Research Article

Robust Neuronal Discrimination in Primary Auditory Cortex Despite Degradations of Spectro-temporal Acoustic Details: Comparison Between Guinea Pigs with Normal Hearing and Mild Age-Related Hearing Loss

YONANE AUSHANA,^{1,2,3} SAMIRA SOUFFI,^{1,2,3} JEAN-MARC EDELINE,^{1,2,3} CHRISTIAN LORENZI,⁴ AND CHLOÉ HUETZ^{1,2,3}

¹Paris-Saclay Institute of Neurosciences (Neuro-PSI), CNRS UMR 9197, Orsay, France

²Université Paris-Sud, 91405, Orsay cedex, France

³Université Paris-Saclay, 91405, Orsay cedex, France

⁴Laboratoire des Systèmes Perceptifs, UMR CNRS 8248, Département d'Etudes Cognitives, Ecole Normale Supérieure (ENS), Paris Sciences & Lettres Research University, 75005, Paris, France

Received: 14 March 2017; Accepted: 11 December 2017; Online publication: 4 January 2018

ABSTRACT

This study investigated to which extent the primary auditory cortex of young normal-hearing and mild hearing-impaired aged animals is able to maintain invariant representation of critical temporal-modulation features when sounds are submitted to degradations of fine spectro-temporal acoustic details. This was achieved by recording ensemble of cortical responses to conspecific vocalizations in guinea pigs with either normal hearing or mild age-related sensorineural hearing loss. The vocalizations were degraded using a tone vocoder. The neuronal responses and their discrimination capacities (estimated by mutual information) were analyzed at single recording and population levels. For normal-hearing animals, the neuronal responses decreased as a function of the number of the vocoder frequency bands, so did their discriminative capacities at the single recording level. However, small neuronal populations were found to be robust to the degradations induced by the vocoder. Similar robustness was obtained when broadband noise was added to exacerbate further the spectro-temporal distortions pro-

duced by the vocoder. A comparable pattern of robustness to degradations in fine spectro-temporal details was found for hearing-impaired animals. However, the latter showed an overall decrease in neuronal discrimination capacities between vocalizations in noisy conditions. Consistent with previous studies, these results demonstrate that the primary auditory cortex maintains robust neural representation of temporal envelope features for communication sounds under a large range of spectro-temporal degradations.

Keywords: vocoder, envelope, fine structure, electrophysiology, auditory cortex, neural discrimination performance, spike timing

INTRODUCTION

Over the last few decades, a wealth of psychophysical studies conducted with young and older normal-hearing and hearing-impaired human subjects have explored the mechanisms of *robust* speech perception, that is the ability to maintain high levels of speech recognition despite sometimes dramatic degradations of the speech signal caused by interfering sound

Correspondence to: Jean-Marc Edeline · Paris-Saclay Institute of Neurosciences (Neuro-PSI), CNRS UMR 9197 · Orsay, France. Telephone: (+33) 1 69 15 49 72; email: jean-marc.edeline@u-psud.fr

sources such as noise, reverberation, or distortions caused by the transmission channel (e.g., the speech processor of a cochlear implant). Research efforts have mostly concentrated on the search for the spectro-temporal acoustic cues used by human subjects, and the neural codes responsible for maintaining invariant neural representations of the speech signal. A large part of this investigation was achieved using vocoders (Dudley 1939), that is signal-processing devices designed to degrade selectively specific acoustic features of the speech signal (e.g., Drullman 1995; Shannon et al. 1995; Friesen et al. 2001; Smith et al. 2002; Nelson et al. 2003; Baskent 2006; Zeng et al. 2005; Lorenzi et al. 2006; Sheft et al. 2008; for a recent review, see Shamma and Lorenzi 2013).

Vocoders decompose incoming sounds into frequency bands mimicking the spectral decomposition performed by the cochlea. The resulting bandpass filtered signals are—in turn—decomposed into a temporal-envelope component (i.e., an amplitude modulator, AM) and a temporal-fine-structure component (i.e., a frequency-modulated (FM) sinusoid). For each band, the temporal-fine-structure component is degraded (by replacing it by a pure tone or a band of noise) and then amplitude modulated by the corresponding (intact or modified) temporal-envelope component. The resulting AM carriers are finally added up and presented for discrimination or identification to human subjects.

Overall, these behavioral studies showed repeatedly that relatively slow (< 16 Hz) AM cues within a limited number of frequency bands (4 to 8) yield nearly perfect speech identification performance in quiet (e.g., Drullman 1995; Shannon et al. 1995; Smith et al. 2002). However, faster AM cues, FM cues, and fine spectral cues were shown to be required for accurate speech identification in steady or fluctuating noise, when speech is presented against a competing speech signal or when speech is temporally interrupted (e.g., Friesen et al. 2001; Gnansia et al. 2009, 2010; Zeng et al. 2005). These findings have been interpreted as indicating that AM cues (the temporal envelope) convey important phonetic information (that is, timbre information) whereas FM and fine spectral cues (the acoustic fine structure) convey “grouping” or “segregation” cues, that is cues used to separate speech from interfering sound sources and integrate partial speech information (e.g., Gnansia et al. 2010; Apoux and Healy 2011; Apoux et al. 2013). One possibility to provide evidence for the relative roles of AM cues, FM cues, and fine spectral cues in speech intelligibility can be to obtain neural data in conditions where some of these cues are present (AM cues) while others are absent (FM cues). As a first attempt, the primary auditory cortex, where neurons are more

sensitive to slow than to fast AM modulations (Liégeois-Chauvel et al. 2004), appears to be a structure which deserves in-depth investigation.

Very few studies have evaluated the neuronal responses to vocoded stimuli either in human or in animal. In human, magnetoencephalographic data indicate that the cortical entrainment is robust to natural and vocoded stimuli (up to four spectral channels) but it is largely reduced when vocoded stimuli are presented in background noise (Ding et al. 2014). In animals, only three electrophysiological studies have assessed the neural responses to vocoded speech or animal vocalizations (Nagarajan et al. 2002; Ranasinghe et al. 2012; Ter-Mikaelian et al. 2013). Altogether, these three studies found that temporal patterns of auditory cortical responses are extremely robust to reduction in the number of frequency bands, suggesting that the temporal code plays a critical role in the robust representation of communication signals cues. For instance, Nagarajan et al. (2002) found that patterns of neural discharges did not differ considerably as a function of number of frequency bands when the number of bands was varied between 2 and 16.

Elderly persons often experience important difficulties understanding speech in adverse listening situations (e.g., Duquesnoy and Plomp 1980; Takahashi and Bacon 1992), sometimes in the absence of elevated audiometric thresholds (for recent reviews, see Füllgrabe and Moore 2014; Schoof and Rosen 2014). One possible explanation for this is that aging per se alters supra-threshold auditory processing (for a review, see Fitzgibbons and Gordon-Salant 2010). Consistent with this idea, older listeners showed deficits in using complex AM patterns to identify speech accurately (Souza and Kitch 2001; Souza and Boike 2006; Schwartz et al. 2008; Sheldon et al. 2008). However, in contrast with these studies, recent work by Schoof and Rosen (2014) and by Paraouty and Lorenzi (2017) did not find any effect of aging on AM sensitivity. Moreover, additional studies have found that sensitivity to FM cues (He et al. 2007; Grose and Mamo 2012; Wallaert et al. 2016) and binaural temporal fine structure cues decline with increasing age (Ross et al. 2007; Grose and Mamo 2010; Hopkins and Moore 2011; Moore et al. 2012a, b; Füllgrabe 2013; King et al. 2014; for a review, see Grose et al. 2015). On the other hand, several studies indicate that frequency selectivity does not change with age provided that audiometric thresholds remain normal (Lutman et al. 1991; Peters and Moore 1992; Hopkins and Moore 2011). These findings are consistent with neurophysiological data suggesting that aging impairs the processing of neural envelope (Frisina and Frisina 1997; Walton et al. 1998) and temporal fine structure information (Ross et al. 2007;

Clinard et al. 2010) in the auditory midbrain and cortex. However, this is at odds with the higher and more precise MEG responses obtained in mild age-related hearing loss human subjects (Presacco et al. 2016; Alain et al. 2014), and also with the relatively subtle effects of aging on the spectral and temporal tuning characteristics of the cortical neurons of macaques (Engle and Recanzone 2013; Overton and Recanzone 2016).

Here, we investigated the cortical representation of unprocessed and vocoded conspecific vocalizations in young guinea pigs with normal hearing (NH) and older guinea pigs exhibiting mild age-related hearing loss (mARHL). Neuronal responses were recorded in the primary auditory cortex (ACx) of these guinea pigs during presentation of whistles (the alarm call for this species) left as such (unprocessed) or processed via a tone-excited vocoder using 10, 20, or 38 frequency bands. We opted for a tone vocoder rather than a noise vocoder because the latter strongly distort the signal's temporal envelope characteristics (Whitmal III et al. 2007) and noise carriers introduce (non-informative) random envelope fluctuations interfering with signal envelopes (Stone et al. 2011).

The data showed that despite a large decrease in response strength, both the temporal reliability and the discrimination performance based on temporal spike patterns were relatively robust to the vocoding process. Importantly, this held true irrespective of age and presence of mild hearing loss. Still, discrimination performance was more affected for mARHL than for NH animals when background noise (added after the spectral degradation) was added to the vocoded stimuli. These data may provide neurophysiological bases for the difficulties in understanding speech in adverse listening situations typically reported by elderly people.

MATERIALS AND METHODS

Subjects

These experiments were performed under the national license A-91-557 (project 2014-25, authorization 05202.02) and using the procedures nos. 32-2011 and 34-2012 validated by the Ethic committee no. 59 (CEEA Paris Centre et Sud). All surgical procedures were performed in accordance with the guidelines established by the European Communities Council Directive (2010/63/EU Council Directive Decree).

