serval
serveur académique lausannois

UNIL | Université de Lausanne
Faculté de biologie
et de médecine

Invited review

Roaring lions and chirruping lemurs: how the brain encodes sound objects in space

Running title: **Sound objects in space**

Stephanie Clarke & Eveline Geiser

Service de Neuropsychologie et de Neuroréhabilitation, Centre Hospitalier Universitaire

Vaudois and Université de Lausanne, Lausanne, Switzerland.

Corresponding author:
Prof. Stephanie Clarke
Service de Neuropsychologie et de Neuroréhabilitation,
CHUV, Université de Lausanne,
1011 Lausanne, Switzerland.
Stephanie.Clarke@chuv.ch

28 pages

3 figures (all in color)

Conflicts of interest: none

.

Abstract

The dual-stream model of auditory processing postulates separate processing streams for sound meaning and for sound location. The present review draws on evidence from human behavioral and activation studies as well as from lesion studies to argue for a position-linked representation of sound objects that is distinct both from the position-independent representation within the ventral/What stream and from the explicit sound localization processing within the dorsal/Where stream.

Abbreviations

AA = anterior auditory area

AEP = auditory evoked potential

ALA = anterolateral auditory area

IID = interaural intensity difference

ITD = interaural time difference

LA = lateral auditory area

fMRI = functional magnetic resonance imaging

EEG = electroencephalography

MEG = magnetoencephalography

1. Introduction

While crossing the African continent, Dr. David Livingstone described the sound-dominated jungle at night: "The nights are warmer now ... A new world awakes and comes forth, more numerous, if we may judge by the noise it makes ... Lions and hyenas roar around us, and sometimes come disagreeably near, though they have never ventured into our midst. Strange birds sing their agreeable songs, while others scream and call harshly ... Marvellous insect-sounds fall upon the ear ... A little lemur was once seen to leap about from branch to branch … it chirruped like a bird" (Livingstone, 1875, p. 153). David Livingstone's description illustrates the importance of the correct perception of sound source and origin not only for orientation but, ultimately, for survival in the wild. In the near absence of visual information, the identity of an animal and its location have to be determined from sound alone. Less dramatic situations illustrate the everyday relevance of combined semantic[1] and spatial[2] information. For example, identifying a ring tone as coming from my pocket helps me to identify the phone as my own. Based on experimental research, we know that the combination of information about the identity and location of a sound object helps to separate sound sources and to identify sound objects more readily, particularly in noisy surroundings (Bregman 1990; Yost 1991). The present review addresses a series of seminal studies that explored the processing steps leading from low level analysis of spectro-temporal features to high level semantic processing of environmental sounds. Both high and low level processing steps integrate spatial information derived from binaural cues. We summarize current evidence on hierarchical processing in the auditory cortex, primarily describing the dual-stream model of auditory processing (section 2). We outline neuroimaging evidence for

---

[1] The term "semantic" sometimes has linguistic connotations, but is here used in a broader, nonlinguistic sense, referring to acoustic information or processing that can lead to the identification of sound objects.

[2] The term "spatial" relates here to the positions of sound objects in space. In psychophysical and imaging studies spatial aspects are investigated in free-field conditions, using loudspeakers, or by simulating locations by means of head-related transfer functions or binaural cues. When referring to a specific study spatial cues denote here free-field or head-related transfer function conditions, and binaural cues interaural time or intensity differences.

distinct neuronal populations that combine semantic and spatial information and may form a distinct, position-linked representation of sound objects (section 3). We argue that the combination of semantic and spatial information may take place in an early phase of auditory processing, as suggested by behavioral and neuroimaging evidence on auditory streaming (section 4). Finally, we summarize the evidence for a functional distinction between explicit and implicit processing of spatial cues during auditory processing (section 5).

2. The ventral and dorsal auditory streams

A large body of evidence from human and non-human studies shows that information pertaining to the identity and location of a sound object is processed along two separate anatomical pathways, ventral and dorsal from the primary auditory cortex. The two anatomical pathways and their functional distinctions have been described as the dual-stream[3] model of auditory processing (Rauschecker 1998; Rauschecker and Scott 2009). This seminal work provided evidence of a ventral stream for sound recognition from the anterior part of the auditory cortex to the inferior frontal cortex and a dorsal stream for auditory spatial processing from the posterior part of the auditory cortex to the superior frontal and parietal cortices.

Non-human primate studies that support the dual pathway model include the observation of two distinct anatomical connections to the prefrontal cortex from electrophysiologically identified auditory areas (Romanski et al. 1999a; Romanski et al. 1999b; for review Kaas et al. 1999; Rauschecker and Tian 2000; Kaas and Hackett 2000). Two streams have also been demonstrated in humans. Imaging studies have localized the

---

[3] The term "stream" denotes two different concepts. In the context of the dual-stream model, which is derived from electrophysiological and anatomical studies in animals, "stream" denotes a processing pathway characterized by a specific connection pattern and along which functional specialization proceeds hierarchically (e. g. Rauschecker and Scott 2009). In the context of auditory scene analysis, which is investigated by psychophysical studies, "stream" denotes the percept of a sound coming from a specific sound source (e. g. Bregman 1990; Yost 1991).

recognition and identification of sounds to the anterior temporal convexity and their

localization to the parietal convexity (Anourova et al., 2001; Maeder et al., 2001; Hart et al.

2004; Ahveninen et al. 2006; De Santis et al. 2007a; 2007b; for review Arnott et al. 2004).

This dichotomy has been explicitly shown in studies that used identical sounds for

identification and localization. For example, when listeners discriminate the pitch of band

noise bursts, the superior temporal and inferior frontal cortices are activated. In contrast, when

they process the location of the same sounds, the parietal and superior frontal cortices are

activated (Alain et al. 2001a). The same dichotomy has been demonstrated with meaningful

sounds. Listening to the same or different animal vocalizations activates the anterior superior

temporal gyrus and sulcus, including the planum polare and the lateral part of Heschl's gyrus,

whereas changes in sound location activate the posterior superior temporal gyrus, including

the planum temporale (Altmann et al. 2008).

The selectivity of the ventral stream for the meaning of sounds has been confirmed in

a series of studies using different categories of environmental sounds as stimuli (e.g., tools,

animals, man-made objects, living and non-living sound sources, actions, and musical

instruments; Lewis et al. 2005; Murray et al. 2006; Altmann et al. 2007; Engel et al. 2009;

Leaver and Rauschecker 2010; Lewis et al. 2011). The processing which leads eventually to

sound recognition involves two sequential steps, which are hierarchically organized. First,

spectro-temporal features are processed in early-stage auditory areas (Altmann et al. 2008).

Then, higher-order areas along the ventral stream encode the semantic aspects of sounds

(Bergerbest et al. 2004; Doehrmann et al. 2008; Murray et al. 2008; De Lucia et al. 2010;

Kumar et al. 2007; Griffiths et al. 2007; De Meo et al. 2015).

Although the ventral stream is the key structure for sound recognition, specific sound

categories involve additional parts of the frontal cortex. In particular, action-related

environmental sounds have been shown to activate parts of the motor, premotor and prefrontal

cortices that are classically associated with the dorsal stream (Lahav et al. 2007; Lewis et al. 2005; Pizzamiglio et al. 2005; Gazzola et al. 2006;Hauk et al. 2006; Doehrmann et al. 2008; De Lucia et al. 2009). Furthermore, evoked muscular potentials induced by transcranial magnetic stimulation of the motor cortex are significantly increased when participants listen to action-related compared to action-unrelated sounds (Aziz-Zadeh et al. 2004; Bourquin et al. 2013b). The above findings indicate that brain areas of the dorsal stream are sensitive to the semantic category of auditory stimuli when this category involves motor action.

The functional distinction between the ventral and the dorsal processing stream is further highlighted by the distinct processing deficits that result from brain lesions. Circumscribed lesions centered on the temporal convexity result in deficits in sound recognition, while both sound localization and sound motion perception are preserved; conversely, lesions centered on the right parietal convexity are associated with auditory spatial deficits but not with sound recognition deficits (Clarke et al. 2000; 2002; Bellmann et al. 2001; Adriani et al. 2003a; Thiran and Clarke 2003; Ducommun et al. 2004; Rey et al. 2007; Spierer et al. 2009). It is important to note that these specific deficits are likely not the result of postlesional reorganization. Transient dysfunction induced by repetitive or chronometric single-pulse transcranial magnetic stimulation (TMS) does not allow participants to compensate through neural reorganization. TMS to posterior parietal cortex in healthy individuals disrupts processing of binaural cues (At et al. 2011, see also Lewald et al. 2004). Thus, a spatial processing deficit, similar to the one observed in patients with brain lesions, can be induced by temporally disabling a similar brain region, indicating that the observed deficits in patients are likely not influenced by postlesional reorganization.

