

Auditory imagery: A review and challenges ahead

Mikael Jensen

SSKKII Center for Cognitive Science, Göteborg University, Sweden

This article is a review of the research on auditory imagery during the last three decades. The purpose is to present and summarize the central findings. The research field of auditory imagery is related to research on visual imagery and it is also very close to the field of auditory perception. Though a lot has been done since the 1970s it is still a fairly small field of research, which is a little strange, since both speech and music are central to most of our lives. Relevant research can be found in psychology, neuroscience and cognitive science, which means that today it is possible to get a relatively broad understanding of the phenomenon of auditory imagery. The findings reveal where in the brain auditory perception and auditory imagery probably are processed and show us what areas are common to these processes. They also tell us something about the character of auditory images, and their relationship to working memory. Three major fields will be presented in the review: (1) auditory imagery and environmental sounds, (2) musical imagery, and (3) inner speech (or speech imagery) and subvocalization. The review ends with a discussion and some suggestions for future research that needs to be done.

The research in perception has got an intimate partner in imagery studies (Denis, 2001). Most things that you can perceive you can also imagine. Just as you can see a cat with your eyes you are capable of seeing this or some other cat by visually imagining it. The perception of the cat's meow and the imagery of the same are both auditory even though there is no present cat in the latter case. So imagery is an experience of seeing, hearing, smelling, feeling and so forth without any sensory stimulus from the world. Then where does this experience come from?

The idea of imagery is that it relies on former experiences of the actual world. While remembering something, e.g. a visit on the beach, you can see the sand and the water for your "inner eyes", hear children laugh and the sound of the waves rolling in for your "inner ear"¹ and feel the hot, soft sand around your feet for your "inner skin". A great deal of this memory must have been stored somewhere in your brain when you were at the beach one particular or several times.

The obvious thing then is that visual perception is connected to visual imagery; auditory perception is connected to auditory imagery and so on. Via the perception the brain creates a representation of that special object that is perceived. That particular representation is called an image. You can have an image of the cat. You can have

¹ "inner ear" between quotation marks is distinct from the inner part of the ear, here meaning a mental process.

an image of the meow. You can have an image of touching the fur. Then the image is the thing stored in your brain and retrieved when recalling it or imagining it.

Because this text is a review of research on *auditory imagery*, and the purpose is to present the findings of the last three decades, we first need to take a look at a definition of the central concept. For a more general definition see Denis (2001).

[A]uditory imagery is the introspection persistence of an auditory experience, including one constructed from components drawn from long-term memory, in the absence of direct sensory instigation of that experience (Intons-Peterson, 1992, 46).

The well-known composer Beethoven was deaf the last years of his life. His auditory sensors in the ear gave him no input to the brain. Consequently he could not hear but he went on composing guided by his “inner ear”. He described how he could carry a tune in his mind for a very long time before he wrote it down (Deutsch & Pierce, 1992). This is one example of how imagery can be used. According to the definition above Beethoven could not have been such an exceptional composer, probably not composer at all, if he had been born deaf.

The research on imagery also includes the characteristics of the sounds. They are: pitch, loudness, timbre, temporal extent and tempo (pulse). Maybe rhythm, metric and sequence should be counted in as well but then we talk about more complex sound patterns and special cases. All of these characteristics will be investigated further later on.

Imagery related to different sounds will be presented in three categories: object and environmental sounds (e.g. a doorbell or waves), music (e.g. a melody), and verbal sounds (e.g. speech). These are sounds that constantly surround us. In some way they must influence our understanding of everyday life. But before we examine imagery and different sounds the research on imagery will be put in a historical context. Thereafter perception and imagery need to be understood from the view of neuroscience. Finally, after examining the imagery research, we will take a look at future investigations. What is still to be done?

A BRIEF OVERVIEW

While most researchers, interested in mental imagery, studied and still study visual imagery just a few try to solve the problems in motor-, auditory- and olfactory imagery. Only about 5 percent of the articles published in some of the largest databases, on mental imagery, concern auditory imagery (June 2004). Some more are about motor imagery and a lot less are about olfactory imagery. The rest of the articles, approximately two thirds, deal with visual imagery.

In the 1970s when Shepard and Metzler (1971) studied mental rotation and Kosslyn (1973) studied visual scanning almost nothing was done in auditory imagery research. You could find some article on memory for pitch (see Deutsch, 1972). At the end of the decade Intons-Peterson (1980) worked on loudness in auditory imagery and Geiselman with collaborators (1977; 1980) did some voice imagery.

Not until the 1980s did the research on auditory imagery get started in a broad sense. Some studied speech imagery or inner speech (e.g. MacKay, 1981; Anderson, 1982) as well as subvocalization (e.g. Reisberg and collaborators). Others studied musical imagery, especially pitch (e.g. Weber & Brown, 1986; Halpern, 1989), temporal extent (Halpern, 1988a), tempo (Halpern, 1988b) and timbre (Crowder, 1989; also Pitt & Crowder, 1992).

With this broader research as a background Reisberg (1992) edited a book named *Auditory imagery* to give a review and inspire others. Smith, Reisberg and Wilson developed a theory that subvocalization has a role in more than just working memory (e.g. Smith, Reisberg & Wilson, 1992; Smith, Wilson & Reisberg 1995; see also Baddeley & Logie, 1992). Halpern found a partner in Zatorre who is specialised in music perception and they carried out a series of experiments (1993; 1996; 1999). Two research-groups (Rao et al., 1997; Penhune et al., 1998) looked at imagined pulse (tempo) and in late 1990s Hoshiyama, Gunji and Kakigi (2001) found interest in object-sounds and imagery.