Recordings were obtained in the primary ACx of 16 adult pigmented guinea pigs aged 6 months to 2 years old. Animals weighting from 450 to 1230 g came from our own colony housed in a humidity (50–55 %) and temperature (22–24 °C)-controlled facility on a 12-h/12-h light/dark cycle (light on at 7:30 A.M.) with free

access to food and water. None of the animals used here have been exposed to noise, except for the noise inherent to the colony room (air conditioning and the animals' own vocalizations; level estimated to 60 dB SPL).

Two to 3 days before each experiment, the animal's audiogram was determined by testing auditory brainstem responses (ABR) under isoflurane anesthesia (2.5 %) as described in Gourévitch et al. (2009). The ABR was obtained by differential recordings between two subdermal electrodes (SC25-NeuroService) placed at the vertex and behind the mastoid bone. A software (RTLab, Echodia, Clermont-Ferrand, France) allowed averaging 500 responses during the presentation of nine pure-tone frequencies (between 0.5 and 32 kHz) delivered by a speaker (Knowles Electronics) placed in the animal right ear. The auditory threshold of each ABR is the lowest intensity where a small ABR wave can still be detected (usually wave III). For each frequency, the threshold was determined by gradually decreasing the sound intensity (from 80 dB down to –10 dB SPL).

Based upon ABR and the age of animals, two groups of animals were separated: half of the animals (8/16), aged 6 to 9 months were considered to have a normal audiogram which did not differ from the one published in previous studies using young guinea pigs (Gourévitch et al. 2009; Gourévitch and Edeline 2011). The other half (aged of at least 2 years old) displayed mild, but significant ($p < 0.01$), age-related hearing loss which was present across the entire audible bandwidth (the “hearing spectrum”) of the guinea pig (see the supplementary figure for more details).

Surgical Procedures

The animal was anesthetized by an initial injection of urethane (1.2 g/kg, i.p.) supplemented by additional doses of urethane (0.5 g/kg, i.p.) when reflex movements were observed after pinching the hind paw (usually two to four times during the recording session). A single dose of atropine sulfate (0.06 mg/kg, s.c.) was given to reduce bronchial secretions and a small dose of buprenorphine was administrated (0.05 mg/kg, s.c.) as urethane has no analgic properties. After placing the animal in a stereotaxic frame, a craniotomy was performed above the left temporal cortex. A local anesthetic (Xylocain 2 %) was liberally injected in the wound. The opening was 8 mm wide starting at the intersection point between parietal and temporal bones and 8–10 mm height. The dura above the auditory cortex was removed under binocular control and the cerebrospinal fluid was drained through the cisterna to prevent the occurrence of edema. After the surgery, a pedestal

in dental acrylic cement was built to allow an atraumatic fixation of the animal's head during the recording session. The stereotaxic frame supporting the animal was placed in a sound-attenuating chamber (IAC, model AC1). At the end of the recording session, a lethal dose of pentobarbital (>200 mg/kg, i.p.) was administered to the animal.

Recording Procedures

Data were from multiunits recordings collected in the primary auditory cortex (area AI). Extracellular recordings were obtained from arrays of 16 tungsten electrodes (ϕ : 33 μ m, <1 M Ω) composed of two rows of 8 electrodes separated by 1000 μ m (350 μ m between electrodes of the same row). A silver wire, used as ground, was inserted between the temporal bone and the dura matter on the contralateral side. The location of the primary auditory cortex was estimated based on the pattern of vasculature observed in previous studies (Edeline and Weinberger 1993; Manunta and Edeline 1999; Edeline et al. 2001; Wallace et al. 2000). The raw signal was amplified 10,000 times (TDT Medusa). It was then processed by an RX5 multichannel data acquisition system (TDT). The signal collected from each electrode was filtered (610–10,000 Hz) to extract multi-unit activity (MUA). The trigger level was set for each electrode to select the largest action potentials from the signal. On-line and off-line examination of the waveforms suggests that the MUA collected here was made of action potentials generated by two to six neurons at the vicinity of the electrode. For each experiment, we set the position of the electrode array in such a way that the two rows of eight electrodes sample neurons responding from low to high frequency when progressing in the rostral-caudal direction (see examples of tonotopic gradients recorded with such arrays in Fig. 1 of Gaucher et al. 2012 and in Fig. 6a of Occelli et al. 2016).

Acoustic Stimuli

Acoustic stimuli were generated in Matlab, transferred to a RP2.1-based sound delivery system (TDT) and sent to a Fostex speaker (FE87E). The speaker was placed at 2 cm from the guinea pig's right ear, a distance at which the speaker produced a flat spectrum (± 3 dB) between 140 Hz and 36 kHz. Calibration of the speaker was made using noise and pure tones recorded by a Bruel & Kjaer microphone 4133 coupled to a preamplifier B&K 2169 and a digital recorder Marantz PMD671.

We first determined spectro-temporal receptive fields (STRFs) using 97 or 129 pure-tones frequencies scaled with a gamma function, covering six (0.14–9 or

0.28–18 or 0.56–36 kHz) or eight (0.14–36 kHz) octaves, respectively, and presented at three levels: 85 or 75, 65 or 55, 45 or 35 dB SPL. At a given level, each frequency was repeated eight times at a rate of 2.35 Hz in pseudorandom order. The duration of these tones over half-peak amplitude was 15 ms and the total duration of the tone was 50 ms, so there was no overlap between tones. A set of four conspecific vocalizations was used to assess the neuronal response to communication sounds. These vocalizations were recorded from animals of our colony. Pairs of animals were placed in the acoustic chamber and their vocalizations were recorded by a Bruel & Kjaer microphone 4133 coupled to a preamplifier B&K 2169 and a digital recorder Marantz PMD671. A large set of whistle calls was loaded in the Audition software (Adobe Audition 3) and four representative examples of whistle were selected. As shown in Figure 1a (bottom), despite the fact the maximal energy of the four selected whistle was in the same frequency range (typically between 4 and 26 kHz), they displayed marked differences in their spectrograms. In addition, their temporal envelopes clearly differed as shown on their oscillograms (Fig. 1a top). The four selected whistles were processed by three tone vocoders as previously described (Gnansia et al. 2009, 2010). The three vocoders only differed in terms of the number of bands used to decompose (i.e., filter) the whistles (38, 20, or 10 bands). In the following, the 38 bands are briefly presented, but the same applies for the 20-bands or the 10-bands vocoders. Each digitized signal was passed through a bank of 38 *fourth*-order Gammatones (Patterson 1987) filters with center frequencies uniformly spaced along a Guinea Pig-adapted ERB (Equivalent Rectangular Bandwidth) scale (Sayles and Winter 2010) ranging from 50 to 35,505 Hz. In each frequency band, the temporal envelope was extracted using full-wave rectification and lowpass filtering at 64 Hz with a zero-phase, sixth-order Butterworth filter. The resulting envelopes were used to amplitude modulate sine-wave carriers with frequencies at the center frequency of the Gammatone filters, and with random starting phase. Impulse responses were peak-aligned for the envelope (using a group delay of 16 ms) and the temporal fine structure across frequency channels (Hohmann 2002). The modulated signals were finally weighted and summed over the 38 frequency bands. The weighting compensated for imperfect superposition of the bands' impulse responses at the desired group delay. The weights were optimized numerically to achieve a flat frequency response (Hohmann 2002). Figure 1 displays the waveforms (first row) and the spectrograms (second row) of the four whistles together with the spectrograms of the 38-band vocoded (third row),

the 20-band vocoded (fourth row), and the 10-band vocoded (fifth row) whistles. In the following figures, the unprocessed whistles will be referred as to the original, and the vocoded versions as to Voc38 (Voc20, Voc10 respectively) for the 38-band (20-band, 10-band, respectively) vocoded whistles. It is important to note that the broadband noise added to the stimuli was not processed through the vocoder. Broadband noise was added to the vocoded vocalizations, that is *after* vocoding. Our aim was to introduce additional spectro-temporal distortions and degrade further the acoustic differences between vocalizations. Last, in contrast to previous cortical studies that all have used noise vocoders (Nagarajan et al. 2002; Ranasinghe et al. 2012; Ter-Mikaelian et al. 2013), a tone vocoder was used here, because noise vocoders were found to introduce random (non-informative) intrinsic temporal-envelope fluctuations distorting the signal's AM features (Stone et al. 2011; Shamma and Lorenzi 2013).

Experimental Protocol

Inserting an array of 16 electrodes in the cortical tissue almost systematically induces a deformation of the cortex. A 30-min recovering time lap was allowed for the cortex to return to its initial shape, then the array was slowly lowered. Tests based on measures of STRFs were used to assess the quality of our recordings and to adjust electrodes' depth. The recording depth was 500–1000 μm , which corresponds to layer III and the upper part of layer IV according to Wallace and Palmer (2008). However, as a result of the convex shape of the guinea pig ACx, it is likely that all the electrodes were not exactly at the same depth.

When a clear tuning was obtained for at least 8 of the 16 electrodes, the stability of the tuning was assessed: we required that the recorded neurons displayed at least three successive similar STRFs (each lasting 6 min) before starting the protocol. When the stability was satisfying, the protocol started by presenting the acoustic stimuli in the following order: We first presented the four whistles at 75 dB SPL in their natural versions, followed by the vocoded versions with 38, 20, and 10 bands. The same set of whistle in their natural and vocoded versions were then presented in the presence of continuous white noise background set at 65 dB SPL (this white noise was added after the spectral degradation produced by the vocoding, and refreshed from trial to trial). In all cases, each vocalization was repeated 20 times. Presentation of this entire stimulus set lasted 30 min.