Thus, the dual-stream model suggests parallel processing routes in the brain, one for semantic information and one for spatial. This model is also illustrated by the conscious experience of brain-damaged patients, who can perceive the identity of a sound object but do

not know where its source is located, or, conversely, who can indicate perfectly the location

of an object that they cannot recognize (Clarke et al. 2000; 2002; Adriani et al. 2003b).

However, auditory experience already indicates that the representation of a sound object in

space, such as a chirruping lemur leaping from branch to branch, relies on a close interaction

between semantic and spatial encoding. This interaction is likely to occur at different levels of

processing, from spectro-temporal to semantic and from binaural to high level auditory spatial

encoding, as described in the following section.

3. Position-linked and position-independent representations of sound objects

The dual-stream model implies that semantic representations within the ventral stream

are position-independent, i.e., not modulated by the spatial attributes of the sound objects. The

majority of previous activation studies that investigated the neural representations of sound

identity were not designed to find location-linked object representations (Bergebest et al.

2004; Murray et al. 2008; Altmann et al. 2007; Doehrmann et al. 2008). However, a position-

linked representation of sound objects was recently identified in an AEP study using

repetition priming (Bourquin et al. 2013a, Fig. 1). Repetition priming is often used to identify

neural populations that are sensitive to a specific stimulus characteristic (Grill-Spector et al.

2006). Thus, the use of a repetition priming paradigm allowed us to identify neural

populations that encoded sound objects either independently of spatial attributes or as a

function of their position. Environmental sounds were first presented lateralized to the left or

the right. Then, the lateralization was either held constant or changed in a second presentation

of the same object. Two types of repetition effects were observed in this study. A first

analysis compared the initial vs repeated presentations, independently of whether the

lateralization of a given sound object was held constant or changed between the initial and

repeated presentation. This comparison revealed the classical repetition priming suppression

(e.g. Murray et al. 2008) independent of lateralization, as postulated in the dual-stream model. It involved regions associated with the ventral stream, specifically the right temporo-frontal region at 42-63 ms and the left temporo-parietal region at 165-215 ms. A second analysis compared repetition priming effects which were yielded by pairs with the same vs different lateralization; this analysis revealed time windows and regions, where neuronal populations kept track of the constant vs changed position of a given sound object. Position-linked repetition effects occurred within the left hemisphere, in the posterior part of the superior and middle temporal gyri at 20-39 ms post-stimulus onset and in the left inferior and middle frontal gyri at 143-162 ms. These findings were based on the analysis of distributed source estimations, giving a rough estimate of the cortical region where the repetition-sensitive neural populations are located. These regions may form a network of interconnected areas. The posterior parts of the superior and middle temporal gyri are strongly connected with the planum temporale (Cammoun et al. 2014). These areas are believed to be part of the dorsal stream, as is the homologous region in non-human primates (Romanski et al. 1999a; Tian et al. 2001). The position-linked representation in the posterior parts of the inferior and middle frontal gyri was more surprising. Its homolog in non-human primates contains neurons with non-spatial object selectivity and has been proposed to play a role in non-spatial cognition (Cohen et al. 2009). However, anatomical studies have revealed possible inputs from both the dorsal and the ventral streams. Tracing studies in non-human primates have described an overlap of projections from electrophysiologically identified rostral and caudal belt areas (Romanski et al. 1999a). Furthermore, Brodmann areas 44 and 45, which are part of this region, have been shown to receive partially overlapping projections from the inferior parietal lobule and the superior temporal sulcus (Petrides and Pandya 2009; Frey et al. 2014). Human in vivo and post-mortem tracing studies have investigated the connectivity of areas 44 and 45. Using diffusion fiber tractography, area 44 has been found to be interconnected with the

inferior parietal lobule via the superior longitudinal fasciculus. The adjacent area 45 was shown to be interconnected with the anterior part of the superior temporal gyrus via the extreme capsule (Frey et al. 2008). Human post-mortem tracing studies have revealed long horizontal intrinsic connectivity within the posterior inferior frontal gyrus (Tardif et al. 2007), which may contribute to the integration of spatial and non-spatial inputs. Together, these findings indicate that a specific region, located in the posterior inferior frontal cortex and in the posterior part of the middle temporal gyrus, is likely to integrate both semantic and spatial information, as reflected in its reported location-linked object representation (Bourquin et al. 2013a).

Current evidence suggests that the interaction between semantic and spatial information occurs at different stages of auditory processing. The combined encoding of spectro-temporal and binaural information has been demonstrated in several early-stage auditory areas. Neurons in the macaque's auditory caudal belt areas show combined selectivity for spectro-temporal cues and spatial positions (Tian et al. 2001). In humans, a similar combined selectivity has been found on the supratemporal plane[4]. The anterior lateral area (ALA), a belt area on the antero-lateral part of Heschl's gyrus (Wallace et al. 2002), can process information related to sound identity (Viceic et al. 2006) as well as combination of binaural information with spectro-temporal (Budd et al. 2003; Hall et al. 2005) or sound identity cues (van der Zwaag et al. 1011). The planum temporale, which comprises posterior belt areas, contains neuronal populations that process the identity of sound objects (Da Costa et al. 2015); spatial (Deouell et al. 2007) or binaural cues (Krumbholz et al. 2005); or combined information about frequency and sound location (Shrem and Deouell 2014). In

---

[4] In view of the growing body of studies involving the supratemporal plane, there is a need to establish a common way of identifying early stage auditory areas. The use of tonotopic mapping for the identification of the primary auditory cortex (Saenz and Langer 2013) has revealed robust, mirror-reversed representations on Heschl's gyrus, very likely corresponding to the homologs of AI and R (Da Costa et al. 2011; 2013). The results of histological studies offer the possibility of identifying individual areas, in particular because two independent groups have identified the same layout (Rivier and Clarke 1997; Wallace et al. 2002). The use of functional localizers, especially those that take hierarchical processing into account (Chevillet et al. 2011), is an excellent tool.

humans, belt areas are strongly interconnected, as shown by diffusion spectrum imaging (Cammoun et al. 2014; Fig. 2), and are very likely the starting point of separate processing streams for sound recognition, sound localization, and potentially a combination of the two (Clarke and Morosan 2013). An early interaction between the processing of frequency and binaural cues was suggested by a previous psychophysical study (Tardif et al. 2008), which reported that discriminating near-threshold frequency stimuli was improved by task-irrelevant changes in binaural cues; the reverse effect, improved discrimination of binaural cues by task-irrelevant frequency changes, has been observed as well. In summary, the interaction between frequency and spatial cues, as it occurs on the supratemporal plane, could represent an intermediate step that leads to the location-linked representation of sound objects.

Second, semantic and spatial information can also be combined at later stages, namely in the posterior part of the inferior frontal cortex, which has been shown to hold location-linked object representations (Bourquin et al. 2013a). The later convergence of spatial and semantic representations is supported by the anatomical convergence of the ventral and dorsal streams within this regions in humans (Frey et al. 2008) as well as the presence of monosynaptic inputs from rostral and caudal belt areas (Romanski et al. 1999a) and from the inferior parietal lobule and the superior temporal sulcus (Frey et al. 2014) to the homologous primate region.

4. The interaction between semantic and spatial information – evidence from auditory scene analysis

In real-life auditory environments, listeners rarely encounter one sound source only. What reaches the ear is a mixture of acoustic cues originating, for example, from a bird singing on the far left together with a lemur chirping on the right-hand side. Listeners segment these sounds into perceptual constructs linked to a specific source - the sound object. This

perceptual organization is referred to as streaming (Carlyon 2004; Bregman 1990). Auditory streaming is a basic capacity of the auditory system and is influenced by the acoustic cues of the stimulus as well as by higher-level semantic processes (Billig et al. 2013). Whereas frequency and timbre are some of the most powerful acoustic cues used for auditory streaming (for reviews see e. g. Darwin 1999; Carlyon 2004; Moore and Gockel 2012), spatial cues can play an important role in auditory stream segregation as well. In laboratory settings spatial cues are provided by the positions of loudspeakers or simulated by either head-related transfer functions or binaural time or intensity differences. Knowing from which loudspeaker a sound will be coming facilitated the detection of one amidst other sounds (Best et al. 2007). Similarly, when simultaneous distractors were presented, the encoding of auditory targets was improved by spatial continuity; thus, participants performed better when the target, e.g., spoken digits, is presented at the same position than when the spatial position of the target is changed (Best et al. 2008; Best et al. 2010). In the above experiments listeners were attentive to sound positions. However, binaural cues were shown to facilitate auditory stream segregation even in the absence of attention. Thus, binaural cues, as provided by interaural time differences (ITDs), facilitated speech perception (Schubert 1956; Schubert and Schultz 1962; Carhart et al. 1967; Carhart et al. 1968; Carhart et al. 1969b). Moreover, once listeners perceive speech, ITDs are used to track it over time (Darwin and Hukin 1999). Spatial cues may also influence earlier stages of stream segregation. For example, a difference in the location of targets (animal cries or musical instruments; positions simulated by head-related transfer functions) and distracters enhances the detection of short (250 ms), but not longer (500 ms), sounds (Eramudugolla et al. 2008). Assuming that stream segregation develops over time and that short exposure consequently hinders complete segregation, the authors interpreted their findings as evidence that spatial cues are effective in enhancing particularly

early processes of stream segregation. The evidence described above shows that spatial cues facilitate auditory stream segregation when processed both explicitly and implicitly.

