From bird to sparrow: Learning-induced modulations in fine-grained semantic discrimination

Rosanna De Meo [a,*], Nathalie M.-P. Bourquin [a,1], Jean-François Knebel [a,b,c], Micah M. Murray [a,b,c,d], Stephanie Clarke [a,*]

[a] Neuropsychology and Neurorehabilitation Service, Department of Clinical Neurosciences, Vaudois University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland
[b] Radiology Department, Vaudois University Hospital Center and University of Lausanne, 1011 Lausanne, Switzerland
[c] Electroencephalography Brain Mapping Core, Center for Biomedical Imaging (CIBM) of Lausanne and Geneva, 1011 Lausanne, Switzerland
[d] Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN 37235, USA

ARTICLE INFO

ABSTRACT

Recognition of environmental sounds is believed to proceed through discrimination steps from broad to more narrow categories. Very little is known about the neural processes that underlie fine-grained discrimination within narrow categories or about their plasticity in relation to newly acquired expertise. We investigated how the cortical representation of birdsongs is modulated by brief training to recognize individual species. During a 60-minute session, participants learned to recognize a set of birdsongs; they improved significantly their performance for trained (T) but not control species (C), which were counterbalanced across participants. Auditory evoked potentials (AEPs) were recorded during pre- and post-training sessions. *Pre* vs. *post* changes in AEPs were significantly different between T and C i) at 206–232 ms post stimulus onset within a cluster on the anterior part of the left superior temporal gyrus; ii) at 246–291 ms in the left middle frontal gyrus; and iii) 512–545 ms in the left middle temporal gyrus as well as bilaterally in the cingulate cortex. All effects were driven by weaker activity for T than C species. Thus, expertise in discriminating T species modulated early stages of semantic processing, during and immediately after the time window that sustains the discrimination between human vs. animal vocalizations. Moreover, the training-induced plasticity is reflected by the sharpening of a left lateralized semantic network, including the anterior part of the temporal convexity and the frontal cortex. Training to identify birdsongs influenced, however, also the processing of C species, but at a much later stage. Correct discrimination of untrained sounds seems to require an additional step which results from lower-level features analysis such as apperception. We therefore suggest that the access to objects within an auditory semantic category is different and depends on subject's level of expertise. More specifically, correct intra-categorical auditory discrimination for untrained items follows the temporal hierarchy and transpires in a late stage of semantic processing. On the other hand, correct categorization of individually trained stimuli occurs earlier, during a period contemporaneous with human vs. animal vocalization discrimination, and involves a parallel semantic pathway requiring expertise.

© 2014 Elsevier Inc. All rights reserved.

Introduction

Similar to the segregation in the visual stream, regions within the superior and middle temporal gyri and the inferior frontal gyrus have been identified as candidates for the "what" pathway of the auditory stream which subserves sound identity processing in non-human primates (Romanski et al., 1999; Rauschecker and Tian, 2000; Wessinger et al., 2001; Romanski and Goldman-Rakic, 2002) and humans (Clarke et al., 2000; Alain et al., 2001; Kaiser and Lutzenberger, 2003; Arnott et al., 2004). This network extends from primary (core) fields to regions of the anterior part of the temporal convexity as well as inferior frontal gyrus (Maeder et al., 2001; Bizley and Cohen, 2013). Discrimination capacities along this pathway remain largely underdetermined particularly in humans, and were the focus of the present study.

Sounds of different categories yield partially distinct activation patterns within the temporo-frontal cortex (Scott, 2005). Broad categories of environmental sounds, such as those produced by living vs. non-living sources were shown to activate differentially parts of the temporo-parieto-frontal cortex (Engel et al., 2009; Murray et al., 2006). The comparison between brain networks responsive to specific categories highlighted the contribution of a

* Corresponding authors at: Service de Neuropsychologie et Neuroréhabilitation, Hôpital Nestlé, Avenue Pierre-Decker 5, 1011 Lausanne, Switzerland.
E-mail addresses: Rosanna.De-Meo@chuv.ch (R. De Meo), Stephanie.Clarke@chuv.ch (S. Clarke).
[1] Current address: School of Health – Geneva- University of Applied Sciences Western Switzerland, Avenue de Champel 47, 1206 Geneva, Switzerland.

predominantly left hemisphere network centered on the supratemporal plane and the superior temporal gyrus during the analysis of animal vocalizations as compared to tool sounds (Doehrmann et al., 2008; Lewis et al., 2005). Category-specificity, not found in the primary auditory cortex and its vicinity, predominates within the anterior part of the temporal cortex, where category-selective clusters were identified (Leaver and Rauschecker, 2010). These results speak in favor of a hierarchical organization of the ventral auditory pathway, which supports increasingly complex representations of sound objects, including that of narrowly defined categories in the most anterior regions (Rauschecker and Scott, 2009; Leaver and Rauschecker, 2010; Peelle et al., 2010; Romanski et al., 1999). Concurring evidence comes from single-unit recordings in non-human primates, which has shown that stimulus selectivity to monkey calls and other meaningful sounds, as well as response latencies increase progressively from the primary auditory cortex to rostrotemporal areas (Recanzone, 2008; Kikuchi et al., 2010). A series of auditory evoked potential (AEP) studies led to the formulation of the _temporal hierarchy model_ (De Lucia et al., 2010; reviewed in Spierer et al., 2011). The differential processing between sounds of living vs. man-made categories was shown to start as early as 70 ms post stimulus onset, with differential activity within the right temporal and left inferior frontal cortices (Murray et al., 2006). Starting at approximately 170 ms, neural activity within the right superior temporal cortex distinguishes the (non-verbal) vocalizations of humans vs. animals (Renvall et al., 2012; De Lucia et al., 2010). Within the category of man-made sounds, those which typically cue for a responsive action by the listener (e.g. a ringing telephone) and those which do not yield different neural activity in the premotor and inferior prefrontal cortex starting at 300 ms (De Lucia et al., 2009).