Data Analysis

Quantification of Responses to Pure Tones The STRFs derived from MUA were obtained by constructing post-stimulus time histograms for each frequency with 1 ms time bins. The firing rate evoked by each frequency was quantified by summing all the action potentials from the tone onset up to 100 ms after this onset. Thus, STRFs are matrices of 100 bins in abscissa (time) multiplied by 97 or 129 bins in ordinate (frequency). All STRFs were smoothed with a uniform 5×5 bin window.

For each STRF, the Best Frequency (BF) was defined as the frequency where the highest firing rate was recorded. Peaks of significant response were automatically identified using the following procedure: A positive peak in the MU-based STRF was defined as a contour of firing rate above the average level of the baseline activity (estimated from the ten first milliseconds of STRFs at all intensity levels) plus six times the standard deviation of the baseline activity. For each electrode, to determine the $Q_{20\text{dB}}$ and $Q_{40\text{dB}}$, we estimated the lowest intensity for which significant STRFs were observed. At this intensity, the BF was considered to be the Characteristic Frequency (CF). At 20 and 40 dB above this intensity, the low and high frequency limits of the STRFs were determined to compute the $Q_{20\text{dB}}$ and $Q_{40\text{dB}}$.

The upper and lower frequency bounds of the STRF were also used to count the number of frequency bands of the vocoding stimuli that were falling within the STRF. For this analysis, only the 10-band vocoder was considered. The possible values ranged from zero to four bands falling into the neurons' STRF.

Quantification of Responses Evoked by Vocalizations The responses to vocalizations were quantified using two parameters: (i) the firing rate of the evoked response, which corresponds to the total number of action potentials occurring during the presentation of the stimulus, divided by its duration; (ii) the spike timing reliability coefficient (CorrCoef) which quantifies the trial-to-trial reliability of the response. This index was computed for each vocalization: it corresponds to the normalized covariance between each pair of spike trains recorded at presentation of this vocalization and was calculated as follows:

$$\text{CorrCoef} = \frac{1}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\sigma x_i x_j}{\sigma x_i \sigma x_j}$$

where N is the number of trials and $\sigma x_i x_j$ is the normalized covariance at zero lag between spike trains x_i and x_j where i and j are the trial numbers.

FIG. 1. Original and vocoded whistles. **a** First row shows the time-waveform of the four whistle stimuli. Their spectrogram is plotted on the second row. **b** The three rows represent the spectrogram of the vocoded version of the four whistle stimuli for each vocoder processing. The first row uses 38 bands, the second 20 bands, the third 10 bands

Spike trains x_i and x_j were previously convolved with a 10-ms width Gaussian window. Based upon computer simulations, we have previously shown that this CorrCoef index is not a function of the neurons' firing rate (see Gaucher et al. 2013a).

As in previous studies (e.g., Gaucher and Edeline 2015), we set up a criterion of an evoked firing rate of at least 1.5 spike s^{-1} and a CorrCoef of at least 0.15 to discard recordings that did not respond to any vocalization. A systematic and meticulous examination of our database has shown that recordings exhibiting only a phasic onset response generated a CorrCoef of at least 0.15, and that cells responding by tonic responses lacking of temporal organization generated at least a firing rate of $1.5 \text{ spikes s}^{-1}$. Only the recordings that matched this criterion for all four whistles in their original and vocoded versions were classified as "responsive" and were included in the present study.

Quantification of Mutual Information from the Responses to Vocalizations The method developed by Schnupp et al. (2006) was used to quantify the amount of information (Shannon 1948) contained in the responses to vocalizations obtained with natural and vocoded stimuli. This method allows quantifying how well the vocalization's identity can be inferred from neuronal responses. Here, "neuronal responses" refers either

to (i) the spike trains obtained from a small group of neurons below one electrode (for the computation of the individual Mutual Information, MI), or to (ii) a concatenation of spike trains simultaneously recorded under several electrodes (for the computation of the population MI). In both cases, the following computation steps were the same. Neuronal responses were represented using different time scales ranging from the duration of the whole response (firing rate) to a 1-ms precision (precise temporal patterns), which allows analyzing how much the spike timing contributes to the information. As this method is exhaustively described in Schnupp et al. (2006) and in Gaucher et al. (2013a), we present below the main principles only.

The method relies on a pattern-recognition algorithm that is designed to "guess which stimulus evoked a particular response pattern" by going through the following steps: From all the responses of a cortical site to the different stimuli, a single response (test pattern) is extracted and represented as a PSTH with a given bin size (different sizes were considered as discussed further below). Then, a mean response pattern is computed from the remaining responses (training set) for each stimulus class. The test pattern is then assigned to the stimulus class of the closest mean response pattern. This operation is

repeated for all the responses, generating a confusion matrix where each response is assigned to a given stimulus class. From this confusion matrix, the Mutual Information (MI) is given by Shannon's formula:

$$MI = \sum_{x,y} p(x,y) \times \log_2 \left(\frac{p(x,y)}{p(x) \times p(y)} \right)$$

where x and y are the rows and columns of the confusion matrix, or in other words, the values taken by the random variables "presented stimulus class" and "assigned stimulus class."

In our case, we used responses to the four whistles and selected the first 264 ms of these responses to work on spike trains of exactly the same duration (the shortest whistle being 264-ms long). In a scenario where the responses do not carry information, the assignments of each response to a mean response pattern is equivalent to chance level (here 0.25 because we used four different stimuli and each stimulus was presented the same number of times) and the MI would be close to zero. In the opposite case, when responses are very different between stimulus classes and very similar within a stimulus class, the confusion matrix would be diagonal and the mutual information would tend to $\log_2(4) = 2$ bits.

As already mentioned, this algorithm was applied with different bin sizes ranging from 264 to 1 ms. For each cortical site, two parameters were extracted from this analysis. First, the value of MI for the largest bin size (264 ms) which represents the MI value when only the firing rate is considered ($MI_{\text{Firing Rate}}$). Second, the MI value obtained for a bin of 8 ms, MI_{Patterns} , which is a temporal resolution leading to the highest values of information in previous studies (Schnupp et al. 2006; Huetz et al. 2009).

The MI estimates are subject to non-negligible positive sampling biases. Therefore, as in Schnupp et al. (2006), we estimated the expected size of this bias by calculating MI values for "shuffled" data, in which the response patterns were randomly reassigned to stimulus classes. The shuffling was repeated 100 times, resulting in 100 MI estimates of the bias (MI_{bias}). These MI_{bias} estimates are then used as estimators for the computation of the statistical significance of the MI estimate for the real (unshuffled) datasets: the real estimate is considered as significant if its value is statistically different from the distribution of MI_{bias} shuffled estimates. Significant MI estimates were computed for MI calculated from neuronal responses under one electrode.

The information carried out by a group of recording was estimated by the population MI ($MI_{\text{Population}}$), using the same method as for the MI_{Patterns} : responses

of several simultaneous recordings were grouped together and considered as a single pattern. To assess the influence of the group size of simultaneous recordings on the information carried by that group ($MI_{\text{Population}}$), the number of sites used for computing $MI_{\text{Population}}$ varied from 2 to the maximal possible size (which is equal to 16 minus the non-responsive sites). As the number of possible combinations could be extremely large (C_n^k , where k is the group size and n the number of responsive sites in a recording session), a threshold was fixed to spare computation time: when the number of possible combinations exceeded 100, 100 combinations were randomly chosen, and the mean of all combinations was taken as the $MI_{\text{Population}}$ for this group size.

Since our electrode arrays can only achieve up to 16 simultaneous recordings (but often less than 12 recordings responding to vocalizations), pseudo-populations (population of neurons which were not simultaneously recorded) were generated as described in the method developed by Ince et al. (2013). Briefly, we sampled responsive recordings from our database to generate pseudo-populations of 2–25 recordings and compute the $MI_{\text{population}}$ for 500 random samplings. These values of $MI_{\text{pseudo-population}}$ were computed for the original and vocoded versions of the vocalizations (38, 20, and 10).

Statistics To assess the significance of the multiple comparisons (vocoding process: four levels; the groups of animals: NH vs. mARHL; two conditions: quiet vs. background noise), we used an analysis of variance (ANOVA) for multiple factors to analyze the whole data set. Follow-up tests were corrected for multiple comparisons using Bonferroni corrections and were considered as significant if their p value was below 0.05. All data are presented as mean values $\pm$ standard error (s.e.m.).

RESULTS

Initially, 880 stable multiunit recordings were obtained from 16 guinea pigs tested with the original and vocoded whistles. Out of them, 581 recordings were classified as "responsive" (see "Materials and Methods") and used in the subsequent analyses. As explained in the "Materials and Methods," half of the animals (8/16, 6–9 months old) had normal audiograms (Fig. 2a, NH, black line) which did not differ from the one published in previous studies for guinea pigs of the same age (Gourévitch et al. 2009; Gourévitch and Edeline 2011). The other half (about 2 years old) displayed mild, but significant ($F(1,14) = 3.12$; $p < 0.01$), age-related hearing loss (Fig. 2a,

mARHL, red line), which was present for across the entire hearing spectrum. Analyzing the cortical STRF obtained from normal-hearing (NH) and mild age-related hearing loss (mARHL) animals revealed that the $Q_{20\text{dB}}$ and $Q_{40\text{dB}}$ values were lowered in the mARHL animals compared to the NH animals (Fig. 2b, respectively $F(1,157)=1.76$ and $F(1,105)=1.37$ $p < 0.05$ in both cases), indicating that hearing loss impacted on the width of frequency tuning in auditory cortex. In the following, we will first focus on the effect of vocoding on the cortical responses in NH animals ($n = 285$ recordings). Then, the influence of background noise on the processing of original and vocoded vocalizations will be presented for both the 285 recordings obtained on NH and also for a set of 296 recordings obtained in mARHL animals.