It is obvious that a new generation of scientists are trying to solve some of the problems connected with auditory imagery and that the result is being available now, in the first decade of the 2000s. Since the 1990s almost every report includes neuro-scanning methods, which has improved the chances of making good estimates. Now in the 2000s we have a pretty good picture of where in the brain the processes of auditory perception and imagery are located. The next step therefore is to look inside the auditory cortex and other relevant neural structures.

AUDITORY CORTEX²

For a long time the auditory system has been neglected in neuroscience and cognitive science (Zatorre, 2003a). It is in the last decade or so that we have seen real results. The reason for this is not just lack of interest. There have been problems with the scanning technique. PET and especially fMRI make too much noise. When you examine the auditory system you can't have unwanted sounds in the experiment. The results would not be reliable. But now the problems are almost solved which means that we can see a fast growing amount of auditory knowledge (Di Salle et al., 2003).

The auditory cortex is located in the superior part of the temporal lobe. It's normally just inside and a little above, slightly behind the ear. The connection between the ear sensors and the auditory cortex is not direct. The auditory nerve leads the impulse to the cochlear nucleus. The left ear is only connected to the left cochlear nu-

² The purpose of this section is not to give a full overview of neuroscience concerning auditory cortex. Other articles have done that. Important, though, is to mention something about the methods used. The most common method is called *subtraction*. In this case the subject is scanned while hearing background sounds or preferably no sound at all. Then the subject is scanned again while listening to a target sound. The brain activity during the first condition is subtracted from the brain activity during the second condition. This might say something about which parts of the brain are processing the information from the target condition. The data must be carefully used and can not be interpreted as some unique conclusion.

cleus and the same goes for the right side. From thereon the auditory information is processed bilaterally. This means that the superior olives (two structures that are direction sensitive) are fed information from both ears. The superior olives send the information to the inferior colliculus passing it further to thalamus and MGB (medial geniculate body). From there the signal reaches the auditory cortex (e.g. Scott & Wise, 2004; Thompson, 2000).

Two-thirds of the primary auditory cortex, PAC, consists of a region called Heschl's gyrus (see fig. 1), a transverse gyrus from the upper side of STG (superior temporal gyrus) into the Sylvian fissure (the lateral fissure) (Zatorre, Belin & Penhune, 2002a; Artacho-Pérula et al., 2004). The PAC is partly tonotopic; i.e. every nerve-end in the ear has a corresponding region in the PAC (Morosan et al., 2001).

Already in the PAC there is a dual pathway. Two subsystems, AI and R (or RL), are specialised in different frequencies. AI is more sensitive to the higher register. R is more sensitive to the lower register. This means that other structures connected to AI, e.g. A4 (or CL), process auditory spatial relations or motion while structures connected to R, e.g. A4 (or AL), process voices or voice-like sounds. These two pathways starting in AI and R are also called the *where* respectively the *what* pathway/stream/system (Rauschecker & Tian, 2003; see also Scott & Wise, 2003; Hickok & Poeppel, 2004; Arnott et al., 2004; Zatorre et al., 2002b).

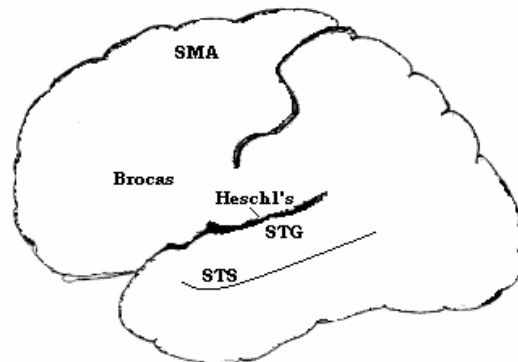


Figure 1. Showing Heschl's gyrus, STG and STS

There seems to be a hemispheric specialisation in certain areas; e.g. the left side is specialised on speech and voices and the right side is specialised on singing and melodies (Scott & Wise, 2004; 2003; Belin, Zatorre & Ahad, 2002; Di Salle et al., 2003; Patel & Balaban, 2001). STG and STS (superior temporal sulcus) in the right side are more sensitive to differences in frequency (pitch) (Zatorre et al., 2002a; Hart, Palmer & Hall, 2002). In planum polare (anterior to Heschl's gyrus), BA 22 and 38, there seems to be a fine-tuning process of frequency, clearly relevant to musical information processing (Brown et al., 2004; Platel et al., 1997). The right anterior STS are active while identifying object sounds (Zatorre, Bouffard, & Belin, 2004). STG in the right hemisphere more often process temporal information (Zatorre et al., 2002a; Scott and Wise, 2003).