The use of spatial cues in auditory streaming may further depend on the nature of the perceived sound object. Specifically, whether a sound has a meaning - a prior semantic representation - may influence the use of spatial cues in stream segregation. Meaningful stimuli include vowels; environmental sounds such as birds, musical instruments, human voices; sounds of a recognizable pitch; and previously heard rhythmic noise busts. Listeners detect and identify these meaningful stimuli more easily when they are spatially distinct from a distractor stimulus. For example, the perception of spoken words was improved by binaural cues (Schubert 1956; Schubert and Schultz 1962; Carhart et al. 1967; Carhart et al. 1968; Carhart et al. 1969b; Darwin and Hukin 1999) and that of whispered vowels by spatial separation of the loudspeakers (Drennan et al. 2003). Similarly, the discrimination of noise-burst rhythms, which were masked by additional noise bursts, has been shown to improve with the spatial separation of the loudspeakers (Middlebrooks and Onsan 2012). In contrast, the effect of spatial cues on the segregation of meaningless, previously unheard sounds is less clear. Meaningless sounds can, for example, contain the basic spectro-temporal structure of natural sounds but lack strong grouping cues (McDermott et al. 2011). Although these stimuli could be streamed based on ITD cues alone, the perceived content of the source has been shown to be inaccurate when other segregation cues, such as harmonicity and common onsets and offsets, are missing (Schwartz et al. 2012). In the experimental setting of that study, subjects were asked to judge whether or not the target, which was presented simultaneously with a distracter, and the probe that followed were the same. Performance was better when the target and distracter differed in ITD and when the lateralization of the target was cued prior to its presentation. These gains were, however, very small compared with the situation when multiple distracters were replaced by repetition of the same distracter. Furthermore, ITD cues

improved the same-different discrimination only when the target and the distracter had minimal spectro-temporal overlap and not when there was more overlap. The small effect of ITD cues on the same-different discrimination contrasts with the accurate localization of targets and distracters. Taken together, these studies suggest that spatial cues differentially contribute to streaming, depending on whether the target is a priori known or unknown. Spatial cues appear to play a minor role in the streaming of novel stimuli that do not have pre-existing semantic representations; they are sufficient for the localization of the stimuli but only marginally improve their identification. This weak influence of spatial cues on the streaming of meaningless sounds contrasts with their much more prominent role in the segregation of sound objects that have a distinct meaning and a prior representation (e. g., Darwin and Hukin 1999; Best et al. 2008). This is evidence for an interaction between semantic and spatial information in the process of sound stream segregation. Further studies are necessary to confirm that the processing of sounds that lack a prior representation is less influenced by spatial cues.

The neural substrate of streaming on the basis of spatial cues is still elusive because neuroimaging studies are relatively scarce. As pointed out in fMRI studies streaming correlates with activation of the supratemporal plane. This is the case for streaming based both on spectral cues (e.g., Micheyl et al. 2007) and on ITD cues (Schadwinkel and Gutschalk 2010a). The latter study localized streaming of complex harmonic tones by ITD cues to the anterolateral part of Heschl's gyrus and provided evidence for a similar time window and brain activation for streaming based on ITD and on differences in fundamental frequencies. The supratemporal plane is also involved in spatial, ITD-related streaming of complex harmonic tones, and two distinct activation patterns have been described (Schadwinkel and Gutschalk 2010b; Schadwinkel and Gutschalk 2011). Transient activity is associated with the perception of streams, that is, with the reversal between a one- and two-stream percept;

sustained activity in Heschl's gyrus and in the planum temporale was shown to be parametrically modulated by the magnitude of the ITD differences. The authors interpreted their finding as evidence of the selective adaptation to the ITD cues underlying streaming. This specific involvement of the planum temporale in streaming based on ITD cues is consistent with its sensitivity to spatial positions as simulated by ITDs. As shown in one of the earlier studies, noise bursts lateralized to right or left hemispace yield stronger activation in the contralateral planum temporale than the same sounds presented centrally (Krumbholz et al. 2005).

The time course of sound source segregation and sound localization has been investigated with EEG and MEG. The streaming of harmonic tone complexes based on ITD has been shown to correlate with enhancement of the P1m at 60-90 ms post-stimulus onset; dipole fitting attributed this difference to the anterolateral part of Heschl's gyrus (Schadwinkel and Gotschalk 2010a). Another series of studies investigated the time course of sound source segregation and sound localization using the dichotic pitch paradigm. This paradigm consists of a narrow frequency band that is lateralized using ITD, while a simultaneously presented broadband noise is not lateralized. At 150-200 ms after stimulus onset, a negative deflection is observed during the processing of ITD in dichotic pitch discrimination (Johnson et al. 2003; Alain et al. 2001b; Alain et al. 2002). The location of the dichotic pitch, in left versus right hemispace, modulates the N2 response that occurs 250-350 ms after stimulus onset (Johnson et al. 2007), and selective attention modulates the neural activity at 400-500 ms (Hautus and Johnson 2005). Taken together, these results indicate that ITD cues are processed not only after information is passed to the dorsal stream but also affect streaming at various earlier stages (at 60-90 ms and at 150-250 ms) that are concurrent with the earliest stages of cortical ITD processing (e.g., Spierer et al. 2008; Tardif et al. 2006),

and even partially precede the differential processing of pitch vs ITD cues (De Santis et al. 2007).

The supratemporal plane has also been shown to contribute to stream segregation in complex auditory surroundings. An fMRI study investigated the ability of listeners to localize environmental sounds (e.g., a dog barking) in a complex setting simulated head-related transfer functions (Zündorf et al. 2013). Participants listened to 5 environmental sounds presented at different positions, with four simultaneous distractors. To assess the effort of sound localization in a complex auditory scene, the authors compared the active localization of a target among distractors to passive exposure to the same stimuli; the resulting activation involved the planum temporale, the anterior insula, the supplementary motor area, the left inferior frontal gyrus, and the right frontal eye field. To visualize the regions involved in separating the sound streams and in extracting the one of interest, the authors compared the localization of a target among distractors to sequential presentation of the targets; the resulting activation involved a small part of the planum temporale and the precuneus. This study and the previous ones highlight the role of the planum temporale and the precuneus in stream segregation. It should be noted, however, that several of the studies had acquired imaging data only of the supratemporal plane and the immediately adjacent region (Schadwinkel and Gutschalk 2010a; Schadwinkel and Gutschalk 2010b; Schadwinkel and Gutschalk 2011); hence, an involvement of other regions, such as the intraparietal sulcus, in streaming by means of spatial cues cannot be excluded. The intraparietal sulcus has been identified as a key structure for streaming, albeit by means of spectral, non-spatial cues (Cusack 2005). Thus, in addition to an early combination of spectro-temporal and spatial cues in the supratemporal plane, an interaction at later stages of the auditory pathways cannot be excluded.

In summary, behavioral and electrophysiological findings underscore the importance of spatial, including binaural, cues for auditory stream segregation. Moreover, existing

evidence points to a possible early combination of spectro-temporal and binaural cues in auditory processing.