FIG. 2. Brainstem and cortical thresholds for normal and mild age-related hearing loss animals. **a** Thresholds (in dB SPL) for the Auditory Brainstem Responses (ABRs) of the Normal Hearing (NH, in dark) and mild Age-Related Hearing Loss (mARHL, in red) animals. For each group, the group data are represented by the thick line with the SEM and the data from each animal are represented by the thin lighter lines. **b** Mean values ($\pm$ sem) of the $Q_{20\text{dB}}$ and $Q_{40\text{dB}}$ of cortical neurons estimated in NH and mARHL animals. The number of recordings in each group is indicated on the bars

Effect of Vocoding on the Responses in the ACx of Normal-Hearing Animals

Figure 3 shows the responses of six simultaneous recordings at presentation of the original and vocoded whistles. Compared with the responses to the original whistles (top row), a strong decrease in response strength (for the four whistles) was observed as the number of frequency bands in the vocoder was reduced. For this set of recordings, the temporal reproducibility (as measured by the CorrCoef) also decreased (from 0.45 for the original vocalization to 0.32 for the 10-band vocalizations).

Over the set of 285 recordings obtained on these animals, vocoding reduced the evoked firing rate and the temporal reliability (Fig. 4a, b, one-way ANOVA, $F(3,1136)=3.36$ for firing rate and $F(3,1136)=2.91$ for CorrCoef, $p < 0.05$ for both). Interestingly, post-hoc pairwise comparisons showed that only the 10-band vocoded whistles elicited significantly lower responses than the original whistles (Fig. 4a, Bonferroni corrected pairwise t test, $p = 0.02$). To assess how the reduction in the number of frequency bands altered the neural discrimination, MI was computed either from the firing rate or from the temporal patterns (see “Materials and Methods”). As shown in previous studies (Schnupp et al. 2006; Gaucher et al. 2013a; Liu and Schreiner 2007; Engineer et al. 2008), temporal spike patterns allowed a much better discrimination between stimuli than the firing rate did: the MI_{Patterns} values (Fig. 4d) are about twice the one of $MI_{\text{Firing Rate}}$ (Fig. 4c). These two measures were impacted by the vocoding (ANOVA, $F(3,515)=3.11$, $p = 0.03$ for $MI_{\text{Firing Rate}}$, $F(3,865)=5.24$, $p = 0.001$ for MI_{Patterns}). However, the vocoding process only induced a small impairment of the neurons coding abilities: the MI computed on temporal patterns was only reduced by ~ 0.2 bits for the vocoded 10-band stimuli compared to the original stimuli (Fig. 4d; note that a perfect discrimination between the four whistles would result in a MI equals to 2 bits).

Vocoding Affected the Cortical Responses Independently of the Neurons’ Receptive Field

To determine if the vocoding process impacts similarly cortical neurons independently of their BF, the recordings were split in four groups according to their BF: BF below 5 kHz ($n = 100$), between 5 and 10 kHz ($n = 61$), between 10 and 15 kHz ($n = 26$), and above 15 kHz ($n = 52$; note that 46 recordings had no clear BF but responded to vocalizations). A two-way ANOVA showed that there was no interaction between the BF factor and the vocoding factor for the three parameters quantifying the neuronal responses (the firing rate, the temporal reliability, and the mutual infor-

FIG. 3. Rasterplots for six-neuron recordings in response to reference and vocoded versions of each whistle. First row shows the six-neuron recordings in response to the original whistles (different whistles on each column). For each subplot, individual neuron's APs are either in red or blue to distinguish between different neurons (1 to 6). Shaded area corresponds to spontaneous activity

before stimulus presentation. Second to fourth rows show the responses to the vocoded whistles (38 to 10 bands). Last row shows the waveforms of the whistles. Mean values of Firing rate (Fr) and CorrCoef (averaged over the six neurons and the four stimuli) are presented at the end of each row

mation; $p > 0.9$ in all cases), indicating that the vocoding did not differentially impact neurons with low, middle, and high frequency BF. This was not totally surprising given that our vocoded stimuli still contained energy in a wide range of frequencies. However, it could be expected that recordings whose frequency receptive field (RF) spanned none of the vocoder's frequency bands would be more impacted by the vocoding process. To investigate this issue, the recordings were classified according to the number of bands of the vocoded stimuli that fell into their RFs at 75 dB. A global ANOVA indicated that all measures of neuronal activity (firing rate, temporal reliability, and mutual information) were impacted by the number of bands falling into the recording's RF (number of bands factor, $p < 0.001$ for Firing Rate, CorrCoef, $MI_{Patterns}$). However, there was no interaction between the "number of bands in RF" factor and the vocoding factor (two-way ANOVA interaction vocoding $\times$ number of bands, $p > 0.7$ for Firing rate, CorrCoef, and $MI_{Patterns}$), indicating that the recordings were similarly impacted by the spectral degradation of the stimulus.

Robust Coding of Vocalizations by Small Neuronal Populations in ACx for Normal-Hearing Animals

As shown above, the parameters quantifying the responses of individual recordings were impacted by the stimulus spectral degradations generated by the vocoder. The next objective was to determine if this effect also impacted the coding abilities at the population level. To explore this issue, mutual information was computed based on the temporal patterns obtained from 2 to 16 simultaneously recordings. Computed on a population, the MI quantifies how well the four whistles can be discriminated based on temporal patterns expressed by this population. Figure 5a presents the value of the $MI_{Population}$ as a function of the number of recordings. Each line represents a given set of simultaneous recordings, from which N recordings (N varying from 2 to 16) were selected using a bootstrap procedure to compute a mean MI value for N recordings (see "Materials and Methods"). During most of the recording sessions, not all 16 electrodes gave significant responses, and, as a

FIG. 4. Group data for normal-hearing animals. Each subplot represents the mean (and standard error of the mean) value of a measure of neuronal activity for original whistles and vocoded versions (Vcoded 38, 20, 10 bands). **a** Firing rate, **b** CorrCoef, **c** $MI_{\text{Firing Rate}}$, **d** MI_{Patterns} . Stars represent a significant difference between original and vocoded (38, 20, or 10), as computed with post-hoc pairwise tests corrected for multiple comparisons with the Bonferroni method

consequence, most of the lines do not reach the right side of Figure 5a (corresponding to 16 simultaneous recordings). Strikingly, for the original and vocoded stimuli, the maximum value of information (2 bits) was reached by many populations of 16 simultaneous recordings. Therefore, when MI values were computed for populations of 12 individual recordings (Fig. 5b, pink line), the mean value was close to maximum and there was no difference between the four groups of stimuli (ANOVA, $F(3,35)=0.9$, $p=0.4$), indicating that even when the stimuli were strongly degraded, populations of 12 simultaneous recordings were, on average, sufficient to discriminate between the four whistles. However, when populations of only eight simultaneous neurons were considered, the vocoding process resulted in a significant decrease of the MI carried by the neuronal responses (Fig. 5b, orange line, ANOVA, $F(3,81)=2.82$, $p<0.05$).

As only a few simultaneous recordings were available for 12 neurons, a pseudo-population analysis was performed (see “Materials and Methods”). Pseudo-populations were constructed from all responsive neurons, regardless of whether or not they were simultaneously recorded (Ince et al. 2013). This analysis allowed creating populations up to 25 neurons. Figure 5c shows the values of $MI_{\text{Pseudo-population}}$ for original (black), 38-band vocoded (blue), 20-band

vocoded (green), and 10-band vocoded (red) responses as a function of the number of neurons included in the pseudo-population. It is obvious from Figure 5c that, for populations made of small numbers of neurons (less than 12), the $MI_{\text{Pseudo-population}}$ decreases as a function of the vocoder’s number of bands. Strikingly, when 25 neurons are included in the population (shaded gray area on Fig. 5c), all curves almost reach 2 bits (the maximum possible MI value, see gray line on Fig. 5d). This suggests that, even with the 10-band vocoded stimuli, 25 neurons are sufficient to perfectly discriminate the four whistles (but see “Discussion” for the limitations).

Impact of Mild Age-Related Hearing Loss on Neural Responses to Vcoded Stimuli in Quiet and Noisy Conditions

The results presented above show that the discrimination between vocalizations performed by ACx neurons is relatively robust to the vocoding process: Despite a significant decrease in evoked activity, temporal reliability, and mutual information (Fig. 4), the decrease in neural performance is modest considering how important stimulus degradations are. Moreover, a perfect discrimination between stimuli can still be achieved at the population level (Fig. 5). The following section now explores to which extent the above conclusions also hold for hearing-impaired animals and in noisy conditions.

As explained at the beginning of the “Results” section, the ABRs for eight animals (2 years old aged) revealed a mild but significant age-related hearing loss (mARHL, Fig. 2). The cortical neurons obtained from these animals (which displayed slightly broader frequency tuning curves) were tested with the same set of stimuli (original and vocoded whistles). The results are displayed in Figure 6a, in red lines for mARHL animals, and in black lines for NH animals. Recordings from mARHL animals showed higher firing rate and temporal reliability compared to the recordings obtained for NH animals (three-way ANOVA, $F(1,4447)=11$ for firing rate, $F(1,4447)=14.6$ for CorrCoef, $p<0.001$). Also, recordings obtained from mARHL animals showed discriminative abilities at the level of individual recordings or neuronal populations comparable to those observed for NH animals ($F(1,3358)=0.41$, $p=0.5$ for MI_{Patterns} , $F(1,161)=1.86$, $p=0.18$ for $MI_{\text{Population}}$).