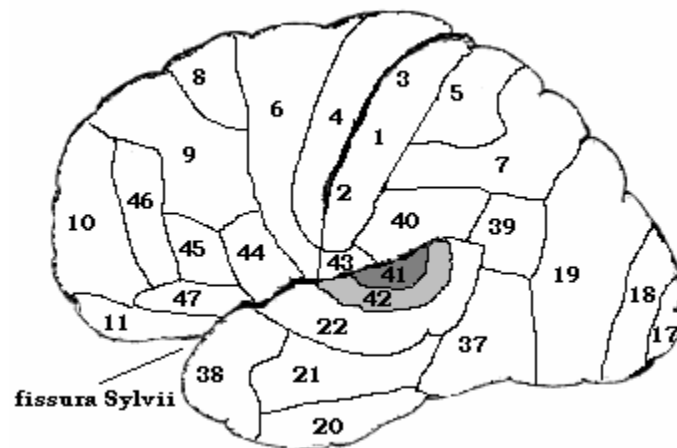


Figure 2. The left side of the brain divided into Brodmann areas

Surrounding the PAC is secondary auditory cortex, SAC, and associative auditory areas. They are localised in Brodmann's area (BA) 41, 42 and 22 (See fig. 2; Husain et al., 2004). According to Penhune, Zatorre & Feindel (1999) rhythm is processed in the anterior part of the SAC, dominantly in the right hemisphere (Liegeois-Chauvel et al., 1998, found a dissociation between rhythm and metre in anterior STG). Sequence information, e.g. links of syllables or notes, is processed in both Broca's area (BA 44, 45 and 6) and in the insula, indicating that it is important for speech, singing and articulation (Tillman, Janata & Bharucha, 2004; Platel et al., 1997). Pulse or tempo is processed in STG, especially PT (BA 22), but the left side is only active in specific conditions (Rao et al., 1997; Penhune, Zatorre & Evans, 1998; Husain et al., 2004; see also Lewis & Miall, 2003, about pulse with visual origin activating STG).

Heschl's gyrus is the most loudness-sensitive area. It is always active from the lowest dB to the highest. Some neighbouring areas (e.g. ALA [anterior lateral area] and PT [planum temporale]) only react when the dB gets really high (Hart et al., 2002).

Finally timbre was assumed to be processed in the right hemisphere (e.g. Platel et al., 1997). Later, two independent studies (Samson, Zatorre & Ramsay, 2002; Menon et al., 2002) showed that timbre is processed bilaterally. Samson et al. had participants with right temporal lobe lesion which revealed disturbed perception in the timbre task. Thus, the right side is needed but a contribution of the left temporal regions is also apparent. Menon et al. found no significant difference between the activity in the left and right hemisphere but found more posterior processing in the left temporal lobe. Halpern et al. (2004) found a somewhat higher activation in the right temporal lobe, especially in PT and posterior STG.

OBJECT AND ENVIRONMENTAL SOUNDS

The first person to examine object or environmental sounds (e.g. trains, wind chimes, popcorn popping) in imagery was Intons-Peterson (1980; also 1992). She wanted to

know if images of object sound encoded loudness. Her experiments gave different results indicating that auditory images contain loudness information but it is not retrieved all the time. In one test the participants were asked to compare two specific, imagined sounds and identify the louder (or softer). As predicted, the judgement time decreased with increasing distance between the loudness ratings. The small differences made it harder to discriminate one sound from the other.

Intons-Peterson, Russel, and Dressel (1992) carried out a study influenced by the one above but this time focused on pitch. Just as in the case of loudness fundamental pitch does not seem to be obligatory in environmental sounds. However, it was easier, i.e. a shorter time was needed, to discriminate two sounds when the pitch differed greatly.

Hoshiyama et al. (2001) tried a cross-modal experiment where the participants first saw (on a video monitor) a hammer strike an anvil but heard no sound. During this passive task their brains were scanned. The second, an active task, was to imagine the sound that would correspond to the collision (still hearing no external sound). The scanning result showed that the imagery task activated STG and inferior frontal regions (in some cases the insula was activated). Neither of these areas was active in the passive task.

MUSICAL IMAGERY

This research field is presented under the captions: pitch, timbre, temporal extent, tempo and neural activation in music imagery.

Pitch

Very few have the ability of absolute pitch, meaning that they can tell the name of the note (e.g. A or C) if they hear a tone. It seems that this talent is a combination of genetic factors and musical training in the childhood (Zatorre, 2003b). Though most of us lack absolute pitch we have a relative, good memory for pitch. Levitin (1994) suggest that we might possess stable representations of pitch, an accurate pitch memory.

Halpern (1989; also 1992) gave her experiment participants the task to imagine the first tone of a particular familiar song. When they had it they were supposed to hum the tone. This was recorded and the tone identified. 48 hours later the test was repeated. The result gave a strong indication that pitch representation is stored, which Levitin later showed more clearly.

In another test the participants used a keyboard (it was placed upside down and hidden from view) to find the starting note. This test was more difficult for the participants (most of them were unfamiliar with the use of keyboards) but the result pointed in the same direction as the first test. Even more tests by Halpern gave a unanimous result. We obviously have stable memory for pitch and it is as clear in perception and singing out loud as in imagery and silent singing.

One interesting result from Halpern's experiments was that musicians did not have an advantage over non-musicians. With the exception of musicians with absolute pitch, musicians have the same memory for pitch as others.

Another way to study pitch is to hear a tone and then imagine it one interval higher (e.g. from C to D). Hubbard and Stoeckig (1992) hypothesised that the participants would react faster if they heard the same note as the one they imagined and slower if it was somewhat higher or lower. According to Hubbard and Stoeckig humans are able to change pitch without the need of perception or the voice.

Timbre

Crowder (1989; see also Crowder & Pitt, 1992) wanted to find out if timbre is an essential part of musical representations and musical imagery. To do this he asked the test subjects to push a button if the second tone they heard was the same as the one they previously heard. It could either be an F, G or an A. The tone was presented by a flute, trumpet or guitar. So there was a 3 x 3-tone arrangement. If the same note was played twice in a row, reaction time was much faster if the same instrument (e.g. two trumpet-blows in G) produced it. This should mean that our memory carries information about timbre as well as pitch.

In the second part of the experiment, the participants heard one of the three notes as above but as a sine wave. They were asked to imagine this note played by one of the three instruments. Thereafter they heard one of the three instruments playing any of the three notes. If the note in the first case was the same as in the latter they should push the button for "same". The result was similar as above. The same note played by the same instrument gave a faster reaction. Just the sine wave would give no difference. The participants clearly could imagine and had representations of timbre (see also Pitt, 1995).

