4.2. Hierarchical clustering and MCA maps of categories

When we compare the Exp CI data (Collett et al., 2016) with Int CI data (Fig. 2A), the results of sound categorisation show a tendency to correspond to the three pre-defined categories of musical, vocal and environmental sounds, though two different stimuli in each group fail to show this correspondence. This tendency differs from the dendrogram of New CI where no visible match to the pre-defined categories of sounds can be determined, 42.6% of the match in the raw data analysis is not enough for the evident separation in the dendrogram. The New CI created four rather heterogeneous categories. Curiously, the New CI put the sounds of door and footsteps into a separate category apart from other environmental sounds. The same category is present in Int CI. However, it is unlikely that this category is formed on the basis of the similarity of sound meaning; in this case it would have been present in Exp CI and in NHS. Compared, for example, with Gygi et al. (2007) who tested 50 sounds, our sample is rather small, which is a certain limitation for broad conclusions. Though dendrograms provide a good visualisation of categorisation strategies, a more precise and objective way to analyse categorisation, in particular for the between-group comparison, is obtained using the maps of the MCA analysis (Fig. 2B), which we will further discuss in regard to each dimension of the MCA map.

The explained by the first two dimensions variance of 23–30% corresponds to the cases of correlation with the coefficients of 0.48–0.55. Though this is not a strong relationship, such moderate effects are often reported in biomedical studies. Another interesting observation with this respect is that the percentage of explained variance is the lowest in New CI and the highest in controls. This may suggest that the percent of explained variance may be a property reflecting the quality of hearing.

Dimension 1 in New CI as well as Dimension 2 in Int CI (Fig. 2B) correspond to the separation between environmental sounds from the combined musical and vocal sounds. However, the distinction is not always well pronounced in New CI: Oboe, Cough and Laugh are positioned rather closely to environmental sounds in New CI. Footsteps and Door sounds in New CI form a separate subgroup among the environmental sounds in the upper right quadrant corresponding to the phenomenon already discussed for the dendrograms.

Thus, our results indicate the importance of environmental sounds for New and Int CI patients. Gygi et al. (2004) found that information within a frequency band of between 1200 and 2400 Hz was linked to the recognition of environmental sounds. Shafiro 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

(Shafiro, 2008).

It was also shown that the spectral and temporal dynamics are important for environmental sound recognition so that more complicated sounds (i.e. with faster changing temporal and spectral information) are more difficult for CI to identify while those with slower temporal changes are more easily identified (Reddy et al., 2009; Reed and Delhorne, 2005). Our results suggest that the distinction between voices and instruments is less subjectively important at the early stages post-implantation given the technical limitations of the implant.

Dimension 1 in Int CI presents another approach to categorisation compared to New CI. It reflects the difference between vocal sounds and non-voice sounds with an exception for the sound Cough, which is grouped with musical sounds but not with voice sounds in Int CI patients. One can hypothesise that Cough, which is a rather abrupt sound, resembles some percussion instruments for these patients.

Similarly to Int CI, in Exp CI users voice and non-voice sounds are separated along the 1st Dimension. However, cough remains somewhere in between, reflecting an uncertainty of the patients concerning this sound. This uncertainty could reflect the property of the sound Cough or the auditory deficit. It is more likely that it is related to the auditory deficit because in normal-hearing controls the same sound was put closely with other vocal sounds into the voice category ([Collett et al., 2016], Fig. 3).

A peculiar feature of Exp CI users is that they put together voice and environmental sounds differentiating them from musical instruments (Dimension 2). It follows that contrary to Int CI users, Dimension 2 reflects mostly the importance of musical instruments for these patients. Thus, one can observe the emergence of the category of musical instruments in Exp CI patients, which is not present in patients at earlier stages post-implantation. This may be the consequence of the distinction between voice and musical instruments, which was absent at the stage of Int CI.