To further test the robustness of cortical coding in NH and mARHL animals, a broadband background noise was added to the original and vocoded whistles (see “Materials and Methods”). The same parameters quantifying neural activity (firing rate, temporal reliability, MI_{Patterns} , and $MI_{\text{Population}}$) were measured in the noisy condition. Figure 6b shows the difference

FIG. 5. Mutual Information (MI) computed for a population of recordings. **a** MI as a function of the size of the population. X-axis represents the number of recordings in the population. Each curve is a given recording (maximum 16 simultaneous recordings). A MI value of 2 corresponds to perfect discrimination of whistles. Shaded colored bars (orange and pink) refer to populations of 8 or 12 simultaneous recordings (see **b**). **b** Mean values of $MI_{Population}$ computed with 12 (pink line) or 8 (orange line) recordings for each set of stimuli (original and vocoded whistles with 38, 20, and 10

bands). **c** $MI_{Pseudo-population}$ as a function of the size of the population. Same axes as in **a**. Red, green, blue, and black curves represent the different stimuli (Original, Voc38, Voc20, Voc10). $MI_{Pseudo-population}$ values for 8 (orange), 12 (pink), and 25 (gray) recordings were extracted to be plotted in **d**. **d** Mean values of $MI_{Pseudo-population}$ computed with 8 (orange line), 12 (pink line), or 25 (gray line) recordings for each set of stimuli (original whistles, vocoded 38, 20, 10 bands)

between noisy and quiet conditions for NH (gray lines) and mARHL (light red lines) animals. As expected, all points are below zero, showing that the addition of background noise reduced the magnitude of all these parameters: the firing rate and temporal reliability decreased (three-way ANOVA, $F(1,4447) = 14$ for firing rate, $F(1,4447) = 349$ for CorrCoef, $p < 0.001$), as well as the MI computed at the individual and population level (three-way ANOVA, $F(1,3358) = 152$ for $MI_{Patterns}$, $F(1,161) = 37$ for $MI_{Population}$, $p < 0.001$). However, for all the measures derived from the neuronal responses, there was no interaction between noise and vocoding (noise $\times$ vocoding, $p > 0.6$ for all measures) suggesting that for NH subjects, the vocoder and background noise did not alter differentially the cortical responses. Interestingly, the effect of background noise was stronger in the recordings collected from mARHL animals compared to those collected in NH animals (red lines are often below the gray ones) both for the temporal reliability and the $MI_{Patterns}$ (three-way ANOVA, interaction noise $\times$ hearing loss, $F(1,4447) = 5$ for CorrCoef, $F(1,3358) = 16$ for $MI_{Patterns}$, $p < 0.03$). It was not the case for the $MI_{Population}$ when computed with 12-neuron populations, but when computed with 8

neurons, the interaction became significant (three-way ANOVA, interaction noise $\times$ hearing loss, $F(1,161) = 3$, $p = 0.08$ for $MI_{Population}(\#neurons = 12)$; $F(1,327) = 7$, $p < 0.01$ for $MI_{Population}(\#neurons = 8)$). However, as in NH animals, there was no interaction between noise and vocoding for the mARHL animals (noise $\times$ vocoding, $p > 0.6$ for all measures, $R^2 < 1.6e-4$). Together, these last two results suggest that hearing loss amplifies the impact of noise on cortical responses to whistles whereas it has no influence on the impact of vocoding.

DISCUSSION

Brief Overview of the Results

The responses of auditory cortex neurons were found to be altered when tested with spectrally degraded vocalizations (in comparison with responses obtained with the original vocalizations). Nevertheless, these alterations were not drastic and were prominent only when the vocoding process was applied to 10 broad frequency bands. For normal-hearing animals, the firing rate and the spike timing reliability decreased, which in turn reduced the discrimination ability of

a Neuronal activity measures for mARHL vs NH animals

b Effect of background noise on neuronal activity measures

FIG. 6. Effects of mild age-related hearing loss and background noise on vocoding. **a** Average neural activity measure (from left to right: firing rate, CorrCoef, MI_{pattern}, MI_{population}) computed from the responses to original whistles and vocoded versions (Vocoded 38, 20, 10 bands) for normal hearing animals (NH, black lines), and mild age-related hearing loss animals (mARHL, red lines). **b** Mean differences

between responses to noisy and quiet whistles. For each recording, the neural activity measure (from left to right: Δ firing rate, Δ CorrCoef, Δ MI_{pattern}, Δ MI_{population}) evoked in the quiet condition is subtracted from its value evoked by the noisy condition. Means of the differences are plotted in gray for NH animals, and in light red for mARHL animals. Bars represent the standard error of the mean

cortical cells at the single recording level. However, when at least 12 neurons were considered, an almost perfect discrimination between the four vocalizations was reached and stayed unaffected by the vocoding process. Data collected on animals exhibiting mild age-related hearing loss showed similar patterns in quiet conditions. In the presence of background noise, the discriminative abilities of cortical cells were more reduced in mild age-related hearing loss animals than in normal-hearing animals, suggesting that the noise impacted more the cortical processing of vocoded vocalizations for the mild age-related hearing loss than for the normal hearing animals. Before discussing the functional implications of these results, several methodological points should be discussed.

Methodological Considerations

First, the data presented here are based on the responses of small groups of neurons (two to five) located at the vicinity of each recording electrodes and we did not perform spike sorting analyses to separate the different units. We consider that this do not invalidate our findings because our recordings technique and conditions allowed detecting relatively subtle effects in previous studies when manipulating either the acoustic properties of the stimuli (Ocelli et al. 2016) or pharmacological agents (Gaucher et al. 2013a; Gaucher and Edeline 2015).

Second, it could be argued that the use of anesthetics can impact the potential effect induced

by the vocoding process on cortical responses. Some studies have reported important differences between anesthetized and unanesthetized animals: for example, the maximum rate of click trains that primary ACx neuron is able to follow (Rennaker et al. 2007), the percentage of bursts in the evoked responses (Massaux and Edeline 2003; Massaux et al. 2004), and the evoked oscillations triggered by acoustic stimuli (Cotillon-Williams and Edeline 2003). However, other studies have reported comparable responses in awake and anesthetized rats for consonant sounds (Engineer et al. 2008) and conspecific vocalizations in guinea pig (Huetz et al. 2009).

Third, the hearing loss displayed by our mild age-related hearing loss animals (20 dB) is modest and not comparable with what was reported in a previous study (Gourévitch and Edeline 2011). Most likely, this is the direct consequence of the fact that animals used in our previous study were much older (3.5 to 4 years in Gourévitch and Edeline 2011 vs. 2 years here). As a consequence, the cortical responses obtained here in our mild age-related hearing loss animals are similar to the responses obtained in normal-hearing animals, which can explain why the vocoding process impacts similarly the two groups of animals.

MI was used here as a metric to quantify the differences between neuronal responses to the four original, or the four vocoded vocalizations. It provides an objective measure to quantify to what extent neuronal responses differ between each other. The MI value can be interpreted as a way to evaluate the discrimination ability of auditory cortical neurons when tested with the original, then with the vocoded stimuli. Obviously, our use of the term “discrimination” does not refer to perceptual abilities. Only behavioral experiments, in humans or animals, can give access to perceptual abilities. To what extent the discriminative abilities of neurons recorded in a given cortical area have causal relationships with the whole perceptual ability is the subject of a vast literature, and is clearly out of the scope of the present study.

Here, the values of MI obtained with the original vocalizations on single recordings are relatively low (below 1 bit) even when computed in terms of spike patterns (the values of $MI_{\text{firing rate}}$ were more than twice lower than the values of MI_{patterns}). This can be explained by the fact we used four versions of the same type of vocalization which had roughly the same spectrograms (even if they were acoustically different). It is only when computing the MI values on populations of simultaneous recordings that an almost perfect discrimination could be achieved (MI value close to 2 in our case). However, our population analysis was suffering from small data sets, due to the fact that only a few numbers of simultaneous recording sets were available. To increase the number of

neuronal populations, analyses were performed following the same method as in Ince et al. (2013). A large number of pseudo-populations (500) was generated from our data set (population of neurons not simultaneously recorded) and simulations were conducted to investigate the impact of vocoding on the discriminative abilities of these pseudo-populations. These analyses confirmed that when 25 neurons are considered, an almost perfect discrimination is possible by ACx neurons, thus suggesting that the discriminative capabilities of cortical cells are robust to the spectral degradations produced on our stimuli set.

One of the main limitations of our study is the small number of stimuli used to test the discriminative performance. Previous studies using larger sets of stimuli have found lower MI values (Gaucher et al. 2013a; Gaucher and Edeline 2015; Chechik et al. 2006). Thus, one can wonder whether using a larger set of stimuli should allow us to reach similar conclusions. The parametric study performed by Ince et al. (2013) indicates that even with larger set of stimuli (up to 50), it is still possible to reach the maximal value of MI (indexing an almost perfect discrimination) with large neuronal populations. In our experimental conditions, if a larger set of stimuli is used, the effects of the vocoding process on the cortical discrimination remain unknown.

Relationship with Human Studies

Vcoders have long been used to characterize the key acoustic features responsible for robust speech perception (Dudley 1939). This literature has largely documented that slow AM fluctuations (<16 Hz) in a limited number of frequency bands (four to eight) is sufficient to maintain an almost perfect identification of speech in quiet (Shannon et al. 1995; Smith et al. 2002). By showing here that auditory cortex neurons are very robust in discriminating between vocoded communication sounds, our results clearly point out that the primary auditory cortex can send relevant (i.e., invariant) cues to downstream brain areas more specialized in speech identification (Belin et al. 2000; Romanski and Averbach 2009).