This experiment is discussed in Smith, Reisberg, and Wilson (1992) to Crowder's advantage. The differences in reaction time disappear if the "inner ear" is blocked; i.e. if the subject perceives irrelevant speech while trying to imagine timbre. That makes the task much harder to perform.

Is timbre easier to handle for musicians? Pitt (1994) showed in his study that non-musicians had some difficulties judging timbre when pitch was varied but even more problems judging pitch when timbre was varied. This was not observed in musicians. Is it possible to find different patterns in neural processing between musicians and non-musicians? As in the case of absolute pitch, training display a different neural activity pattern (Zatorre, 2003b). With this in mind Halpern et al. (2004) PET-experiment show only the neural activity of musicians. During timbre imagery the right posterior STG, left PT, right TPO (temporo-parieto-occipital) and SMA (supplementary motor area, on the top of the brain) was activated. Posterior STG and PT are part of the posterior SAC.

Temporal extent

Halpern (1988a) gave her participants a song title (e.g. Jingle Bells) and the first line of the song. They were asked to say if the word *snow* was higher or lower in pitch than the word *sleigh* (Dashing through the SNOW, in a one-horse open SLEIGH). This means that the participants had to scan through the melody from the beginning to the point where they could answer the question. How long time will it take?

People indeed behave as if they were running songs through their heads. That is, the evidence seems to point toward a representation that codes extension in time that unfolds in real time, that has strong links between adjacent elements, and that is unidirectionally ordered (Halpern, 1992, 10).

So it seems that imagery scanning is performed in real time (i.e. the same as perception) and that the representations have temporal-like characteristics, temporal extent. The next question is where in the brain this type of task is processed. Zatorre and Halpern (1993) found that participants with right temporal lesions had bigger problems in both the perception task and the imagery task (constructed as above) than participants with left temporal lesions and the control group had (see also Halpern, 2001).

In a later PET-study Zatorre et al. (1996) discovered activation in right anterior and posterior STG, in left posterior STG, in mid frontal regions bilaterally, left anterior inferior frontal regions and SMA bilaterally to mention most of them.

Aleman et al. (2000) repeated Halpern's experiments and their study showed that the musicians had better results on both the temporal test and on a non-music related but auditory imagery test. On the visual imagery test there was no significant difference between musicians and non-musicians. Aleman et al. suggest that musicians have better auditory imagery abilities.

Tempo

Music normally has a tempo (or pulse). In one imagery experiment Halpern (1988b) wanted to know if tempo is represented in a similar way as pitch and temporal extent. The subjects sang familiar songs silently in their heads simultaneous to their own finger tapping. This was recorded and the tempo was determined. The procedure was repeated twice. After five days there was another session.

The musicians in the test kept the same tempo from try to try and between sessions, singing the same song. They were also able to vary the tempo between the songs depending on what tempo suited the song best. The non-musicians varied the tempo, singing the same song, from try to try and between the sessions. They did not adjust the tempo to each song as well as the musicians.

In another imagery-test Halpern once more found differences between musicians and non-musicians. During silent singing - ten different familiar songs - the musicians were able to vary the tempo with 153 BPM (between trials, as slow and as fast as possible) while the non-musicians only could vary the tempo with 87 BPM. Together with the previous test Halpern concluded that we have codes for tempo in our repre-

sentations but that it is possible by practice to get them more accurate. Just as musicians do.

Levitin and Cook (1996) found that folk songs, as used in Halpern's study, lack a standard tempo leading to a large variability in tempo. This is why non-musicians have difficulties keeping the pulse. If the subjects instead sing a popular song known from a CD the result will be much better. According to Levitin and Cook, 72 percent of the subjects came within 8 percent of the actual tempo on two consecutive trials. That is considered a very good result and indicates that representations of tempo are stored in long term memory.

A non-musical pulse-test was carried out by Rao et al. (1997). The subjects first heard a pulse with either a 300 ms interval or a 600 ms interval. Thereafter they were asked to synchronise the pulse by tapping with the index finger. In the next test they heard the pulse for a short time and were then exhorted to hold the pulse after the sound stopped. As soon as the guiding pulse went silent the participants increased their tapping tempo, more for the slower pulse.

In a similar experiment with both simple sequences and novel sequences Penhune, Zatorre and Evans (1998) found that the subjects deviated more than 20 percent in some of the trials during both conditions.

In the perception part of the first experiment above only right STG was active; except for in the fast pulse where also the left STG lit up (Rao et al., 1997). In the second experiment the activation looked like the faster condition of the first experiment (Penhune et al., 1998). When the subjects tried to hold the pulse themselves a great net of structures became active. According to Rao et al. the auditory processing was done in STG and right IFG (inferior frontal gyrus) and the timing was processed in putamen, thalamus and SMA. Cerebellum and motor cortex carried through the motor performance (the tapping). Penhune et al. had a different view, meaning that Cerebellum has a central role in forming the temporal structure. Obviously important is also the basal ganglia (including the putamen), SMA and SAC (i.e. in STG in this case).

Neural activation in music imagery

There are several questions concerning where in the brain processing is done. Let's take a look at two of them. The first one: Where are the sounds generated? It is known that SMA is contributing in motor planning, especially when producing verbal sounds. Because of this, Zatorre et al. (1996; see also Halpern & Zatorre, 1999) was surprised that SMA was active during song imagery. (Also in Ackermann & Reicker [2004] there was SMA activation during covert singing.) In their second experiment the melodies were non-verbal. Might SMA be the structure generating (musical) sounds? Halpern (2001) tried to find this out in a TMS-study (transcranial magnetic stimulation) but the results were even more puzzling. Timbre is known not to be involved in subvocalization (e.g. Smith et al., 1992). Thus, Halpern et al. (2004) expected no SMA activation in their timbre imagery experiment. SMA was active and the question comes one step closer to the resolution. It is likely that SMA is part of the planning of or generation of (musical) imagery sounds.