It is evident that spectral degradation by the implant dramatically affects the perception of any melodic sound by CI users such as telephone tunes or any instrumental passages, which rely on fine spectral details that cannot be efficiently transmitted through the implant (Cousineau et al., 2010; Gfeller et al., 2002; Looi et al., 2008). Though cochlear implants can improve the awareness of environmental sounds (Looi and Arnephy, 2010), the identification of such sounds in CI varies from about 45% (Shafiro et al., 2011) to about 80% in experienced CI users (Reed and Delhorne, 2005). However, the performance can be increased by computer-based training (Shafiro et al., 2015), which is reinforced by multimodal integration (Altieri, 2017). It is possible to simulate the perception of spectrally degraded sounds through cochlear implants using a vocoding technique in normal-hearing listeners. Such studies have shown that both normal-hearing listeners perceiving vocoded sounds and CI users have impaired ability to discriminate vocal and environmental sounds (Leech et al., 2009; Loebach and Pisoni, 2008; Massida et al., 2011; Shafiro, 2008). However, given a high diversity of environmental sounds, it remains a problem to determine general acoustic features for their discrimination.

4.3. Distances between voice and non-voice stimuli

To test the importance of voice sound distinction with respect to other sounds, we used a quantitative analysis of MCA maps, which was achieved by calculating the Euclidian distances between voice and non-voice sounds. This information is presented in Fig. 4 with the colour-coded distances between each stimulus and the rest of the stimuli. Comparing different stages post-implantation in Fig. 4A, it is possible to see that from New to Exp CI distances become smaller between voice stimuli (left upper corner) and

between environmental stimuli (right lower corner). This means that vocal and environmental stimuli become grouped closer together. Further details can be viewed in Fig. 4B, which indicates between which sounds there is an increase (green), decrease (red) or no change (yellow) of the distance at different stages post-implantation. Comparing Exp CI vs. New CI users, it is evident that there is an increase of the distance between voice sounds and most of the other sounds depicted by the green columns on the left of the matrix. Interestingly, within the group of vocal sounds there is also an increase of the distance between male and female voices.

This increase of the distance between male and female voices remains even if we include Int CI users in a stricter analysis requiring that distances between New CI, Int CI and Exp CI change in the same direction – either increase or decrease (the right panel in Fig. 4B). Although in this more restrictive analysis there is more yellow code (no distances survived the criterion of the unidirectional change), we observe an increase of the distance between Male voice, Laugh and most of the other sounds depicted by the green columns on the left of the matrix.

The statistical comparison of distances between vocal and non-vocal sounds at each stage post-implantation reveals a gradual increase of these distances from New CI to Exp CI (Figure 5). The increase was significant between New CI and Int CI patients but insignificant between Int CI and Exp CI patients. Thus, the strongest improvement is seen during the first months post-implantation. This conclusion is also supported by the fact that the only significant correlation with post-implantation duration was found in New CI with MCA dimension 1 (Spearman $\rho = 0.6$, $p < 0.05$). As earlier discussed, this dimension reflects the opposition of environmental sounds to voice and music and the found correlation indicates that there is a gradual increase in the capacity of the distinction of environmental sounds during the first six months post-implantation.

Including controls from our previous study (Collett et al., 2016) into this analysis, we can also see that the distance between vocal and non-vocal sounds is significantly higher even in comparison with Exp CI users (Fig. 5). Thus, the distances between vocal and non-vocal sounds reflect the general auditory capacity and are still at the sub-optimal level in Exp CI patients.

4.4. Correlations of MCA dimensions with acoustic properties

Correlations of categorisation measures with acoustic properties are useful to investigate whether patients use the same cues but with better efficiency or whether they switch to different acoustic cues in the course of rehabilitation. In New CI, there were ten significant correlations only for Dimension 1, which reflects the categorisation of environmental sounds. These correlations reflect a certain complex of spectral and temporal properties, which differentiates environmental sounds from music and voice (Gygi et al., 2004). We suggest that New CI patients are most efficient in the separation of the environmental sounds on the basis of acoustic frequency and temporal properties of the sounds. It is likely that they use acoustic hearing more than everyday listening strategies.

In Int CI, the pattern of correlations for Dimension 2 is in most cases similar to the correlations in New CI, Dimension 1. This is easily explained because Dimension 2 in Int CI reflects the same opposition of environmental sounds to voice and music as Dimension 1 in New CI patients. Despite the resemblance of most correlations, the Int CI patients use Med Frequency instead of Max frequency in New CI. Med frequency is a more difficult acoustic feature to extract and analyse; this reflects the increase in the complexity of frequential analysis in Int CI patients.