Despite this accumulating knowledge regarding the representation of different semantic categories of sounds, little is known about the mechanisms involved in the discrimination within narrow categories. Intra-categorical discrimination has been well studied in the visual modality with notably face discrimination. It is well established that faces are processed to a large degree in a specific brain area, called the fusiform face area (FFA), which detects faces and extracts the necessary perceptual information to recognize them (for a review, see Kanwisher and Yovel, 2006). In the auditory modality, the typical example of intra-categorical discrimination is conspecific vocalization discrimination. In humans specific brain regions exhibiting differential responses to voices have been identified within the middle and anterior superior temporal sulcus (Belin et al., 2000, 2002; De Lucia et al., 2010; reviewed in Belin, 2006), and at a specific time-window, i.e. peaking at 320 ms after stimulus onset (Charest et al., 2009; Levy et al., 2001). It has been suggested that the voice-selective areas within the superior temporal sulcus (STS) may represent the counterpart of the face-selective area in human visual cortex (Belin et al., 2000, 2004). However, it is still unclear whether this discrimination relies on domain-specific (Belin et al., 2004; Campanella and Belin, 2007) or domain-general processes (Chartrand et al., 2008; Poremba et al., 2013); the latter of which would arise from learned expertise. One way to address this topic and more generally the issue of how the brain establishes and maintains a categorical representation for a given set of objects is via training-induced plasticity. This has already been proven effective in studies of face discrimination (Tanaka and Curran, 2001) and auditory spatial representations (Spierer et al., 2007). An important consideration in applying this approach is the use of stimuli for which the subjects are not already experts in discriminating (cf. Tanaka and Curran, 2001), which was the tactic adopted in the present study.

We investigated how the cortical representation of a narrow category, here birdsongs, is modulated by training non-experts to recognize individual species. We hypothesized that learning to identify a small number of species will improve their recognition and will modulate the spatio-temporal patterns of neural activity associated with their representation. We expected these modulations to occur along the

ventral auditory stream, most likely within the semantic network in the vicinity of the anterior part of the temporal convexity (Altmann et al., 2007) and lead possibly to a more refined representation of the trained items (Spierer et al., 2007). Two hypotheses were formulated as to the time-point of differential processing of the trained species. First, the newly acquired expertise might be associated with processing changes occurring roughly within the same time period as the discrimination of human vs. non-human vocalizations, i.e. at 169–219 ms post-stimulus onset (De Lucia et al., 2010), assuming that intra- and inter-categorical expertise rely on similar mechanisms. Alternatively, the new expertise might be associated with processing changes occurring much later, assuming that intra-categorical processing constitutes a subsequent hierarchical step.

Material and methods

Participants

Nineteen healthy right-handed (eleven females) native French-speakers, aged 21–33 years (mean age 24 ± 3 years) participated in the experiment. All participants provided written, informed consent to participate in the study; the procedures of which were approved by the Ethics Committee of the University of Lausanne. None had a history of neurological or psychiatric illnesses, and all reported normal hearing. None of the participants had extensive knowledge about birdsongs or particular musical training. Six participants were excluded from the study: because of excessive EEG artifacts (5 subjects) or poor performance at birdsong discrimination (1 subject), leaving 13 subjects (8 females) for behavioral and EEG analyses.

Stimuli

The auditory stimuli were songs of 12 different bird species (16 bit mono; 22050 Hz digitization, processed with Adobe Audition CS5.5) of which 6 were native birds (i.e. ubiquitous in Switzerland: Common Buzzard (_Buteo buteo_), Eurasian Jay (_Garrulus glandarius_), Blue Tit (_Cyanistes coeruleus_), House Sparrow (_Passer domesticus_), Black Woodpecker (_Dryocopus martinus_), Eurasian Wren (_Troglodytes troglodytes_)) and 6 were non-native birds (i.e. not generally encountered in Switzerland: Spotted Flycatcher (_Muscicapa striata_), Northern Mockingbird (_Mimus polyglottos_), Ruddy Shelduck (_Tadorna ferruginea_), Great Crested Grebe (_Podiceps cristatus_), Great Northern Loon (_Gavia immer_), Laughing Kookaburra (_Dacelo novaeguineae_)). For each species, a long excerpt of their song was downloaded from websites specialized in ornithology (www.randonneur.net; www.web-ornitho.com; www.oiseaux-faune.net; www.sound-fishing.net). Then, two sets of sounds were created for each bird: a) a 6 second set (including 50 ms of rising and decay time); and b) a second set composed of 6 short excerpts of 1 second in duration (including 50 ms rising and decay time), which had been extracted from the 6 second sound files. All sounds were normalized according to their peak level. From set b), 3 excerpts for each species were used during the _Pre- and Post-training sessions_ and the other 3 excerpts during the training tests (see procedure below). Half of set a), 3 native and 3 non-native species, were used for _Training_ condition ("_T_") and the other 3 natives and 3 non-natives as _Control_ condition ("_C_"). The _T_ and _C_ lists were counterbalanced across participants.

Analysis of sounds

The following procedure was used to minimize the possibility that acoustical features did not confound the effect of learning. First, the stimuli were subdivided into two sets, each of which was composed of 3 native and 3 non-native species, and contained three 1 second excerpts per species. The species and excerpts were randomly assigned to _set 1 and set 2_. The power spectrogram of each selected excerpt was performed using a short-time Fourier transform (STFT) with a time-