Second question: Where do we “hear” the sounds? In Zatorre et al. (1996), Halpern and Zatorre (1999) and Schürmann et al. (2002) both anterior and posterior STG were active during musical imagery. Halpern et al. (2004) found no anterior activation in their timbre imagery study. This should indicate that posterior STG (also posterior SAC, BA 22) is the seat of musical sound processing during imagery. It is however confusing that the right STG is active in some cases but not others and the left STG is active in some cases but not others. Sometimes they are active at the same time.

SUBVOCALIZATION AND INNER SPEECH

The term subvocalization is connected to the working-memory model presented by Baddeley and Hitch (1974). To remember something we repeat it silently in our head; we use our “inner voice” or subvocalise. In this rehearsal process we repeat the word (or sound) to our “inner ear”. The difference between subvocalization and inner speech is that the latter can be drawn directly from long-term memory without engaging the phonological loop or need any kind of articulation.

Some believe that inner speech is the same as thinking. To say this you have to be convinced that thinking is only verbal. Inner speech or speech imagery are verbal thoughts but there are other kinds of thinking. These verbal thoughts might be involved in problem solving, calculating or planning but in this text that is on the border of imagery research.

Inner speech differs from overt speech in the sense that the latter is (1) audible to others and (2) using articulation. This means that inner speech only needs to use a phonological code (MacKay, 1992). We never seem to commit articulatory mistakes while speaking silent and people with injured articulation areas in the brain are still able to speak silently.

This “inner voice” (or verbal thought) apparently lacks pitch, timber, prosody and other effects that can vary in people’s audible voices. Therefore it is sex neutral or hermaphrodite (Geiselman & Bjork, 1980). Though it is still possible to imagine words spoken in a male or female voice.

In a PET experiment by Ryding, Brådvik, and Ingvar (1996) they found a different neural activation pattern during inner speech compared to audible speech. There was left-sided activation in regions related to speech perception and speech motor control including Broca’s area and SMA. They also found activation in a right dorsolateral prefrontal area (see also Numminen & Curio, 1999; compare with Jonides, 2000).

In a comparison overt speech uses anterior insula (known to be involved in articulation) while inner speech does not. Both use left SMA, left motor cortex and right cerebellum in the sound-production (Ackermann & Riecker, 2004).

When schizophrenic patients have their episodes of auditory hallucinations – they hear someone talking to them – there are strong indications that they themselves are using speech imagery, though not aware of it (Smith, 1992). During an episode the same neural structures are active as in speech imagery (e.g. Thompson, 2000).

Because there are sounds that we can't reproduce with the voice (e.g. waves crashing on the beach or a symphony orchestra tuning up) it is unlikely that auditory imagery is based on subvocalization (Baddeley, 1999). Timbre is also known to be a sound characteristic that we subvocally can't handle (Smith, Reisberg & Wilson, 1992). In spite of this Smith et al. (1992; 1995; also Reisberg, 2000) give subvocalization an important role in auditory imagery. Of importance in this case is the supposed neural connection between the "inner voice" and the "inner ear".

The area called the phonological store (in the working-memory model) is synonymous with the "inner ear". It is localised to posterior STG (BA 22) by some (e.g. Smith et al., 1995; Ackermann & Reicker, 2004; Hickok & Poeppel, 2004; see also Husain et al., 2004), BA 39/40 by some (e.g. Gisselgård et al., 2004) and in Broca's area, BA 44/45 by others (e.g. Baddeley, 2003 [for other opinion more similar to the first above, see Baddeley, 1999, 309]). BA 44/45, though, has to low memory capacity to be the phonological store (Gisselgård et al., 2004). According to Hickok and Poeppel (2004) Spt (in BA 39/40) is not a store but an interface connecting the "inner voice" to the "inner ear". (For a fourth suggestion see Logie et al., 2003.)

Subvocal rehearsal is also synonymous with the "inner voice". This one is localised in SMA and premotor cortex, both in BA 6, and Broca's area, BA 44/45 (Gisselgård et al., 2004; Hickok & Poeppel, 2004; Ackermann & Reicker, 2004; Jonides, 2000). Much is also pointing to BA 46/9 being an important structure during subvocal rehearsal (Braver et al., 1997; Jonides, 2000; Tillman et al., 2003).

There is no controversy in claiming that subvocalization is important to auditory imagery. We can subvocally sing a melody, pronounce letter/number-strings as "NEI 4 IOS" silent (Smith et al., 1992) and make onomatopoeic sounds. Inner speech uses less or other neural structures just as generation of environmental sounds do. Subvocalization might still be substantial for auditory imagery but it has its limitations.

After a series of experiments Reisberg et al. (1989) suggested that there are two types of auditory imagery: *pure* imagery and *enacted* imagery. Pure imagery does not need articulation, thus not involving subvocalization. This includes environmental sounds (e.g. glass breaking) and timbre. Enacted imagery draws on subvocalization.

Pure and enacted imagery have two steps in the system in common. According to Reisberg (2000) the phonological store (buffer) and the central executive (see Baddeley & Hitch, 1974) are used by both of them. If the phonological store is busy processing incoming information from the auditory perception system it will interrupt on both pure and enacted imagery.