While Max frequency is used in New CI and Med Frequency in Int CI, these simply frequential types of analysis are not

significantly used by Exp CI but are replaced by the much more complex analysis of pitch (mean pitch salience) as indicated in (Collett et al., 2016). Mean pitch salience demonstrates the highest correlation in controls as well (Collett et al., 2016) meaning that Exp CI patients become closer to normal-hearing subjects in their strategies of sound categorisation. In general, we observe an increase of the complexity of the used acoustic features in Exp CI patients as witnessed by the correlations with acoustic features for both Dimension 1 and Dimension 2. This can also be related to the improvement in speech perception, which is correlated with the perception of environmental sounds in CI patients (Shafiro et al., 2011).

The highest correlations in New and Int CI were for the Mean Peak measurement (Fig. 5B and C), which is a temporal acoustic property. Compared with other groups, Exp CI patients used a different strategy to categorise sounds, which is closer to controls: the highest correlation was for the spectral property - mean pitch salience (Fig. 4D) with Dimension 2, in which musical instruments were categorised.

It is known that the strategy used by a subject can depend on the experimental task. For example, it was demonstrated that asking subjects to find similarities between sounds favours the usage of the acoustic listening mode (Goldstone, 1994; Gygi et al., 2007). We believe that our task was not subject to this bias because of categorisation being a more general procedure, which implies the analysis of both similarities and dissimilarities between objects.

All of the CI patients used the mean BPF, the frequential measure, which was not significantly used by controls. On the contrary, the temporal 'wav std peak' measure, was significantly used by all groups of patients and by the controls. This means that the temporal analysis strategy of CI patients is close to controls. Lastly, all the correlation analyses showed a lack of relationship between the categorisation performance and all the patients' history including deafness duration, CI duration, speech performance, and age of the patients.

A rather high variability in patients' age and deafness duration within each group is a potentially confounding factor. The absence of correlations can also be related to the fact that cochlear implants are mostly tuned for speech processing, and also that the general adaptation strategy of patients is mostly for speech comprehension. The absence of correlation with the performance in our test and speech perception scores can be explained by the fact that acoustic properties of our stimulation were rather different from speech sound properties. Moreover, there was high variability in patients' scores, even for speech scores. Moreover, identification scores and categorisation strategies may not reflect the same cognitive ability. For non-linguistic sounds in CI users, a sound identification score of $49 \pm 13.5\%$ correct was reported, which was much lower than their ability to identify the correct category: $71 \pm 11.5\%$ correct (Inverso and Limb, 2010). This underlies the high adaptive role of categorisation in CI patients, which can help even when the exact sound is not identified.

The mean values of sound identification in CI patients are rather low in our study compared with Inverso and Limb (2010); these mean values reflect the high variability between patients. Given high variability, no conclusion can be based on these mean values, moreover, though the mean values are different, the variance of sound identification performance (SD) in our study overlaps with the data in Inverso and Limb (2010).

Studies on CI patients have highlighted that the average recognition accuracy is about 70%–80% when the adults are evaluated using a closed-set recognition test (Reed and Delhorne, 2005; Inverso and Limb, 2010), however the performance falls drastically with an open-set or nearly open-set response format (Shafiro et al., 2011; Inverso and Limb, 2010). See Table 1 in Liu et al. (2013) for the

comparison of different studies using the open-set and closed-set tests for the recognition of environmental sounds in CI patients. Usually in CI patients, sound identification is done for speech recognition based on words correct, which is quite different from the broader task of recognition of any environmental sound (speech being part of environmental sounds). This can explain the lower results than for word identification.

It is potentially important that the only correlation we found was in New CI between post-implantation duration and the usage of the MCA dimension 1, which reflects the opposition of environmental sounds to voice and music. This correlation may indicate that the crucial period for the distinction of environmental sounds is during the first months post-implantation.

4.5. Conclusions

We found that there was an early deficit of categorisation, especially for vocal sounds, which improved approximately six months post-implantation with a change of categorisation strategy based on different acoustic cues. It follows from our study that patients during the rehabilitation period tend to rely more on the “acoustic listening” mode when they do not perceive meaningful sounds, and this may explain the variability in the groups of CI patients. With the progress of rehabilitation, this mode is gradually replaced by the “everyday listening” mode, which extracts information about the meaning of objects in the environment and their actions (Gygi et al., 2007; Inverso and Limb, 2010).