Elderly persons often experience important difficulties understanding speech in adverse listening situations sometimes in the absence of elevated audiometric thresholds (for recent reviews, see Füllgrabe and Moore 2014; Schoof and Rosen 2014). Potentially, this can result from an alteration of supra-threshold auditory processing (for a review, see Fitzgibbons and Gordon-Salant 2010), which can be explained if high-threshold auditory nerve fibers are the first to be impacted during aging (Sergeyenko et al. 2013). Initially, it was found that older listeners showed deficits in using complex AM patterns to

identify speech accurately (Souza and Kitch 2001; Souza and Boike 2006; Schwartz et al. 2008; Sheldon et al. 2008). This is possibly due to reduced sensitivity to AM cues or reduced central ability to make optimal use of AM cues (e.g., Takahashi and Bacon 1992; He et al. 2008; Füllgrabe et al. 2015). However, several studies did not find any effect of aging on AM sensitivity (Schoof and Rosen 2014; Paraouty et al. 2016; Paraouty and Lorenzi 2017), suggesting that the processing of AM information is roughly comparable for younger and older listeners exhibiting similar audiometric thresholds. When elderly subjects show sensorineural hearing loss, AM sensitivity is generally improved compared to normal (Wallaert et al. 2017), presumably because of the loss in the fast-acting amplitude compression applied by outer hair cells in the cochlea (Moore et al. 1996). Thus, the potentially detrimental effects of aging on AM processing may be confounded with (and counter-balanced by) the loss of compression associated with mild hearing loss. This may explain why previous MEG studies found enhanced cortical responses to speech for older subjects with mild hearing loss compared to young and older subjects with normal hearing (Sörös et al. 2009; Alain et al. 2014; Presacco et al. 2016). More specifically, the synchronization of MEG responses with the speech envelope was more precise for older subjects both in quiet and in noise (Presacco et al. 2016). Other studies found higher amplitude for the waves P1 (Heschl's gyrus) and N1 (Planum temporale) in hearing-impaired older subjects than in normal hearing subjects (Alain et al. 2014; Sörös et al. 2009). This may also explain the increase in evoked firing rate and temporal reliability observed in our cortical recordings from animals with mild aged-related hearing loss.

Relationship with Previous Electrophysiological Studies at the Cortical Level

As a general rule, previous cortical studies reported that the responses of cortical neurons to communication sounds are particularly robust to spectral alterations produced by vocoders, even when the vocoders only incorporated four or two bands. This was the case in the marmoset (Nagarajan et al. 2002) and gerbil (Ter-Mikaelian et al. 2013) auditory cortex at presentation of conspecific vocalizations; this was also the case in the rat auditory cortex at presentation of vocoded consonants and vowels (Ranasinghe et al. 2012). Consistent with these findings, the present study showed that the degradation produced by the vocoder had little impact on cortical responses. Even when 10-band vocoded stimuli were presented, the decrease in MI at the level of individual recordings was modest (0.2 bit on average, see Fig. 4). Only one

of the previous studies (Nagarajan et al. 2002) has evaluated the addition of broadband noise on the conspecific vocalizations and found little effect of masking noise up to a signal-to-noise ratio of +10 dB (but note that the vocoded vocalizations were not presented in the background noise in their study). This is consistent with other findings showing robust representation of speech in the primary auditory cortex in the presence of white/pink noise and in the presence of convolutive reverberations (Mesgarani et al. 2014).

So far, none of the previous cortical studies have evaluated to what extent age-related hearing loss can potentiate the degradation produced by the vocoding process. In the present study, we found that in the case of mild age-related hearing loss (about 20 dB), this hearing loss did not contribute much to the alterations of responses produced by the vocoder (Fig. 6a). However, it seems that in noisy conditions, the discrimination capabilities of cortex neurons were slightly lower in mild age-related hearing loss animals compared with normal ones (Fig. 6b). Obviously this effect should be confirmed by investigating more in depth the effects of other types of background noise and testing several levels of signal-to-noise ratio.

What Are the Acoustic Features Triggering Cortical Responses to Original and Vcoded Vocalizations?

As shown here, populations of cortical neurons are able to discriminate between vocalizations that have been severely degraded by vocoding. What are the acoustic cues allowing for this robust discrimination? As a matter of fact, it is now well documented that auditory cortex neurons do not respond to vocalizations based on their spectral content. Several studies have clearly showed that the cortical responses to natural stimuli cannot be predicted based upon the spectral content of the vocalizations (e.g., Machens et al. 2004; Laudanski et al. 2012; Bitterman et al. 2008; Chechik and Nelken 2012; review in Huetz et al. 2011; Gaucher et al. 2013b); even responses from inferior colliculus neurons cannot be predicted from spectral content (Woolley et al. 2006). In contrast, auditory cortex neurons are highly sensitive to the slope of changes in sound level (Fishbach et al. 2001; Heil 1997a, b) and as a consequence are sensitive to the temporal envelope of natural stimuli. Thus, one explanation for the present results is that despite the severe degradations in fine spectro-temporal details produced by vocoding, auditory cortex neurons can still discriminate between the four whistles based upon the differences in gross temporal envelope cues in each frequency band. One can envision that even if a much larger set of stimuli was used, a good cortical

discrimination between vocoded stimuli is still possible provided that these stimuli differ in terms of gross temporal-envelope cues.

Conclusions and Perspectives

The present study shows that both for normal-hearing and hearing-impaired animals, auditory cortex neurons can discriminate between spectrally similar communication sounds in the absence of relevant temporal-fine-structure cues (i.e., FM cues and fine spectral details). These results suggest that the critical cue for discriminating such sounds corresponds to the relatively slow temporal-envelope fluctuations (AM) of these sounds within a limited number of frequency bands. More severe conditions of stimuli presentation and degree of hearing loss should be tested in future work before generalizing these results. Also, it will be of great interest to study other cortical areas to determine to which extent the effects reported in AI impact neuronal responses in belt, parabelt, and downstream cortical areas.

ACKNOWLEDGEMENTS

Special thanks to Fabien Lhericel and Céline Dubois for taking care of the guinea pig colony and to Dr. Grégoire Levasseur for helping generating high quality figures.

Funding Information This work was supported by grants from the National Research Agency (ANR Blanc2011, HearFin and ANR Blanc2014, Heart) to CL and JME. This work was also supported by ANR-11-0001-02 PSL* and ANR-10-LABX-0087. SS was supported by the Fondation pour la Recherche Médicale (FRM grant number ECO20160736099 to [SS]).

REFERENCES

- ALAIN C, ROYE A, SALLOUM C (2014) Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Front Syst Neurosci* 8:8
- APOUX F, HEALY EW (2011) Relative contribution of target and masker temporal fine structure to the unmasking of consonants in noise. *J Acoust Soc Am* 130(6):4044–4052. <https://doi.org/10.1121/1.3652888>
- APOUX F, YOHO SE, YOUNGDAHL CL, HEALY EW (2013) Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners. *J Acoust Soc Am* 134(3):2205–2212. <https://doi.org/10.1121/1.4816413>
- BASKENT D (2006) Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *J Acoust Soc Am* 120(5):2908–2925. <https://doi.org/10.1121/1.2354017>
- BELIN P, ZATORRE RJ, LAFAILLE P, AHAD P, PIKE B (2000) Voice-selective areas in human auditory cortex. *Nature* 403(6767):309–312. <https://doi.org/10.1038/35002078>
- BITTERMAN Y, MUKAMEL R, MALACH R, FRIED I, NELKEN I (2008) Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451(7175):197–201. <https://doi.org/10.1038/nature06476>
- CHECHIK G, NELKEN I (2012) Auditory abstraction from spectro-temporal features to coding auditory entities. *Proc Natl Acad Sci U S A* 109(46):18968–18973. <https://doi.org/10.1073/pnas.1111242109>
- CHECHIK G, ANDERSON MJ, BAR-YOSEF O, YOUNG ED, TISHBY N, NELKEN I (2006) Reduction of information redundancy in the ascending auditory pathway. *Neuron* 51(3):359–368. <https://doi.org/10.1016/j.neuron.2006.06.030>
- CLINARD CG, TREMBLAY KL, KRISHNAN AR (2010) Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. *Hear Res* 264(1–2):48–55. <https://doi.org/10.1016/j.heares.2009.11.010>
- COTILLON-WILLIAMS N, EDELINE JM (2003) Evoked oscillations in the thalamo-cortical auditory system are present in anesthetized but not in unanesthetized rats. *J Neurophysiol* 89(4):1968–1984. <https://doi.org/10.1152/jn.00728.2002>
- DING N, CHATTERJEE M, SIMON JZ (2014) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *Neuroimage* 88:41–46. <https://doi.org/10.1016/j.neuroimage.2013.10.054>
- DRULLMAN R (1995) Temporal envelope and fine structure cues for speech intelligibility. *J Acoust Soc Am* 97(1):585–592. <https://doi.org/10.1121/1.413112>
- DUDLEY H (1939) Remaking speech. *J Acoust Soc Am* 11(2):169–177. <https://doi.org/10.1121/1.1916020>
- DUQUESNOY AJ, PLOMP R (1980) Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis. *J Acoust Soc Am* 68(2):537–544. <https://doi.org/10.1121/1.384767>
- EDELINE JM, WEINBERGER NM (1993) Receptive field plasticity in the auditory cortex during frequency discrimination training: selective retuning independent of task difficulty. *Behav Neurosci* 107(1):82–103. <https://doi.org/10.1037/0735-7044.107.1.82>
- EDELINE JM, DUTRIEUX G, MANUNTA Y, HENNEVIN E (2001) Diversity of receptive field changes in auditory cortex during natural sleep. *Eur J Neurosci* 14(11):1865–1880. <https://doi.org/10.1046/j.0953-816x.2001.01821.x>
- ENGINEER CT, PEREZ CA, CHEN YH, CARRAWAY RS, REED AC, SHETAKE JA, JAKKAMSETTI V, CHANG KQ, KILGARD MP (2008) Cortical activity patterns predict speech discrimination ability. *Nat Neurosci* 11(5):603–608. <https://doi.org/10.1038/nn.2109>
- ENGLE JR, RECANZONE GH (2013) Characterizing spatial tuning functions of neurons in the auditory cortex of young and aged monkeys: a new perspective on old data. *Front Aging Neurosci* 4:36. <https://doi.org/10.3389/fnagi.2012.00036>
- FISHBACH A, NELKEN I, YESHURUN Y (2001) Auditory edge detection: a neural model for physiological and psychoacoustical responses to amplitude transients. *J Neurophysiol* 85(6):2303–2323
- FITZGIBBONS PJ, GORDON-SALANT S (2010) Age-related differences in discrimination of temporal intervals in accented tone sequences. *Hear Res* 264(1–2):41–47. <https://doi.org/10.1016/j.heares.2009.11.008>
- FRIESEN LM, SHANNON RV, 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(2):1150–1163. <https://doi.org/10.1121/1.1381538>
- FRISINA DR, FRISINA RD (1997) Speech recognition in noise and presbycusis: relations to possible neural mechanisms. *Hear Res* 106(1–2):95–104. [https://doi.org/10.1016/S0378-5955\(97\)00006-3](https://doi.org/10.1016/S0378-5955(97)00006-3)
- FÜLLGRABE C (2013) Age-dependent changes in temporal-fine-structure processing in the absence of peripheral hearing loss.