Enacted imagery uses one more subsystem which is responsible for the articulation being part of subvocalization. If the articulation is blocked (e.g. biting a pencil) a letter/number-string like "D2R" will become difficult to pronounce subvocally. On the contrary there are no problems imagining pure images during these conditions.

Words that rhyme (TAPE and FAIP) also seem to rely on subvocalization. Homophones (RAIN and REIGN) on the other hand are not disturbed either by blocked articulation or by perception input (blocking the "inner ear") (Reisberg, 2000). How is

this possible? Inner speech, in the perspective of this two-path system, then, must be considered a kind of pure imagery.

THE KNOWLEDGE TODAY

Summing it up we seem to have representations for loudness, pitch, temporal extent, tempo (pulse), timbre among others which we can retain during auditory imagery. In some cases these characteristics are not as obligatory as in environmental sounds. In other cases the representations are accurate. We have also seen that auditory imagery can work inter-modally, e.g. viewing a monitor and imagining the corresponding sound. It is probable that there are two kinds of auditory imagery. Reisberg and collaborators suggest that we are able to use both pure and enacted imagery where the latter is synonymous with subvocalization.

An important question is; how much neural substrate has perception and imagery in common? This depends on the imagery task but generally imagery always activates some part of the auditory cortex (for a good example where parts of PAC and SAC were active see Yoo, Lee & Choi, 2001). A lesion in auditory cortex impairing perception also impairs imagery in a corresponding way. Much points to the posterior STG, an associative auditory area, as accommodating the “inner ear”. There is still no sign on Heschl’s gyrus activation during imagery (Halpern & Zatorre, 1999; Halpern et al., 2004). In visual imagery the primary visual cortex has been activated several times (e.g. Kosslyn, 1994/1999; Le Bihan, Klein & Dohi, 2002).

There are indications that SMA is an important area in auditory imagery (it is also active in visual imagery [e.g. Ganis, Thompson & Kosslyn, 2004] and motor imagery [e.g. Grézes & Decety, 2002]). A lesion in SMA leads to mutism (Brown et al., 2004). Does this implicate loss of imagery ability? Halpern (2001) found no evidence for this but could not really draw any conclusions. We have seen that subvocalization including melodies and timbre activate SMA. Areas that also could be involved with imagery sound generation are IFG, Brocas area and Insula.

In the big picture nothing is against auditory imagery as a phenomenon. The research has not yet met any heavy criticism, or a vivid debate as in the visual imagery debate. Why is this?

CHALLENGES AHEAD

The major criticism so far is that too little research has been done. Some areas are still underrepresented. Let’s take a look at those areas:

1. Inner speech or verbal thoughts need more research, both behavioural and neural.
2. Object and environmental sounds in imagery needs more research. Especially where in the brain it is processed. Is sma involved?

3. No one except Intons-Peterson and collaborates have studied loudness. This area could be important because loudness is probably processed in the pac. Maybe this is how imagery tasks can activate Heschl's gyrus.
4. No one has yet developed an artificial neural network (ANN) model trying to simulate or explain the neural processes involved in auditory imagery.
5. What can we say about the differences and similarities between auditory imagery and visual imagery? Are there for example individual dominance towards one kind of mental imagery?
6. Finally, no one has developed a model or theory as comprehensive as Kosslyn's (1994/1999) protomodel in visual imagery. Intons-Peterson (1992) discussed possibilities and Reisberg (2000) has sketched on a model but there is much to be done.

Everyone in this field of research should discuss the borders of auditory imagery (what is and what is not auditory imagery?) as well as the reliability and possibility of imagery research.

REFERENCES

- Ackermann, H. & Riecker, A. (2004) *The contribution of the insula to motor aspects of speech production: A review and a hypothesis*. *Brain and Language*, 89, 320–328.
- Aleman, A. et al. (2000) *Music training and mental imagery ability*. *Neuropsychologia*, 38, 1664–1668.
- Anderson, R. A. (1982) *Speech imagery is not always faster than visual imagery*. *Memory & Cognition*, 10, 371–380.
- Arnott, S. R. et al. (2004) *Assessing the auditory dual-pathway model in humans*. *NeuroImage*, 22, 401–408.
- Artacho-Péruña, E. et al. (2004) *Quantitative estimation of the primary auditory cortex in human brains*. *Brain Research*, 1008, 20–28.
- Baddeley, A. D. (1999) *Essentials of human memory*. East Sussex, Storbritannien: Psychology press. ISBN 0-86377-545-4.
- Baddeley, A. D. (2003) *Working memory and language: an overview*. *Journal of communication disorders*, 36, 189–208.
- Baddeley, A. D. & Hitch, G. J. (1974) *Working memory*. In Bower G. A. (ed.) (1974) *The psychology of learning and motivation*. New York, USA: Academic press.
- Baddeley, A. D. & Logie, R. (1992) *Auditory imagery and working memory*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Belin, P., Zatorre, R. J. & Ahad, P. (2002) *Human temporal-lobe response to vocal sounds*. *Cognitive Brain Research*, 13, 17–26.
- Braver, T. S. et al. (1997) *A parametric study of prefrontal cortex involvement in human working memory*. *NeuroImage*, 5, 49–62.
- Brown, S. et al. (2004) *The song system of the human brain*. *Cognitive Brain Research*, 20, 363–375.
- Crowder, R. G. (1989) *Imagery for musical timbre*. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 472–478.
- Crowder, R. G. & Pitt, M. A. (1992) *Research on memory/imagery for musical timbre*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Denis, M. (2001) *Mental imagery, psychology of*. Article in *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier Science Ltd. ISBN 0-08-043076-7.
- Deutsch, D. (1972) *Mapping of interactions in pitch memory store*. *Science*, 175, 1020–1022.