5. *Functional dissociation between explicit and implicit processing of spatial cues – evidence from patient studies*

Paradigms investigating the use of explicit processing of spatial cues in the dorsal stream require listeners to localize a sound or to discriminate the positions of two successive sounds, as demonstrated in pioneering studies using free-field presentations (Makous and Middlebrooks 1990; Mills 1958) or in later clinical studies using binaural cues (e. g. Clarke et al. 2002). In contrast, implicit processing can be tested by spatial release from masking (SRM). This paradigm evaluates the intelligibility of speech, sound detection, or other psychoacoustic measures in conditions in which the spatial separation of the target and the distracters is modulated. It has been used for speech with spatial information provided in free field condition or simulated by head-related impulse responses or by binaural cues (Hawley et al. 1999; Hawley et al. 2004; Culling et al. 2004); for music in free field condition (Saupe et al. 2010); as well as for environmental sounds using ITD cues (Thiran and Clarke 2003). If explicit and implicit processing of spatial cues rely on the same substrate within the dorsal stream, both aspects should be similarly affected by brain lesions. However, several patient studies have reported differential effects on the explicit and implicit processing of spatial cues, suggesting a functional dissociation.

Patients with lesions centered on the parietal cortex often report striking difficulties in their everyday lives. Crossing a street can be a challenge for them. They are unable to track a vehicle using only auditory cues, which forces them to compensate by visually checking (Thiran and Clarke 2003; Clarke and At 2013). In a group of unknown people, these patients are unable to determine where a voice comes from by audition only; instead, they have to look

for moving lips. Thus, patients with parietal lesions are impaired in explicit spatial sound localization. Despite these difficulties, some patients are able to use implicitly ITD cues for auditory stream segregation. This was the case in two patients, one with a lesion of the right inferior colliculus (Litovsky et al. 2002) and the other with a large right hemispheric stroke (Thiran and Clarke 2003). The latter was examined 10 years after the insult and presented with spatial deafness, i.e., a complete inability to localize sounds or to compare their positions, either on the basis of binaural cues or in free-field conditions); despite this profound deficit, she was able to use binaural cues implicitly for the parsing of auditory scenes, as demonstrated in two types of tasks. She displayed SRM similar to healthy listeners. That is, a sound that was made unrecognizable by a simultaneous masking sound became intelligible with increasing binaural cue difference between the target and the masker. Moreover, she displayed auditory hemineglect. That is, in a test of diotic listening (i.e., two words were presented simultaneously with different ITDs such that one was lateralized to the right and the other to the left), she presented a left-sided extinction, indicating that the two words were segregated before one of them was neglected. Both, the SRM and the auditory hemineglect observed in that patient indicate that the implicit processing of binaural cues can be preserved while their explicit processing is impaired. A similarly preserved implicit use of binaural cues in diotic listening in conjunction with impairment of explicit sound localization was found in three other patients (Spierer et al. 2007). The patient studies described above illustrate the frequent dissociation between impaired explicit and preserved implicit use of binaural cues for the parsing of an auditory scene.

Explicit auditory spatial deficits can vary from mild to very severe. A mild deficit can refer to difficulties localizing sounds, while a severe deficit refers to an inability to distinguish between sounds coming from the left and the right (spatial deafness). These two extremes might reflect two stages of spatial processing: i) the precise computation of spatial

17

coordinates; and ii) the construction of a global auditory spatial representation, as suggested by electrophysiological, TMS and lesion studies using ITD cues (Lewald et al., 2002; Magezi and Krumbholz, 2010; Spierer et al. 2009). An analysis of the deficit profiles of a large group of patients has shown that all degrees of explicit processing deficits can coincide with either preserved or impaired implicit use of binaural cues (Duffour-Nikolov et al. 2012). This observation could indicate that there are three processing mechanisms: two explicit processing mechanisms for binaural cues and a third independent mechanism for the implicit processing of binaural cues.

It is worth mentioning that intact implicit processing together with impaired explicit processing was much more frequently observed in the study population than the opposite dissociation. Preserved sound localization in conjunction with disturbed implicit use of binaural cues was observed in only one patient (Duffour-Nikolov et al. 2012). However, this finding might be confounded by the patient selection. For practical reasons, patients with sound recognition deficits were not included in that study (to avoid false negatives in the spatial release from masking task). This may have led to the exclusion of patients who had, in addition to the sound recognition deficit, difficulties in using binaural cues implicitly for sound object segregation. It is likely that the two deficits, impaired spatial release from masking and impaired sound recognition, coincide in many patients, as both processes rely on the posterior superior temporal plane. Consequently, a sound identification deficit could frequently be associated with an impairment of sound segregation. This hypothesis is supported by a previous report that impairment of the semantic identification of sound objects is always associated with impairment of sound object categorization and/or sound object segregation based on non-spatial cues, whereas deficits in sound object segregation are not always associated with impairments in sound identification (Clarke et al. 1996). Thus, the infrequent observation of preserved explicit processing together with disturbed implicit

processing of binaural cues does not necessarily mean that this dissociation occurs less frequently.

The patient studies cited above clearly indicate a functional dissociation between the explicit and implicit processing of spatial cues. Patients frequently display impaired explicit sound localization and preserved implicit use of spatial cues, as shown by spatial release from masking or auditory hemineglect. This functional dissociation was already suggested in very early research studies with healthy listeners. One study indicated that binaural manipulations that do not produce clear lateralization improve speech intelligibility (Licklider, 1948). A series of studies then confirmed this finding using the inversion of speech waveforms, a manipulation that renders a sound un-localizable while simultaneously preserving the sound cues relevant for speech recognition. Interestingly, a speech sound that is masked by noise becomes intelligible when inversed (Carhart et al. 1969a; 1969b; Carhart et al., 1967; Carhart et al. 1968; Levitt and Rabiner 1967; Schubert 1956; Schubert and Schultz 1962). That is, a change in binaural characteristics affects auditory streaming even when the binaural stimulus characteristics do not allow the explicit localization of sounds. This finding had already indicated that the implicit processing of binaural cues does not depend on explicit processing or on a coherent spatial representation and supports the conclusion of a functional distinction between explicit and implicit spatial processing.

6. Concluding remarks

Taken together, the present body of experimental work indicates that auditory objects can be processed by distinct neural networks (Fig. 3). Identifying known sounds relies on processing within the ventral stream (Lewis et al. 2005; Murray et al. 2006; Altmann et al. 2007; Engel et al. 2009; Leaver and Rauschecker 2010; Lewis et al. 2011), which carries a position-independent representation of sound objects (Bourquin et al. 2013a). Localizing

objects engages parts of the dorsal stream within the parieto-frontal cortex (Zündorf et al.
2013). In addition to the position-independent representation of the meaning in the ventral
stream and to the spatial representation in the dorsal stream, specific neural populations
respond to sound objects as a function of their specific position. The current evidence for a
position-linked representation has been provided by studies that used spatially presented or
lateralized sound objects as stimuli, by auditory streaming based on ITD or effective location
cues, and by dissociation profiles in brain-damaged patients.

Neural populations that keep track of changes in the positions of individual sound
objects were identified on the left temporo-frontal convexity (Bourquin et al. 2013a); the two
regions involved, the posterior part of the superior and middle temporal gyri and the left
posterior inferior frontal cortex are known to be interconnected (Frey et al. 2008), and may
correspond to the early-stage and higher-order levels of the position-linked representation of
sound objects. The supratemporal plane has been shown to process binaural cues in the
context of auditory streaming (Schadwinkel and Gutschalk 2011), potentially as early as 60
ms after a sound is encountered (Johnson et al. 2007). The posterior inferior frontal cortex,
which has also been implicated in position-linked sound processing, is within a region known
to be selective for non-spatial processing (Maeder et al. 2001; Cohen et al. 2009) but receives
convergent input from the ventral and dorsal streams (Romanski et al. 1999a; Frey et al. 2014;
Frey et al. 2008). The contribution of the parietal cortex to position-linked representations of
sound objects cannot currently be excluded, but this nevertheless seems unlikely because
lesion studies have repeatedly shown that patients with parietal lesions tend to have sound
lateralization deficits (Spierer et al. 2009), often with preserved implicit use of binaural cues
for sound object segregation (Duffour-Nikolov et al. 2012). Activation studies of streaming
based on binaural cues suggest a possible role of the supratemporal plane in implicit
processing of spatial cues, but evidence from patient studies is currently not available.

References

Adriani, M., Bellmann, A., Meuli, R., Fornari, E., Frischknecht, R., Bindschaedler, C., Rivier, F., Thiran, J. P., Maeder, P., & Clarke, S. (2003). Unilateral hemispheric lesions disrupt parallel processing within the contralateral intact hemisphere: an auditory fMRI study. *Neuroimage, 20 Suppl 1*, S66-74. doi:S1053811903005196 [pii]

Ahveninen, J., Jaaskelainen, I. P., Raij, T., Bonmassar, G., Devore, S., Hamalainen, M., Levanen, S., Lin, F. H., Sams, M., Shinn-Cunningham, B. G., Witzel, T., & Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America, 103*(39), 14608-14613. doi:0510480103 [pii]

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001a). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences of the United States of America, 98*(21), 12301-12306. doi:10.1073/pnas.211209098 [doi]

Alain, C., Arnott, S. R., & Picton, T. W. (2001b). Bottom-up and top-down influences on auditory scene analysis: evidence from event-related brain potentials. *Journal of Experimental Psychology.Human Perception and Performance, 27*(5), 1072-1089.