- Am J Audiol 22(2):313–315. [https://doi.org/10.1044/1059-0889\(2013/12-0070\)](https://doi.org/10.1044/1059-0889(2013/12-0070)).
- FÜLLGRABE C, MOORE BC (2014) Effects of age and hearing loss on stream segregation based on interaural time differences. *J Acoust Soc Am* 136(2):EL185–EL191. <https://doi.org/10.1121/1.4890201>
- FÜLLGRABE C, MOORE BC, STONE MA (2015) 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>
- GAUCHER Q, EDELINE JM (2015) Stimulus-specific effects of noradrenaline in auditory cortex: implications for the discrimination of communication sounds. *J Physiol Lond* 593(4):1003–1020. <https://doi.org/10.1113/jphysiol.2014.282855>
- GAUCHER Q, EDELINE JM, GOURÉVITCH B (2012) How different are the local field potentials and spiking activities? Insights from multi-electrodes arrays. *J Physiol Paris* 106(3–4):93–103. <https://doi.org/10.1016/j.jphysparis.2011.09.006>
- GAUCHER Q, HUETZ C, GOURÉVITCH B, EDELINE J-M (2013A) Cortical inhibition reduces information redundancy at presentation of communication sounds in the primary auditory cortex. *J Neurosci* 33(26):10713–10728. <https://doi.org/10.1523/JNEUROSCI.0079-13.2013>
- GAUCHER Q, HUETZ C, GOURÉVITCH B, LAUDANSKI J, OCCELLI F, EDELINE JM (2013B) How do auditory cortex neurons represent communication sounds? *Hear Res* 305:102–112. <https://doi.org/10.1016/j.heares.2013.03.011>
- GNANSIA D, PEAN V, MEYER B, LORENZI C (2009) Effects of spectral smearing and temporal fine structure degradation on speech masking release. *J Acoust Soc Am* 125(6):4023–4033. <https://doi.org/10.1121/1.3126344>
- GNANSIA D, PRESSNITZER D, PEAN V, MEYER B, LORENZI C (2010) Intelligibility of interrupted and interleaved speech for normal-hearing listeners and cochlear implantees. *Hear Res* 265(1–2):46–53. <https://doi.org/10.1016/j.heares.2010.02.012>
- GOURÉVITCH B, EDELINE JM (2011) Age-related changes in the guinea pig auditory cortex: relationship with peripheral changes and comparison with tone-induced hearing loss. *Eur J Neurosci* 34(12):1953–1965. <https://doi.org/10.1111/j.1460-9568.2011.07905.x>
- GOURÉVITCH B, DOISY T, AVILLAC M, EDELINE JM (2009) Follow-up of latency and threshold shifts of auditory brainstem responses after single and interrupted acoustic trauma in guinea pig. *Brain Res* 1304:66–79. <https://doi.org/10.1016/j.brainres.2009.09.041>
- GROSE JH, MAMO SK (2010) Processing of temporal fine structure as a function of age. *Ear Hear* 31(6):755–760. <https://doi.org/10.1097/AUD.0b013e3181e627e7>
- GROSE JH, MAMO SK (2012) Frequency modulation detection as a measure of temporal processing: age-related monaural and binaural effects. *Hear Res* 294(1–2):49–54. <https://doi.org/10.1016/j.heares.2012.09.007>
- GROSE JH, MAMO SK, BUSS E, HALL JW (2015) Temporal processing deficits in middle age. *Am J Audiol* 24(2):91–93. https://doi.org/10.1044/2015_AJ-A14-0053
- HE NJ, MILLS JH, DUBNO JR (2007) Frequency modulation detection: effects of age, psychophysical method, and modulation waveform. *J Acoust Soc Am* 122(1):467–477. <https://doi.org/10.1121/1.2741208>
- HE NJ, MILLS JH, AHLSTROM JB, DUBNO JR (2008) Age-related differences in the temporal modulation transfer function with pure-tone carriers. *J Acoust Soc Am* 124(6):3841–3849. <https://doi.org/10.1121/1.2998779>
- HEIL P (1997A) Auditory cortical onset responses revisited. I. First-spike timing. *J Neurophysiol* 77(5):2616–2641
- HEIL P (1997B) Auditory cortical onset responses revisited. II. Response strength. *J Neurophysiol* 77(5):2642–2660
- HOHMANN V (2002) Frequency analysis and synthesis using a Gammatone filterbank. *Acust Acta Acust* 88:433–442
- HOPKINS K, MOORE BC (2011) The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J Acoust Soc Am* 130(1):334–349. <https://doi.org/10.1121/1.3585848>
- HUETZ C, PHILIBERT B, EDELINE JM (2009) A spike-timing code for discriminating conspecific vocalizations in the thalamocortical system of anesthetized and awake guinea pigs. *J Neurosci* 29(2):334–350. <https://doi.org/10.1523/JNEUROSCI.3269-08.2009>
- HUETZ C, GOURÉVITCH B, EDELINE J-M (2011) Neural codes in the thalamocortical auditory system: from artificial stimuli to communication sounds. *Hear Res* 271(1–2):147–158. <https://doi.org/10.1016/j.heares.2010.01.010>
- INCE RAA, PANZERI S, KAYSER C (2013) Neural codes formed by small and temporally precise populations in auditory cortex. *J Neurosci* 33(46):18277–18287. <https://doi.org/10.1523/JNEUROSCI.2631-13.2013>
- KING A, HOPKINS K, PLACK CJ (2014) The effects of age and hearing loss on interaural phase difference discrimination. *J Acoust Soc Am* 135(1):342–351. <https://doi.org/10.1121/1.4838995>
- LAUDANSKI J, EDELINE JM, HUETZ C (2012) Differences between spectro-temporal receptive fields derived from artificial and natural stimuli in the auditory cortex. *PLoS One* 7(11):e50539. <https://doi.org/10.1371/journal.pone.0050539>
- LIÉGEAIS-CHAUVEL C, LORENZI C, TRÉBUCHON A, RÉGIS J, CHAUVEL P (2004) Temporal envelope processing in the human left and right auditory cortices. *Cereb Cortex* 14(7):731–740. <https://doi.org/10.1093/cercor/bhh033>
- LIU RC, SCHREINER CE (2007) Auditory cortical detection and discrimination correlates with communicative significance. *PLoS Biol* 5(7):e173. <https://doi.org/10.1371/journal.pbio.0050173>
- LORENZI C, GILBERT G, CARN H, GARNIER S, MOORE BCJ (2006) Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci U S A* 103(49):18866–18869. <https://doi.org/10.1073/pnas.0607364103>
- LUTMAN ME, GATEHOUSE S, WORTHINGTON AG (1991) Frequency resolution as a function of hearing threshold and age. *J Acoust Soc Am* 89(1):320–328. <https://doi.org/10.1121/1.400513>
- MACHENS CK, WEHR MS, ZADOR AM (2004) Linearity of cortical receptive fields measured with natural sounds. *J Neurosci* 24(5):1089–1100. <https://doi.org/10.1523/JNEUROSCI.4445-03.2004>
- MANUNTA Y, EDELINE JM (1999) Effects of noradrenaline on frequency tuning of auditory cortex neurons during wakefulness and slow-wave sleep. *Eur J Neurosci* 11(6):2134–2150. <https://doi.org/10.1046/j.1460-9568.1999.00633.x>
- MASSAUX A, EDELINE JM (2003) Bursts in the medial geniculate body: a comparison between anesthetized and unanesthetized states in guinea-pig. *Exp Brain Res* 153(4):573–578. <https://doi.org/10.1007/s00221-003-1516-3>
- MASSAUX A, DUTRIEX G, COTILLON-WILLIAMS N, MANUNTA Y, EDELINE JM (2004) Auditory thalamus bursts in anesthetized and non-anesthetized states: contribution to functional properties. *J Neurophysiol* 91(5):2117–2134. <https://doi.org/10.1152/jn.00970.2003>
- MESGARANI N, DAVID SV, FRITZ JB, SHAMMA SA (2014) Mechanisms of noise robust representation of speech in primary auditory cortex. *Proc Natl Acad Sci U S A* 111(18):6792–6797. <https://doi.org/10.1073/pnas.1318017111>
- MOORE BC, HAFTER ER, GLASBERG BR (1996) The probe-signal method and auditory-filter shape: results from normal- and hearing-impaired subjects. *J Acoust Soc Am* 99(1):542–552. <https://doi.org/10.1121/1.414512>