- Deutsch, D. & Pierce, J. R. (1992) *The climate of auditory imagery and music*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Di Salle, F. et al. (2003) *fMRI of the auditory system: understanding the neural basis of auditory gestalt*. *Magnetic Resonance Imaging*, 21, 1213-1224.
- Ganis, G., Thompson, W. L. & Kosslyn, S. M. (2004) *Brain areas underlying visual mental imagery and visual perception: an fMRI study*. *Cognitive Brain Research*, 20, 226-241.
- Geiselman, R. E. & Bjork, R. A. (1980) *Primary versus secondary rehearsal in imagined voices: Differential effects on recognition*. *Cognitive Psychology*, 12, 188-205.
- Geiselman, R. E. & Glenny, J. (1977) *Effects of imagining speakers' voices on the retention of words presented visually*. *Memory & Cognition*, 5, 499-504.
- Gisselgård, J. et al. (2003) *The irrelevant speech effect: a PET study*. *Neuropsychologia*, 41, 1899-1911.
- Grézes, J. & Decety, J. (2002) *Does visual perception of object afford action: Evidence from a neuroimaging study*. *Neuropsychologia*, 40, 212-222.
- Halpern, A. R. (1988a) *Mental scanning for auditory imagery for songs*. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 434-443.
- Halpern, A. R. (1988b) *Perceived and imagined tempos of familiar songs*. *Music perception*, 6, 193-202.
- Halpern, A. R. (1989) *Memory for the absolute pitch of familiar songs*. *Memory and Cognition*, 17, 572-581.
- Halpern, A. R. (1992) *Musical aspects of auditory imagery*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Halpern, A. R. (2001) *Cerebral substrates of musical imagery*. In *The biological foundations of music*. The Annals of New York Academy of Sciences, 930, 179-192.
- Halpern, A. R. & Zatorre R. J. (1999) *When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies*. *Cerebral Cortex*, 9, 697-704.
- Halpern, A. R. et al. (2004) *Behavioral and neural correlates of perceived and imagined musical timbre*. *Neuropsychologia*, 42, 1281-1292.
- Hart, H. C., Palmer, A. R. & Hall, D. A. (2002) *Heschl's gyrus is more sensitive to tone level than non-primary auditory cortex*. *Hearing Research*, 171, 177-190.
- Hickok, G. & Poeppel, D. (2004) *Dorsal and Ventral streams: a framework for understanding aspects of the functional anatomy of language*. *Cognition*, 92, 67-99.
- Hubbard, T. L. & Stoeckig, K. (1992) *The representation of pitch in musical images*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Husain, F. T. et al. (2004) *Relating neuronal dynamics for auditory object processing to neuroimaging activity: a computational modeling and an fMRI study*. *NeuroImage*, 21, 1701-1720.
- Intons-Peterson, M. J. (1980) *The role of loudness in auditory imagery*. *Memory & Cognition*, 8, 385-393.
- Intons-Peterson, M. J. (1992) *Components of auditory imagery*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Intons-Peterson, M. J., Russell, W. & Dressel, S. (1992) *The role of pitch in auditory imagery*. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 233-240.
- Jonides, J. (2000) *Mechanisms of verbal working memory revealed by neuroimaging studies*. In Landau et al. (ed.) (2000) *Perception, Cognition, and Language. Essays in honor of Henry and Lila Gleitman*. Cambridge, Massachusetts: A Bradford Book. ISBN 0-262-12228-6.
- Kosslyn, S. M. (1973) *Scanning visual images: Some structural implications*. *Perception and Psychophysics*, 14, 90-94.
- Kosslyn, S. M. (1999) *Image and Brain. The Resolution of the Imagery Debate*. Massachusetts, USA: A Bradford Book, the MIT Press. Fourth Printing. First printing, 1994. ISBN 0-262-61124-4.
- Le Bihan, D., Klein, I. & Dohi, M. (2002) *Making Area V1 Glow in Visual Imagery*. In Galaburda, A. M., Kosslyn, S. M. & Christen, Y. (ed.) (2002) *The languages of the brain*. Massachusetts, USA: Harvard University Press. ISBN 0-674-00772-7.
- Levitin, D. J. (1994) *Absolute memory for musical pitch: Evidence from the production of learned melodies*. *Perception & Psychophysics*, 56, 414-423.