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *The Journal of the Acoustical Society of America, 111*(2), 990-995.

Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex (New York, N.Y.: 1991), 17*(11), 2601-2608. doi:bhl167 [pii]

Altmann, C. F., Henning, M., Doring, M. K., & Kaiser, J. (2008). Effects of feature-selective attention on auditory pattern and location processing. *Neuroimage, 41*(1), 69-79. doi:10.1016/j.neuroimage.2008.02.013 [doi]

Anourova, I., Nikouline, V. V., Ilmoniemi, R. J., Hotta, J., Aronen, H. J., & Carlson, S. (2001). Evidence for dissociation of spatial and nonspatial auditory information processing. *Neuroimage, 14*(6), 1268-1277. doi:10.1006/nimg.2001.0903 [doi]

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage, 22*(1), 401-408. doi:10.1016/j.neuroimage.2004.01.014 [doi]

At, A., Spierer, L., & Clarke, S. (2011). The role of the right parietal cortex in sound localization: a chronometric single pulse transcranial magnetic stimulation study. *Neuropsychologia, 49*(9), 2794-2797. doi:10.1016/j.neuropsychologia.2011.05.024 [doi]

Aziz-Zadeh, L., Iacoboni, M., Zaidel, E., Wilson, S., Mazziotta, J. (2004). Left hemisphere motor facilitation in response to manual action sounds. Eur J Neurosci 19, 2609–2612. doi:10.1111/j.0953-816X.2004.03348.x

Bellmann, A., Meuli, R., & Clarke, S. (2001). Two types of auditory neglect. *Brain : A Journal of Neurology, 124*(Pt 4), 676-687.

Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. (2004). Neural correlates of auditory repetition priming: reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience, 16*(6), 966-977. doi:10.1162/0898929041502760 [doi]

Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology : JARO, 8*(2), 294-304. doi:10.1007/s10162-007-0073-z [doi]

Best, V., Ozmeral, E. J., Kopco, N., & Shinn-Cunningham, B. G. (2008). Object continuity enhances selective auditory attention. *Proceedings of the National Academy of Sciences of the United States of America, 105*(35), 13174-13178. doi:10.1073/pnas.0803718105 [doi]

Best, V., Shinn-Cunningham, B. G., Ozmeral, E. J., & Kopco, N. (2010). Exploring the benefit of auditory spatial continuity. *The Journal of the Acoustical Society of America, 127*(6), EL258-64. doi:10.1121/1.3431093 [doi]

Billig, A.J., Davis, M.H., Deeks, J.M., Monstrey, J., Carlyon, R.P. (2013). Lexical influences on auditory streaming. Curr Biol 23, 1585–1589. doi:10.1016/j.cub.2013.06.042

Bourquin, N. M., Murray, M. M., & Clarke, S. (2013a). Location-independent and location-linked representations of sound objects. *Neuroimage, 73*, 40-49. doi:10.1016/j.neuroimage.2013.01.026 [doi]

Bourquin, N. M., Simonin, A., & Clarke, S. (2013b). Repetition-induced plasticity of motor representations of action sounds. *Brain Topography, 26*(1), 152-156. doi:10.1007/s10548-012-0260-z [doi]

Bregman, A. S. (1990). *Auditory scene analysis*. Cambridge, MA: MIT Press.

Budd, T. W., Hall, D. A., Goncalves, M. S., Akeroyd, M. A., Foster, J. R., Palmer, A. R., Head, K., & Summerfield, A. Q. (2003). Binaural specialisation in human auditory cortex: an fMRI investigation of interaural correlation sensitivity. *Neuroimage, 20*(3), 1783-1794. doi:S1053811903004658 [pii]

Cammoun, L., Thiran, J. P., Griffa, A., Meuli, R., Hagmann, P., & Clarke, S. (2014). Intrahemispheric cortico-cortical connections of the human auditory cortex. *Brain Structure & Function,* doi:10.1007/s00429-014-0872-z [doi]

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969a). Perceptual masking in multiple sound backgrounds. *The Journal of the Acoustical Society of America, 45*(3), 694-703.

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969b). Release from multiple maskers: effects of interaural time disparities. *The Journal of the Acoustical Society of America, 45*(2), 411-418.

Carhart, R., Tillman, T. W., & Johnson, K. R. (1967). Release of masking for speech through interaural time delay. *The Journal of the Acoustical Society of America, 42*(1), 124-138.

Carhart, R., Tillman, T. W., & Johnson, K. R. (1968). Effects of interaural time delays on masking by two competing signals. *The Journal of the Acoustical Society of America, 43*(6), 1223-1230.

Carlyon, R. P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8*(10), 465-471. doi:10.1016/j.tics.2004.08.008 [doi]

Clarke, S., & At, A. (2013). Sound localization disorders and auditory neglect. In G. G. Celesia (Ed.), *Disorders of peripheral and central auditory processing. Handbook of clinical neurophysiology (Book 10)* (1st ed., pp. 423-434) Elsevier.

Clarke, S., & Morosan, P. (2013). Architecture, connectivity, and transmitter receptors of human auditory cortex. In D. Poeppel, & T. Overath (Eds.), *The human auditory cortex* (1.th ed., pp. 11-38). New York: Springer Science+Business Media.

Clarke, S., Bellmann Thiran, A., Maeder, P., Adriani, M., Vernet, O., Regli, L., Cuisenaire, O., & Thiran, J. P. (2002). What and where in human audition: selective deficits following focal hemispheric lesions. *Experimental Brain Research, 147*(1), 8-15. doi:10.1007/s00221-002-1203-9 [doi]

Clarke, S., Bellmann, A., De Ribaupierre, F., & Assal, G. (1996). Non-verbal auditory recognition in normal subjects and brain-damaged patients: evidence for parallel processing. *Neuropsychologia, 34*(6), 587-603. doi:0028-3932(95)00142-5 [pii]

Clarke, S., Bellmann, A., Meuli, R. A., Assal, G., & Steck, A. J. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia, 38*(6), 797-807. doi:S0028-3932(99)00141-4 [pii]

Cohen, Y.E., Russ, B.E., Davis, S.J., Baker, A.E., Ackelson, A.L., Nitecki, R. (2009). A functional role for the ventrolateral prefrontal cortex in non-spatial auditory cognition. Proc Natl Acad Sci U S A 106, 20045–20050. doi:10.1073/pnas.0907248106

Culling, J. F., Hawley, M. L., & Litovsky, R. Y. (2004). The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources. *The Journal of the Acoustical Society of America, 116*(2), 1057-1065.

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience, 17*(4), 641-651. doi:10.1162/0898929053467541 [doi]

Da Costa, S., Bourquin, N.M.-P., Knebel, J.-F., Saenz, M., van der Zwaag, W., Clarke, S., (2015) Representation of sound objects within early-stage auditory areas: a repetition effect study using 7T fMRI. PLOS ONE, in press.

Darwin, C. J., & Hukin, R. W. (1999). Auditory objects of attention: the role of interaural time differences. Journal of Experimental Psychology.Human Perception and Performance, 25(3), 617-629.

De Lucia, M., Camen, C., Clarke, S., & Murray, M. M. (2009). The role of actions in auditory object discrimination. *Neuroimage, 48*(2), 475-485. doi:10.1016/j.neuroimage.2009.06.041 [doi]

De Lucia, M., Cocchi, L., Martuzzi, R., Meuli, R. A., Clarke, S., & Murray, M. M. (2010). Perceptual and semantic contributions to repetition priming of environmental sounds. *Cerebral Cortex (New York, N.Y.: 1991), 20*(7), 1676-1684. doi:10.1093/cercor/bhp230 [doi]

De Meo, R, Bourquin, NM, Knebel, JF, Murray, MM, Clarke, S (2015) From bird to sparrow: learning-induced modulations in fine-grained semantic discrimination. Neuroimage, in press.