Our study permitted us to estimate the individual strategies of sound categorisation and demonstrated that there is an increase in the separation of the category of vocal sounds from other sounds during the experience with cochlear implants. This is the first study to show the increase of such a separation during the rehabilitation period post-implantation, and it emphasises the importance of the period of the first six months post-implantation. This result is in line with studies which indicate 6 months as the critical time before which the auditory skills have the greatest improvement and after which they either reach a plateau or improve much more slowly (Massida et al., 2011; Rouger et al., 2007).

For clinical perspectives categorisation tests, which can be quite short, can complement the usual identification tests especially in the situations when the hearing improvement with CI using sound identification tests is not clear or unstable. FST task can be used for rehabilitation programs to improve sound discrimination. As our present data support the view that auditory categorisation strategies are heavily influenced by the identification of sounds, it is possible that the context or auditory environment will also affect the categorisation strategy of listeners. Therefore, by imposing to the subject a specific set of context (or categories) to perform the FST, we can probably improve the identification scores of sounds. Ongoing data are exploring such issue by creating training protocols based on FST of a high number of natural sounds, which are to be sorted with respect to different a-priori categories; a method that can easily be implemented as a home training program for CI deaf patients.

Author contributions

KS - data analysis and writing the article, EC - stimulus preparation, data collection and analysis, GP - experimental methodology, ET - help with patient selection and clinical supervision, OD, MM - clinical supervision, writing the article, PB - supervision, writing the article.

Acknowledgments

We would like to thank all participants for their time as well as the staff at the ORL Hopital Purpan for helping to accommodate and arrange testing, especially ML Laborde, M Tartarye and G Iversenc. We also thank Pr Qannari, Julien Tardieu as well as Olivier Macherey for their help and comments, and Didier Caron & Julien Labourse (Advanced Bionics) for their help and support, Jean-Pierre Jaffrézou for his reading and comments on the manuscript.

This work was supported by a doctoral subvention to E Collett (Advanced Bionics SARL, France), recurrent funding of the CNRS and “Agir pour l’Audition” (#APA-RD2015-6B).