- Levitin, D. J. & Cook, P. R. (1996) *Memory for musical tempo: Additional evidence that auditory memory is absolute*. *Perception & Psychophysics*, 58, 927–935.
- Lewis, P. A. & Miall, R. C. (2003) *Brain activation patterns during measurement of sub- and supra-second intervals*. *Neuropsychologia*, 41, 1583–1592.
- Liegeois-Chauvel, C. et al. (1998) *Contribution of different cortical areas in the temporal lobes to music processing*. *Brain*, 121, 1853–1867.
- Logie, R. H. et al. (2003) *Brain activation and the phonological loop: The impact of rehearsal*. *Brain and Cognition*, 53, 293–296.
- MacKay, D. G. (1981) *The problems of rehearsal or mental practice*. *Journal of Motor Behavior*, 13, 274–285.
- MacKay, D. G. (1992) *Constraints on theories of inner speech*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Menon, V. et al. (2002) *Neural correlates of timbre change in harmonic sounds*. *NeuroImage*, 17, 1742–1754.
- Morosan, P. (2001) *Human Primary Auditory Cortex: Cytoarchitectonic Subdivisions and Mapping into a Spatial Reference System*. *NeuroImage*, 13, 684–701.
- Numminen, J. & Curio, G. (1999) *Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex*. *Neuroscience Letters*, 272, 29–32.
- Patel, A. D. & Balaban, E. (2001) *Human pitch perception is reflected in the timing of stimulus-related cortical activity*. *Natur Neuroscience*, 4 (8), 839–844.
- Penhune, V. B., Zatorre, R. J. & Evans, A. C. (1998) *Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction*. *Journal of Cognitive Neuroscience*, 10 (6), 752–765.
- Penhune, V. B., Zatorre, R. J. & Feindel, W. H. (1999) *The role of auditory cortex in retention of rhythmic patterns as studied in patients with temporal lobe removals including Heschl's gyrus*. *Neuropsychologia*, 37, 315–331.
- Pitt, M. A. (1994) *Perception of pitch and timbre by musically trained and untrained listeners*. *Journal of Experimental Psychology: Human Perception and Performance*, 20, 976–986.
- Pitt, M. A. (1995) *Evidence for a central representation of instrument timbre*. *Perception and Psychophysics*, 57, 43–55.
- Pitt, M. A. & Crowder, R. G. (1992) *The role of spectral and dynamic cues in imagery for musical timbre*. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 728–738.
- Platel, H. et al. (1997) *The structural components of music perception. A functional anatomical study*. *Brain*, 120, 229–243.
- Rao, S. M. et al. (1997) *Distributed neural systems underlying the timing of movements*. *Journal of Neuroscience*, 17 (14), 5528–5535.
- Rauschecker, J. P. & Tian, B. (2003) *Processing of "what" and "where" in auditory association cortex*. *International Congress Series*, 1250, 37–51.
- Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0-8058-0556-7.
- Reisberg, D. (2000) *The detachment gain: the advantage of thinking out loud*. In Landau et al. (ed.) (2000) *Perception, Cognition, and Language. Essays in honor of Henry and Lila Gleitman*. Cambridge, Massachusetts: A Bradford Book. ISBN 0-262-12228-6.
- Reisberg, D. et al. (1989) *Enacted auditory images are ambiguous; Pure auditory images are not*. *Quarterly Journal of Experimental Psychology*, 41A, 619–641.
- Ryding, E., Brådvik, B. & Ingvar, D. H. (1996) *Silent speech activates prefrontal cortical regions asymmetrically, as well as speech-related areas in the dominant hemisphere*. *Brain and Language*, 52, 435–451.
- Samson, S., Zatorre, R. J. & Ramsay, J. O. (2002) *Deficits of musical timbre perception after unilateral temporal-lobe lesion revealed with multidimensional scaling*. *Brain*, 125, 511–523.
- Schürmann, M. et al. (2002) *Mind's ear in musician: where and when in the brain*. *NeuroImage*, 16, 434–440.
- Scott, S. K. & Wise, R. J. S. (2003) *PET and fMRI studies of the neural basis of speech perception*. *Speech Communication*, 41, 23–34.
- Scott, S. K. & Wise, R. J. S. (2004) *The functional neuroanatomy of prelexical processing in speech perception*. *Cognition*, 92, 13–45.

- Shepard, R. R. & Metzler, J. (1971) *Mental rotation of three-dimensional objects*. *Science*, 171, 701–703.
- Smith, J. D. (1992) *The auditory hallucinations of schizophrenia*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0–8058–0556–7.
- Smith, J. D., Reisberg, D. & Wilson, M. (1992) *Subvocalization and auditory imagery: interactions between the inner ear and inner voice*. In Reisberg, D. (ed.) (1992) *Auditory imagery*. New Jersey, USA: Lawrence Erlbaum Associates, Inc. Publishers. ISBN 0–8058–0556–7.
- Smith, J. D., Wilson, M. & Reisberg, D. (1995) *The role of subvocalization in auditory imagery*. *Neuropsychologia*, 33 (11), 1433–1454.
- Thompson, R. F. (2000) *The Brain. A Neuroscience Primer*. New York: Worth Publishers. ISBN 0–7167–3226–2.
- Tillman, B., Janata, P. & Bharucha, J. J. (2003) *Activation of the inferior frontal cortex in musical priming*. *Cognitive Brain Research*, 16, 145–161.
- Weber, R. J. & Brown, S. (1986) *Musical imagery*. *Music Perception*, 3, 411–426.
- Yoo, S-S, Lee, C. U. & Choi, B. G. (2001) *Human brain mapping of auditory imagery: event-related functional MRI study*. *Neuroreport*, 12 (14), 3045–3049.
- Zatorre, R. J. (2003a) *Sound analysis in auditory cortex*. *Trends in Cognitive Sciences*, 26 (5), 229–230.
- Zatorre, R. J. (2003b) *Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function*. *Nature Neuroscience*, 6, 692–695.
- Zatorre, R. J., Belin, P. & Penhune, V. B. (2002a) *Structure and function of auditory cortex: music and speech*. *Trends in Cognitive Sciences*, 6 (1), 37–46.
- Zatorre, R. J. et al. (2002b) *Where is 'where' in the human auditory cortex?* *Nature neuroscience*, 5, 905–909.
- Zatorre, R. J., Bouffard, M. & Belin, P. (2004) *Sensitivity to auditory object features in human temporal neocortex*. *The Journal of Neuroscience*, 24, 3637–3642.
- Zatorre, R. J. & Halpern, A. R. (1993) *Effect of unilateral temporal-lobe excision on perception and imagery of songs*. *Neuropsychologia*, 31, 221–232.
- Zatorre, R. J. et al. (1996) *Hearing in the mind's ear: A PET investigation of musical imagery and perception*. *Journal of Cognitive Neuroscience*.