Deouell, L.Y., Heller, A.S., Malach, R., D'Esposito, M., Knight, R.T., 2007. Cerebral responses to change in spatial location of unattended sounds. Neuron 55, 985–996. doi:10.1016/j.neuron.2007.08.019

De Santis, L., Clarke, S., & Murray, M. M. (2007). Automatic and intrinsic auditory "what" and "where" processing in humans revealed by electrical neuroimaging. *Cerebral Cortex (New York, N.Y.: 1991), 17*(1), 9-17. doi:bhj119 [pii]

De Santis, L., Spierer, L., Clarke, S., & Murray, M. M. (2007). Getting in touch: segregated somatosensory what and where pathways in humans revealed by electrical neuroimaging. *Neuroimage, 37*(3), 890-903. doi:S1053-8119(07)00486-7 [pii]

Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia, 46*(11), 2776-2786. doi:10.1016/j.neuropsychologia.2008.05.011 [doi]

Drennan, W. R., Gatehouse, S., & Lever, C. (2003). Perceptual segregation of competing speech sounds: the role of spatial location. *The Journal of the Acoustical Society of America, 114*(4 Pt 1), 2178-2189.

Ducommun, C. Y., Michel, C. M., Clarke, S., Adriani, M., Seeck, M., Landis, T., & Blanke, O. (2004). Cortical motion deafness. *Neuron, 43*(6), 765-777. doi:10.1016/j.neuron.2004.08.020 [doi]

Ducommun, C. Y., Murray, M. M., Thut, G., Bellmann, A., Viaud-Delmon, I., Clarke, S., & Michel, C. M. (2002). Segregated processing of auditory motion and auditory location: an ERP mapping study. *Neuroimage, 16*(1), 76-88. doi:10.1006/nimg.2002.1062 [doi]

Duffour-Nikolov, C., Tardif, E., Maeder, P., Thiran, A. B., Bloch, J., Frischknecht, R., & Clarke, S. (2012). Auditory spatial deficits following hemispheric lesions: dissociation of explicit and implicit processing. *Neuropsychological Rehabilitation, 22*(5), 674-696. doi:10.1080/09602011.2012.686818 [doi]

Engel, L. R., Frum, C., Puce, A., Walker, N. A., & Lewis, J. W. (2009). Different categories of living and non-living sound-sources activate distinct cortical networks. *Neuroimage, 47*(4), 1778-1791. doi:10.1016/j.neuroimage.2009.05.041 [doi]

Eramudugolla, R., McAnally, K. I., Martin, R. L., Irvine, D. R., & Mattingley, J. B. (2008). The role of spatial location in auditory search. *Hearing Research, 238*(1-2), 139-146. doi:S0378-5955(07)00245-6 [pii]

Frey, S., Campbell, J.S.W., Pike, G.B., Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. J Neurosci 28, 11435–11444. doi:10.1523/JNEUROSCI.2388-08.2008

Frey, S., Mackey, S., Petrides, M. (2014). Cortico-cortical connections of areas 44 and 45B in the macaque monkey. Brain Lang 131. doi:10.1016/j.bandl.2013.05.005

Grill-Spector, K., Golarai, G., Gabrieli, J., 2008. Developmental neuroimaging of the human ventral visual cortex. Trends Cogn Sci 12, 152–62.

Griffiths, T. D., Kumar, S., Warren, J. D., Stewart, L., Stephan, K. E., & Friston, K. J. (2007). Approaches to the cortical analysis of auditory objects. *Hearing Research, 229*(1-2), 46-53. doi:S0378-5955(07)00019-6 [pii]

Hall, D. A., Barrett, D. J., Akeroyd, M. A., & Summerfield, A. Q. (2005). Cortical representations of temporal structure in sound. *Journal of Neurophysiology, 94*(5), 3181-3191. doi:00271.2005 [pii]

Hart, H. C., Palmer, A. R., & Hall, D. A. (2004). Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. *Human Brain Mapping, 21*(3), 178-190. doi:10.1002/hbm.10156 [doi]

Hauk, O., Shtyrov, Y., Pulvermuller, F., 2006. The sound of actions as reflected by mismatch negativity: rapid activation of cortical sensory-motor networks by sounds associated with finger and tongue movements. Eur J Neurosci 23, 811–821. doi:10.1111/j.1460-9568.2006.04586.x

Hautus, M. J., & Johnson, B. W. (2005). Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *The Journal of the Acoustical Society of America, 117*(1), 275-280.

Hawley, M. L., Litovsky, R. Y., & Colburn, H. S. (1999). Speech intelligibility and localization in a multi-source environment. *The Journal of the Acoustical Society of America, 105*(6), 3436-3448.

Hawley, M. L., Litovsky, R. Y., & Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: effect of location and type of interferer. *The Journal of the Acoustical Society of America, 115*(2), 833-843.

Johnson, B. W., Hautus, M. J., Duff, D. J., & Clapp, W. C. (2007). Sequential processing of interaural timing differences for sound source segregation and spatial localization: evidence from event-related cortical potentials. *Psychophysiology, 44*(4), 541-551. doi:PSYP535 [pii]

Johnson, B. W., Hautus, M., & Clapp, W. C. (2003). Neural activity associated with binaural processes for the perceptual segregation of pitch. *Clinical Neurophysiology : Official Journal of the International Federation of Clinical Neurophysiology, 114*(12), 2245-2250. doi:S1388245703002475 [pii]

Kaas, J. H., & Hackett, T. A. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 11793-11799. doi:10.1073/pnas.97.22.11793 [doi]

Kaas, J. H., Hackett, T. A., & Tramo, M. J. (1999). Auditory processing in primate cerebral cortex. *Current Opinion in Neurobiology, 9*(2), 164-170. doi:S0959-4388(99)80022-1 [pii]

Krumbholz, K., Schonwiesner, M., von Cramon, D.Y., Rubsamen, R., Shah, N.J., Zilles, K., Fink, G.R. (2005). Representation of interaural temporal information from left and right auditory space in the human planum temporale and inferior parietal lobe. Cereb Cortex 15, 317–324. doi:10.1093/cercor/bhh133

Kumar, S., Stephan, K. E., Warren, J. D., Friston, K. J., & Griffiths, T. D. (2007). Hierarchical processing of auditory objects in humans. *PLoS Computational Biology, 3*(6), e100. doi:06-PLCB-RA-0480R3 [pii]

Lahav, A., Saltzman, E., Schlaug, G., 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci 27, 308–314. doi:10.1523/JNEUROSCI.4822-06.2007

Leaver, A. M., & Rauschecker, J. P. (2010). Cortical representation of natural complex sounds: effects of acoustic features and auditory object category. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 30*(22), 7604-7612. doi:10.1523/JNEUROSCI.0296-10.2010 [doi]

Levitt, H., & Rabiner, L. R. (1967). Predicting binaural gain in intelligibility and release from masking for speech. *The Journal of the Acoustical Society of America, 42*(4), 820-829.

Lewald, J., Foltys, H., & Topper, R. (2002). Role of the posterior parietal cortex in spatial hearing. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 22*(3), RC207. doi:20026072 [pii]

Lewald, J., Wienemann, M., & Boroojerdi, B. (2004). Shift in sound localization induced by rTMS of the posterior parietal lobe. *Neuropsychologia, 42*(12), 1598-1607. doi:10.1016/j.neuropsychologia.2004.04.012 [doi]

Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Janik, J. J., & DeYoe, E. A. (2005). Distinct cortical pathways for processing tool versus animal sounds. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 25*(21), 5148-5158. doi:25/21/5148 [pii]

Lewis, J. W., Talkington, W. J., Puce, A., Engel, L. R., & Frum, C. (2011). Cortical networks representing object categories and high-level attributes of familiar real-world action sounds. *Journal of Cognitive Neuroscience, 23*(8), 2079-2101. doi:10.1162/jocn.2010.21570 [doi]

Licklider, J. R. (1948). The influence of interaural phase relations upon the masking of sepech by white noise. *Journal of the ACoustical Society of America, 20*, 150-159.

Litovsky, R. Y., Fligor, B. J., & Tramo, M. J. (2002). Functional role of the human inferior colliculus in binaural hearing. *Hearing Research, 165*(1-2), 177-188. doi:S0378595502003040 [pii]

Livingstone, D. (1875). *A popular account of Dr. Livingstone's expedition to the Zambesi and its tributaries: and of the discovery of Lakes Shirwa and Nyassa, 1858-1864*. London: J. Murray.

Maeder, P.P., Meuli, R.A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J.P., Pittet, A., Clarke, S., 2001. Distinct pathways involved in sound recognition and localization: a human fMRI study. Neuroimage 14, 802–816. doi:10.1006/nimg.2001.0888

Magezi, D. A., & Krumbholz, K. (2010). Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *Journal of Neurophysiology, 104*(4), 1997-2007. doi:10.1152/jn.00424.2009 [doi]

Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America, 87*(5), 2188-2200.