References

- Abdi, H., Valentin, D., Chrea, C., 2007. Analyzing assessors and products in sorting tasks: DISTATIS, theory and application. *Food Qual. Prefer.* 18, 1–16.
- Altieri, N., 2017. Commentary: environmental sound training in cochlear implant users. *Front. Neurosci.* 11. <https://doi.org/10.3389/fnins.2017.00036>.
- Barone, P., Deguine, O., 2011. Multisensory processing in cochlear implant listeners. In: Zeng, F.-G., Popper, A.N., Fay, R.R. (Eds.), *Auditory Prostheses*. Springer New York, New York, NY, pp. 365–381. https://doi.org/10.1007/978-1-4419-9434-9_15.
- Belin, P., Fecteau, S., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. *Trends Cognit. Sci.* 8, 129–135. <https://doi.org/10.1016/j.tics.2004.01.008>.
- Berland, A., Gaillard, P., Guidetti, M., Barone, P., 2015. Perception of everyday sounds: a developmental study of a free sorting task. *PLoS One* 10, e0115557. <https://doi.org/10.1371/journal.pone.0115557>.
- Bonebright, T.L., 1996. An investigation of data collection methods for auditory stimuli: paired comparisons versus a computer sorting task. *Behav. Res. Meth. Instrum. Comput.* 28, 275–278. <https://doi.org/10.3758/BF03204780>.
- Cabrera, L., Tsao, F.-M., Gnansia, D., Bertoni, J., Lorenzi, C., 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. <https://doi.org/10.1121/1.4887444>.
- Cadoret, M., Lâa, S., Pagès, J., 2009. A factorial approach for sorting task data (FAST). *Food Qual. Prefer.* 20, 410–417.
- Chollet, S., Lelièvre, M., Abdi, H., Valentin, D., 2011. Sort and beer: everything you wanted to know about the sorting task but did not dare to ask. *Food Qual. Prefer.* 22, 507–520. <https://doi.org/10.1016/j.foodqual.2011.02.004>.
- Collett, E., Marx, M., Gaillard, P., Roby, B., Fraysse, B., Deguine, O., Barone, P., 2016. Categorization of common sounds by cochlear implanted and normal hearing adults. *Hear. Res.* 335, 207–219. <https://doi.org/10.1016/j.heares.2016.03.007>.
- Copeland, B.J., Pillsbury, H.C., 2004. Cochlear implantation for the treatment of deafness. *Annu. Rev. Med.* 55, 157–167. <https://doi.org/10.1146/annurev.med.55.091902.105251>.
- Cousineau, M., Demany, L., Meyer, B., Pressnitzer, D., 2010. What breaks a melody: perceiving F0 and intensity sequences with a cochlear implant. *Hear. Res.* 269, 34–41. <https://doi.org/10.1016/j.heares.2010.07.007>.
- De Lucia, M., Tzovara, A., Bernasconi, F., Spierer, L., Murray, M.M., 2012. Auditory perceptual decision-making based on semantic categorization of environmental sounds. *Neuroimage* 60, 1704–1715. <https://doi.org/10.1016/j.neuroimage.2012.01.131>.
- Doucet, M.E., Bergeron, F., Lassonde, M., Ferron, P., Lepore, F., 2006. Cross-modal reorganization and speech perception in cochlear implant users. *Brain* 129, 3376–3383. <https://doi.org/10.1093/brain/awl264>.
- Friesen, L.M., Shannon, R.V., Baskent, D., Wang, X., 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, 1150–1163.
- Füllgrabe, C., Moore, B.C.J., Stone, M.A., 2014. Age-group differences in speech identification despite matched audiometrically normal hearing: contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6, 347. <https://doi.org/10.3389/fnagi.2014.00347>.
- Gaver, W.W., 1993. What in the world do we Hear?: an ecological approach to auditory event perception. *Ecol. Psychol.* 5, 1–29. https://doi.org/10.1207/s15326969eco0501_1.
- Gfeller, K., Witt, S., Woodworth, G., Mehr, M.A., Knutson, J., 2002. Effects of frequency, instrumental family, and cochlear implant type on timbre recognition and appraisal. *Ann. Otol. Rhinol. Laryngol.* 111, 349–356. <https://doi.org/10.1177/000348940211100412>.
- Giordano, B.L., McDonnell, J., McAdams, S., 2010. Hearing living symbols and nonliving icons: category specificities in the cognitive processing of environmental sounds. *Brain Cognit.* 73, 7–19. <https://doi.org/10.1016/j.bandc.2010.01.005>.
- Goldstone, R.L., 1994. The role of similarity in categorization: providing a ground-work. *Cognition* 52, 125–157.
- 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./Revue Can. Psychol. Exp.* 61, 54–63. <https://doi.org/10.1037/cjep2007006>.
- Gygi, B., Kidd, G.R., Watson, C.S., 2007. Similarity and categorization of environmental sounds. *Percept. Psychophys.* 69, 839–855.
- 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.
- Healy, E.W., Yoho, S.E., Apoux, F., 2013. Band importance for sentences and words reexamined. *J. Acoust. Soc. Am.* 133, 463–473. <https://doi.org/10.1121/1.4770246>.