- MOORE BC, GLASBERG BR, STOEV M, FÜLLGRABE C, HOPKINS K (2012A) The influence of age and high-frequency hearing loss on sensitivity to temporal fine structure at low frequencies (L). *J Acoust Soc Am* 131(2):1003–1006. <https://doi.org/10.1121/1.3672808>
- MOORE BC, VICKERS DA, MEHTA A (2012B) The effects of age on temporal fine structure sensitivity in monaural and binaural conditions. *Int J Audiol* 51(10):715–721. <https://doi.org/10.3109/14992027.2012.690079>
- NAGARAJAN SS, CHEUNG SW, BEDENBAUGH P, BEITEL RE, SCHREINER CE, MERZENICH MM (2002) Representation of spectral and temporal envelope of twitter vocalizations in common marmoset primary auditory cortex. *J Neurophysiol* 87(4):1723–1737. <https://doi.org/10.1152/jn.00632.2001>
- NELSON PB, JIN SH, CARNEY AE, NELSON DA (2003) Understanding speech in modulated interference : cochlear implant users and normal-hearing listeners. *J Acoust Soc Am* 115:2286–2294
- OCCELLI F, SUIED C, PRESSNITZER D, EDELINE JM, GOURÉVITCH B (2016) A neural substrate for rapid timbre recognition? Neural and behavioral discrimination of very brief acoustic vowels. *Cereb Cortex* 26(6):2483–2496. <https://doi.org/10.1093/cercor/bhv071>
- OVERTON JA, RECANZONE GH (2016) Effects of aging on the response of single neurons to amplitude-modulated noise in primary auditory cortex of rhesus macaque. *J Neurophysiol* 115(6):2911–2923. <https://doi.org/10.1152/jn.01098.2015>
- PARAOUTY N, LORENZI C (2017) Using individual differences to assess modulation-processing mechanisms and age effects. *Hear Res* 344:38–49. <https://doi.org/10.1016/j.heares.2016.10.024>
- PARAOUTY N, EWERT SD, WALLAERT N, LORENZI C (2016) Interactions between amplitude modulation and frequency modulation processing: effects of age and hearing loss. *J Acoust Soc Am* 140(1):121–131. <https://doi.org/10.1121/1.4955078>
- PATTERSON RD (1987) A pulse ribbon model of monaural phase perception. *J Acoust Soc Am* 82(5):1560–1586. <https://doi.org/10.1121/1.395146>
- PETERS RW, MOORE BCJ (1992) Auditory filters and aging: filters when auditory thresholds are normal. In: Cazals Y, Demany L, Horner K (eds) *Auditory physiology and perception*. Pergamon, Oxford, pp 179–185. <https://doi.org/10.1016/B978-0-08-041847-6.50026-5>
- PRESACCO A, SIMON JZ, ANDERSON S (2016) Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J Neurophysiol* 116(5):2346–2355. <https://doi.org/10.1152/jn.00372.2016>
- RANASINGHE KG, VRANA WA, MATNEY CJ, KILGARD MP (2012) Neural mechanisms supporting robust discrimination of spectrally and temporally degraded speech. *J Assoc Res Otolaryngol* 13(4):527–542. <https://doi.org/10.1007/s10162-012-0328-1>
- RENNAKER RL, CAREY HL, ANDERSON SE, SLOAN AM, KILGARD MP (2007) Anesthesia suppresses nonsynchronous responses to repetitive broadband stimuli. *Neuroscience* 145(1):357–369. <https://doi.org/10.1016/j.neuroscience.2006.11.043>
- ROMANSKI LM, AVERBECK BB (2009) The primate cortical auditory system and neural representation of conspecific vocalizations. *Annu Rev Neurosci* 32(1):315–346. <https://doi.org/10.1146/annurev.neuro.051508.135431>
- ROSS B, FUJIOKA T, TREMBLAY KL, PICTON TW (2007) Aging in binaural hearing begins in mid-life: evidence from cortical auditory-evoked responses to changes in interaural phase. *J Neurosci* 27(42):11172–11178. <https://doi.org/10.1523/JNEUROSCI.1813-07.2007>
- SAYLES M, WINTER IM (2010) Equivalent-rectangular bandwidth of single units in the anesthetized guinea-pig ventral cochlear nucleus. *Hear Res* 262(1–2):26–33. <https://doi.org/10.1016/j.heares.2010.01.015>
- SCHNUPP JW, HALL TM, KOKELAAR RF, AHMED B (2006) Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *J Neurosci* 26(18):4785–4795. <https://doi.org/10.1523/JNEUROSCI.4330-05.2006>
- SCHOOF T, ROSEN S (2014) The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Front Aging Neurosci* 6:307. <https://doi.org/10.3389/fnagi.2014.00307>
- SCHVARTZ KC, CHATTERJEE M, GORDON-SALANT S (2008) Recognition of spectrally degraded phonemes by younger, middle-aged, and older normal-hearing listeners. *J Acoust Soc Am* 124(6):3972–3988. <https://doi.org/10.1121/1.2997434>
- SERGEYENKO Y, LALL K, LIBERMAN MC, KUJAWA SG (2013) Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *J Neurosci* 33(34):13686–13694. <https://doi.org/10.1523/JNEUROSCI.1783-13.2013>
- SHAMMA S, LORENZI C (2013) On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *J Acoust Soc Am* 133(5):2818–2833. <https://doi.org/10.1121/1.4795783>
- SHANNON CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- SHANNON RV, ZENG FG, KAMATH V, WYGONSKI J, EKELID M (1995) Speech recognition with primary temporal cues. *Science* 270(5234):303–304. <https://doi.org/10.1126/science.270.5234.303>
- SHEFT S, ARDOINT M, LORENZI C (2008) Speech identification based on temporal fine structure cues. *J Acoust Soc Am* 124(1):562–575. <https://doi.org/10.1121/1.2918540>
- SHELDON S, PICHORA-FULLER MK, SCHNEIDER BA (2008) Effect of age, presentation method, and learning on identification of noise-vocoded words. *J Acoust Soc Am* 123(1):476–488. <https://doi.org/10.1121/1.2805676>
- SMITH ZM, DELGUTTE B, OXENHAM AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416(6876):87–90. <https://doi.org/10.1038/416087a>
- SÖRÖS P, TEISMANN IK, MANEMANN E, LÜTKENHÖNER B (2009) Auditory temporal processing in healthy aging: a magnetoencephalographic study. *BMC Neurosci* 10(1):34. <https://doi.org/10.1186/1471-2202-10-34>
- SOUZA PE, BOIKE KT (2006) Combining temporal-envelope cues across channels: effects of age and hearing loss. *J Speech Lang Hear Res* 49(1):138–149. [https://doi.org/10.1044/1092-4388\(2006/011\)](https://doi.org/10.1044/1092-4388(2006/011))
- SOUZA PE, KITCH V (2001) The contribution of amplitude envelope cues to sentence identification in young and aged listeners. *Ear Hear* 22(2):112–119. <https://doi.org/10.1097/00003446-200104000-00004>
- STONE MA, FÜLLGRABE C, MACKINNON RC, MOORE BCJ (2011) The importance for speech intelligibility of random fluctuations in “steady” background noise. *J Acoust Soc Am* 130(5):2874–2881. <https://doi.org/10.1121/1.3641371>
- TAKAHASHI GA, BACON SP (1992) Modulation detection, modulation masking, and speech understanding in noise in the elderly. *J Speech Hear Res* 35(6):1410–1421. <https://doi.org/10.1044/jsrh.3506.1410>
- TER-MIKHAELIAN M, SEMPLE MN, SANES DH (2013) Effects of spectral and temporal disruption on cortical encoding of gerbil vocalizations. *J Neurophysiol* 110(5):1190–1204. <https://doi.org/10.1152/jn.00645.2012>
- WALLACE MN, PALMER AR (2008) Laminar differences in the response properties of cells in the primary auditory cortex. *Exp Brain Res* 184(2):179–191. <https://doi.org/10.1007/s00221-007-1092-z>
- WALLACE MN, RUTKOWSKI RG, PALMER AR (2000) Identification and localisation of auditory areas in guinea pig cortex. *Exp Brain Res* 132(4):445–456. <https://doi.org/10.1007/s002210000362>

- WALLAERT N, MOORE BC, LORENZI C (2016) Comparing the effects of age on amplitude modulation and frequency modulation detection. *J Acoust Soc Am* 139(6):3088–3096. <https://doi.org/10.1121/1.4953019>
- WALLAERT N, MOORE BC, EWERT SD, LORENZI C (2017) Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation. *J Acoust Soc Am* 141(2):971–980. <https://doi.org/10.1121/1.4976080>
- WALTON JP, FRISINA RD, O'NEILL WE (1998) Age-related alteration in processing of temporal sound features in the auditory midbrain of the CBA mouse. *J Neurosci* 18(7):2764–2776
- WHITMAL NA, POISSANT SF, FREYMAN RL, HELFER KS (2007) Speech intelligibility in cochlear implant simulations: effects of carrier type, interfering noise, and subject experience. *J Acoust Soc Am* 122(4):2376–2388. <https://doi.org/10.1121/1.2773993>
- WOOLLEY SM, GILL PR, THEUNISSEN FE (2006) Stimulus-dependent auditory tuning results in synchronous population coding of vocalizations in the songbird midbrain. *J Neurosci* 26(9):2499–2512. <https://doi.org/10.1523/JNEUROSCI.3731-05.2006>
- ZENG FG, NIE K, STICKNEY GS, KONG YY, VONGPHOE M, BHARGAVA A, WEI C, CAO K (2005) Speech recognition with amplitude and frequency modulations. *Proc Natl Acad Sci* 102(7):2293–2298. <https://doi.org/10.1073/pnas.0406460102>