McDermott, J. H., Wrobleski, D., & Oxenham, A. J. (2011). Recovering sound sources from embedded repetition. *Proceedings of the National Academy of Sciences of the United States of America, 108*(3), 1188-1193. doi:10.1073/pnas.1004765108 [doi]

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J.P., Tian, B., Courtenay Wilson, E. (2007). The role of auditory cortex in the formation of auditory streams. Hear Res 229, 116–131. doi:10.1016/j.heares.2007.01.007

Middlebrooks, J. C., & Onsan, Z. A. (2012). Stream segregation with high spatial acuity. *The Journal of the Acoustical Society of America, 132*(6), 3896-3911. doi:10.1121/1.4764879 [doi]

Mills, A. W. (1958). On the minimum audible angle. *Journal of the Acoustical Society of America, 30*, 237-246.

Moore, B. C., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society of London.Series B, Biological Sciences, 367*(1591), 919-931. doi:10.1098/rstb.2011.0355 [doi]

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 26*(4), 1293-1302. doi:26/4/1293 [pii]

Murray, M. M., Camen, C., Spierer, L., & Clarke, S. (2008). Plasticity in representations of environmental sounds revealed by electrical neuroimaging. *Neuroimage, 39*(2), 847-856. doi:S1053-8119(07)00799-9 [pii]

Petrides, M., Pandya, D.N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. PLoS Biol 7. doi:10.1371/journal.pbio.1000170

Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'Amico, S., Di Russo, F., 2005. Separate neural systems for processing action- or non-action-related sounds. Neuroimage 24, 852–861. doi:10.1016/j.neuroimage.2004.09.025

Rauschecker, J. P. (1998). Parallel processing in the auditory cortex of primates. *Audiology & Neuro-Otology, 3*(2-3), 86-103. doi:aud03086 [pii]

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nature Neuroscience, 12*(6), 718-724. doi:10.1038/nn.2331 [doi]

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 11800-11806. doi:10.1073/pnas.97.22.11800 [doi]

Rey B, Frischknecht R, Maeder P, Clarke S (2007) Patterns of recovery following focal hemispheric lesions: Relationship between lasting deficit and damage to specialized networks. Restorative Neurology and Neuroscience, 25: 285-294.

Rivier, F., & Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage, 6*(4), 288-304. doi:S1053-8119(97)90304-9 [pii]

Romanski, L. M., Tian, B., Fritz, J., Mishkin, M., Goldman-Rakic, P. S., & Rauschecker, J. P. (1999a). Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat.Neurosci., 2*(12), 1131-1136.

Romanski, L. M., Bates, J. F., & Goldman-Rakic, P. S. (1999b). Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. *The Journal of Comparative Neurology, 403*(2), 141-157. doi:10.1002/(SICI)1096-9861(19990111)403:2<141::AID-CNE1>3.0.CO;2-V [pii]

Saupe, K., Koelsch, S., & Rubsamen, R. (2010). Spatial selective attention in a complex auditory environment such as polyphonic music. *The Journal of the Acoustical Society of America, 127*(1), 472-480. doi:10.1121/1.3271422 [doi]

Schadwinkel, S., & Gutschalk, A. (2010a). Activity associated with stream segregation in human auditory cortex is similar for spatial and pitch cues. *Cerebral Cortex (New York, N.Y.: 1991), 20*(12), 2863-2873. doi:10.1093/cercor/bhq037 [doi]

Schadwinkel, S., & Gutschalk, A. (2010b). Functional dissociation of transient and sustained fMRI BOLD components in human auditory cortex revealed with a streaming paradigm based on interaural time differences. *The European Journal of Neuroscience, 32*(11), 1970-1978. doi:10.1111/j.1460-9568.2010.07459.x [doi]

Schadwinkel, S., & Gutschalk, A. (2011). Transient bold activity locked to perceptual reversals of auditory streaming in human auditory cortex and inferior colliculus. *Journal of Neurophysiology, 105*(5), 1977-1983. doi:10.1152/jn.00461.2010 [doi]

Schubert, D. (1956). Some preliminary experiments on binaural time delay and intelligibility. *Journal of the Acaoustical Society of America, 28*, 895-901.

Schubert, D., & Schultz, M. C. (1962). Some aspects of binaural signal selection. *Journal of the Acoustical Society of America, 34*, 844-849.

Schwartz, A., McDermott, J. H., & Shinn-Cunningham, B. (2012). Spatial cues alone produce inaccurate sound segregation: the effect of interaural time differences. *The Journal of the Acoustical Society of America, 132*(1), 357-368. doi:10.1121/1.4718637 [doi]

Shrem, T., & Deouell, L. Y. (2014). Frequency-dependent auditory space representation in the human planum temporale. *Frontiers in Human Neuroscience, 8*, 524. doi:10.3389/fnhum.2014.00524 [doi]

Spierer, L., Bellmann-Thiran, A., Maeder, P., Murray, M. M., & Clarke, S. (2009). Hemispheric competence for auditory spatial representation. *Brain 132*(7), 1953-1966. doi:10.1093/brain/awp127. Epub 2009 May 28.

Spierer, L., Meuli, R., & Clarke, S. (2007). Extinction of auditory stimuli in hemineglect: Space versus ear. *Neuropsychologia, 45*(3), 540-551. doi:S0028-3932(06)00153-9 [pii]

Spierer, L., Murray, M. M., Tardif, E., & Clarke, S. (2008). The path to success in auditory spatial discrimination: electrical neuroimaging responses within the supratemporal plane predict performance outcome.. *Neuroimage, 41*(2), 493-503. doi:0.1016/j.neuroimage.2008.02.038. Epub 2008 Mar 5.

Tardif, E., Murray, M. M., Meylan, R., Spierer, L., & Clarke, S. (2006). The spatio-temporal brain dynamics of processing and integrating sound localization cues in humans. *Brain Research, 1092*(1), 161-176. doi:S0006-8993(06)00849-3 [pii]

Tardif, E., Probst, A., Clarke, S. (2007). Laminar specificity of intrinsic connections in Broca's area. Cereb Cortex 17, 2949–2960. doi:10.1093/cercor/bhm021

Tardif, E., Spierer, L., Clarke, S., & Murray, M. M. (2008). Interactions between auditory 'what' and 'where' pathways revealed by enhanced near-threshold discrimination of frequency and position. *Neuropsychologia, 46*(4), 958-966. doi:10.1016/j.neuropsychologia.2007.11.016 [doi]

Thiran, A. B., & Clarke, S. (2003). Preserved use of spatial cues for sound segregation in a case of spatial deafness. *Neuropsychologia, 41*(9), 1254-1261. doi:S0028393203000149 [pii]

Tian, B., Reser, D., Durham, A., Kustov, A., & Rauschecker, J. P. (2001). Functional specialization in rhesus monkey auditory cortex. *Science (New York, N.Y.), 292*(5515), 290-293. doi:10.1126/science.1058911 [doi]

van der Zwaag, W., Gentile, G., Gruetter, R., Spierer, L., & Clarke, S. (2011). Where sound position influences sound object representations: a 7-T fMRI study. *Neuroimage, 54*(3), 1803-1811. doi:10.1016/j.neuroimage.2010.10.032 [doi]

Viceic, D., Fornari, E., Thiran, J. P., Maeder, P. P., Meuli, R., Adriani, M., & Clarke, S. (2006). Human auditory belt areas specialized in sound recognition: a functional magnetic resonance imaging study. *Neuroreport, 17*(16), 1659-1662. doi:10.1097/01.wnr.0000239962.75943.dd [doi]

Wallace, M. N., Johnston, P. W., & Palmer, A. R. (2002). Histochemical identification of cortical areas in the auditory region of the human brain. *Experimental Brain Research, 143*(4), 499-508. doi:10.1007/s00221-002-1014-z [doi]

Yost, W. A. (1991). Auditory image perception and analysis: the basis for hearing. *Hearing Research, 56*(1-2), 8-18.