- Husson, F., Josse, J., Le, S., Mazet, J., 2014. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R.
- Inverso, Y., Limb, C.J., 2010. Cochlear implant-mediated perception of nonlinguistic sounds. *Ear Hear.* 31, 505–514. <https://doi.org/10.1097/AUD.0b013e3181d99a52>.
- Kaiser, A.R., Kirk, K.I., Lachs, L., Pisoni, D.B., 2003. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *J. Speech Lang. Hear. Res.* 46, 390–404.
- Leech, R., Gygi, B., Aydelott, J., Dick, F., 2009. Informational factors in identifying environmental sounds in natural auditory scenes. *J. Acoust. Soc. Am.* 126, 3147–3155. <https://doi.org/10.1121/1.3238160>.
- Lemaitre, G., Heller, L.M., 2012. Auditory perception of material is fragile while action is strikingly robust. *J. Acoust. Soc. Am.* 131, 1337–1348. <https://doi.org/10.1121/1.3675946>.
- Liberman, A.M., Cooper, F.S., Shankweiler, D.P., Studdert-Kennedy, M., 1967. Perception of the speech code. *Psychol. Rev.* 74, 431–461.
- Liu, S.-Y., Liu, T.-C., Teng, Y.-L., Lee, L.-A., Lai, T.-J., Wu, C.-M., 2013. Environmental sounds recognition in children with cochlear implants. *PLoS One* 8. <https://doi.org/10.1371/journal.pone.0066100>.
- Loebach, J.L., Pisoni, D.B., 2008. Perceptual learning of spectrally degraded speech and environmental sounds. *J. Acoust. Soc. Am.* 123, 1126–1139. <https://doi.org/10.1121/1.2823453>.
- Looi, V., Arnephy, J., 2010. Environmental sound perception of cochlear implant users. *Cochlear Implants Int.* 11, 203–227. <https://doi.org/10.1002/cii.428>.
- Looi, V., McDermott, H., McKay, C., Hickson, L., 2008. Music perception of cochlear implant users compared with that of hearing aid users. *Ear Hear.* 29, 421–434. <https://doi.org/10.1097/AUD.0b013e31816a0d0b>.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 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. <https://doi.org/10.1073/pnas.0607364103>.
- Marx, M., James, C., Foxton, J., Capber, A., Fraysse, B., Barone, P., Deguine, O., 2015. Speech prosody perception in cochlear implant users with and without residual hearing. *Ear Hear.* 36, 239–248. <https://doi.org/10.1097/AUD.000000000000105>.
- Massida, Z., Belin, P., James, C., Rouger, J., Fraysse, B., Barone, P., Deguine, O., 2011. Voice discrimination in cochlear-implanted deaf subjects. *Hear. Res.* 275, 120–129. <https://doi.org/10.1016/j.heares.2010.12.010>.
- Massida, Z., Marx, M., Belin, P., James, C., Fraysse, B., Barone, P., Deguine, O., 2013. Gender categorization in cochlear implant users. *J. Speech Lang. Hear. Res.* 56, 1389–1401. [https://doi.org/10.1044/1092-4388\(2013\)12-0132](https://doi.org/10.1044/1092-4388(2013)12-0132).
- Møller, A.R., 2006. History of cochlear implants and auditory brainstem implants. *Adv. Oto-Rhino-Laryngol.* 64, 1–10. <https://doi.org/10.1159/000094455>.
- Reddy, R.K., Ramachandra, V., Kumar, N., Singh, N.C., 2009. Categorization of environmental sounds. *Biol. Cybern.* 100, 299–306. <https://doi.org/10.1007/s00422-009-0299-4>.
- Reed, C.M., Delhorne, L.A., 2005. Reception of environmental sounds through cochlear implants. *Ear Hear.* 26, 48–61.
- Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Barone, P., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7295–7300. <https://doi.org/10.1073/pnas.0609419104>.
- Shafiro, V., 2008. Identification of environmental sounds with varying spectral resolution. *Ear Hear.* 29, 401–420. <https://doi.org/10.1097/AUD.0b013e31816a0cf1>.
- Shafiro, V., Gygi, B., Cheng, M.-Y., Vachhani, J., Mulvey, M., 2011. Perception of environmental sounds by experienced cochlear implant patients. *Ear Hear.* 32, 511–523. <https://doi.org/10.1097/AUD.0b013e3182064a87>.
- Shafiro, V., Sheft, S., Kuvadia, S., Gygi, B., 2015. Environmental sound training in cochlear implant users. *J. Speech Lang. Hear. Res.* 58, 509. https://doi.org/10.1044/2015_JSLHR-H-14-0312.
- Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Summerfield, Q., 1992. Lipreading and audio-visual speech perception. *Phil. Trans. Biol. Sci.* 335, 71–78. <https://doi.org/10.1098/rstb.1992.0009>.
- Tyler, R.S., Parkinson, A.J., Woodworth, G.G., Lowder, M.W., Gantz, B.J., 1997. Performance over time of adult patients using the Ineraid or nucleus cochlear implant. *J. Acoust. Soc. Am.* 102, 508–522.
- Zeng, F.-G., Nie, K., Stickney, G.S., Kong, Y.-Y., Vongphoe, M., Bhargava, A., Wei, C., Cao, K., 2005. Speech recognition with amplitude and frequency modulations. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2293–2298. <https://doi.org/10.1073/pnas.0406460102>.