Zundorf, I. C., Lewald, J., & Karnath, H. O. (2013). Neural correlates of sound localization in complex acoustic environments. *PloS One, 8*(5), e64259. doi:10.1371/journal.pone.0064259 [doi]
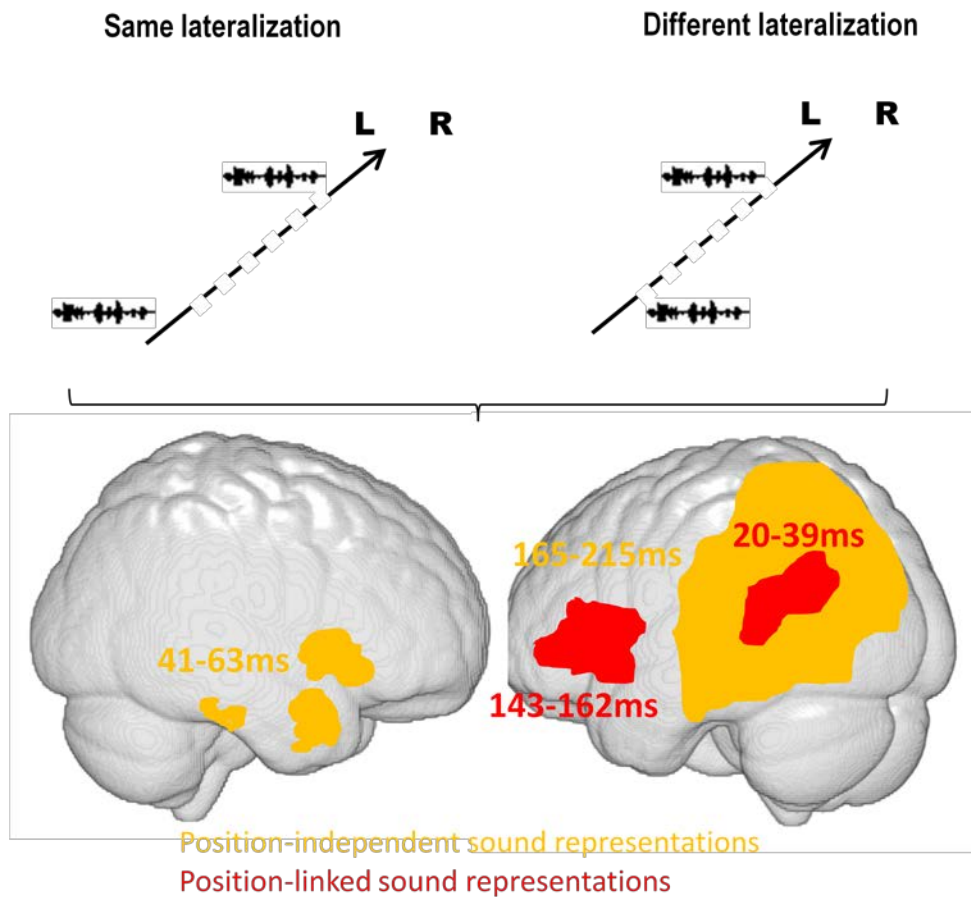
Figure 1. Schematic depiction of position-independent and position-linked representations of sound objects, as revealed by auditory evoked potentials in a repetition priming paradigm (adapted from the graphical abstract in Bourquin et al. (2013**a**). Top) Experimental conditions: Sounds of the same objects presented either with the same lateralization (left) or with different lateralization (right); L = left, R = right. Bottom) Position-linked (red) and position-independent (yellow) representations and their respective time windows post-stimulus onset. The former was revealed by differential repetition effects for stimulus pairs with the same vs different lateralization, the latter by repetition suppression which occurred independently of whether the lateralization of a given sound object was held constant or not.
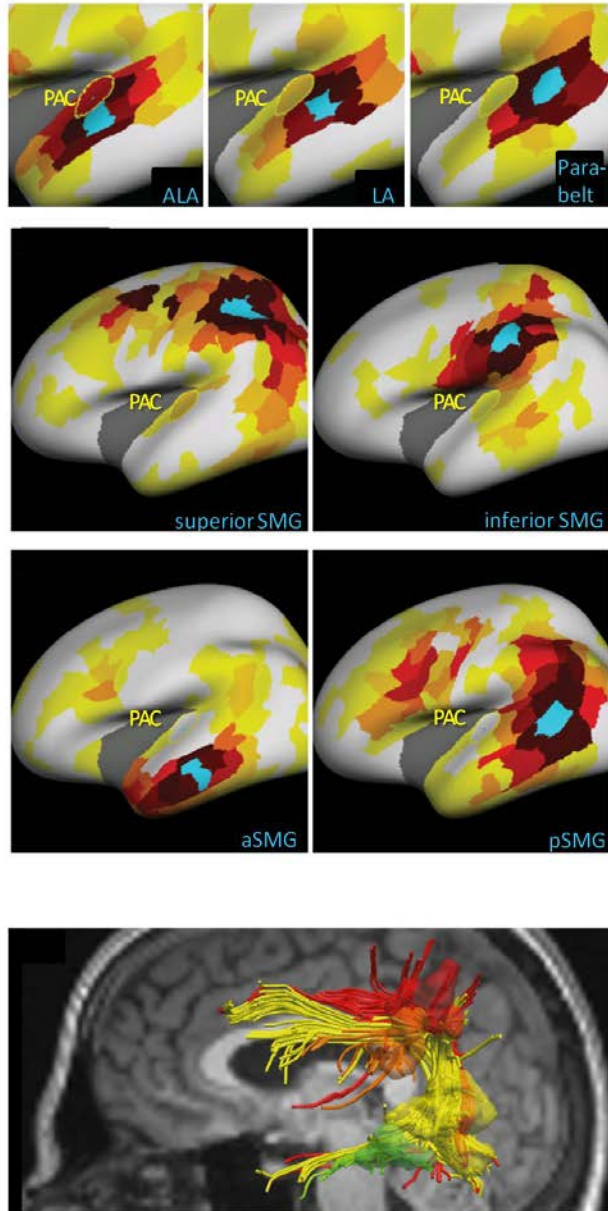
Figure 2. Cortico-cortical connections of the (left side) human auditory belt (ALA, LA) and parabelt areas (top panel) and of discrete regions of the middle temporal and supramarginal gyri (middle panel; MTG and SMG, respectively) as well as cortico-cortical fibers originating in the MTG and SMG regions (bottom panel), as revealed by in vivo DSI tractography. In the top and middle panels, cyan regions mark the region of origin of the connections shown; brown-to-yellow represents the strength of the connections; the primary auditory cortex (PAC) is outlined in yellow. In the bottom panel, red and orange show the connectivity of two regions on the supramarginal gyrus and green and yellow the connectivity of two regions on the middle temporal gyrus. The cortical representation is a latero-superior view of the partially inflated brain; individual areas have been defined by means of coordinates published previously (Rivier and Clarke 1997; Wallace et al. 2002). The PAC was considered to be co-extensive with the medial two-thirds of Heschl's gyrus, in agreement with previous 7T and 3T activation studies, which demonstrated a constant relationship between this part of Heschl's gyrus and the two main mirror-reversed tonotopic maps (Da Costa et al. 2011; Da Costa et al. 2013). Adapted from Cammoun et al. (2014).
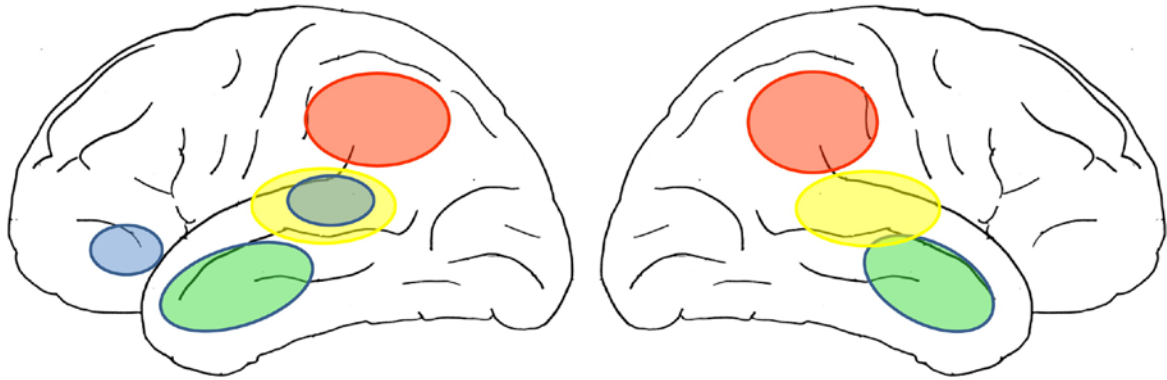
Figure 3. Schematic representation of the neural networks involved in auditory processing that selectively underlie sound recognition (green) and sound localization (red). Regions comprising neural populations for spatial or non-spatial processing in close proximity: yellow. A left hemispheric network supporting a position-linked representation of sound objects, involving the posterior part of the superior and middle temporal gyrus and the posterior part of the inferior frontal cortex: blue (Bourquin et al. 2013b).