

Brain activity during selective listening to natural speech

Kimmo Alho ¹, Victor A. Vorobyev ²

¹Department of Psychology, PO Box 9, FIN-00014 University of Helsinki, Finland, ²Department of Psychology, FIN-20014 University of Turku, Finland

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Brain activity associated with selective auditory attention
4. Brain activity associated with speech listening
5. Brain activity associated with selective listening to speech
5. Summary and conclusions
6. Acknowledgements
7. References

1. ABSTRACT

Human brain functions involved in selective attention to particular sounds have been studied extensively with non-invasive measurements of electro-magnetic and hemodynamic brain activity. Here we review studies indicating that selection of the attended sounds for further processing occurs in the auditory cortex. The exact locus of this selection process in the auditory cortex appears to depend on the auditory attribute, i.e., location or pitch, separating the attended sounds from the irrelevant ones. Recent neuroimaging studies extend this finding from processing of non-speech sounds to attentional selection of relevant speech differing by its location or speaker identity from concurrent irrelevant speech. These studies suggest also that selective listening to speech depends less on prefrontal control functions than other kinds of listening tasks demanding selective attention.

2. INTRODUCTION

A listener in the middle of a noisy crowd of chatting people may still selectively attend to the voice of a particular speaker. Some 50 years ago, a number of experiments demonstrated that a difference in the location (e.g., left vs. right ear), pitch (e.g., a male vs. female voice), or both between the attended speech and concurrent irrelevant speech facilitates such selective listening compared with a situation where the relevant and irrelevant speech messages are separated only by their semantic contents, i.e., two different messages are spoken simultaneously by the same voice in the same location (1-5). These studies also showed that such selective listening tasks are quite easy to perform and that the irrelevant speech messages are not processed much. For example, a listener attending to a speech message delivered to one ear and repeating (“shadowing”) it aloud word by word does

Selective listening to speech

not necessarily notice when the language of the irrelevant speech message delivered to the opposite ear is changed or when the irrelevant speech message is momentarily reversed, i.e., presented backwards (1). These findings supported the so-called early-selection or filtering theories of selective attention (6).

However, the early studies on selective listening demonstrated that even irrelevant speech is processed to some degree during selective listening to another speech message. For instance, a listener attending to a message delivered to one ear may notice physical changes or novel sounds in a to-be-ignored input delivered to the other ear, e.g., a change from a male voice to a female voice or tones superimposed on irrelevant speech (1). Moreover, it was demonstrated that significant words in the irrelevant message, for example, the listener's own name (7,8) or a word semantically related to the attended speech (9,10), may sometimes catch the listener's attention. Such findings led to a modified filter theory proposing that the attentional filter attenuates but does not totally prevent the processing of irrelevant sounds (9,11). Therefore irrelevant speech may sometimes activate word representations in a mental dictionary, especially if these representations are highly primed, such priming being involuntary and unconscious.

However, it was also demonstrated (12) that words in irrelevant speech delivered to one ear that are semantically related to the shadowed words delivered to the other ear distract shadowing more in the beginning of the task than later during the task performance. Such results suggest that attentional selection of relevant speech or other sounds is a dynamic process that becomes more efficient during the performance of a selective listening task.

3. BRAIN ACTIVITY ASSOCIATED WITH SELECTIVE AUDITORY ATTENTION

Brain activity associated with selective attention to sounds has been studied extensively with event-related brain potentials (ERPs) elicited by attended and unattended sounds in the scalp-recorded electroencephalogram (EEG). In a paradigm developed by Hillyard and colleagues (13), a human listener is presented with tones delivered randomly to the left and right ear or with tones varying randomly in pitch (14), or both (15), and the listener is instructed to selectively attend to tones with a designated location and/or pitch in order to find deviant target tones (e.g., tones with a slightly higher pitch or shorter duration) among otherwise constant attended tones. These studies showed that ERPs elicited by the attended tones are negatively displaced in relation to ERPs to the unattended tones. This attentional effect begins at the time zone of the N1 deflection (peak latency about 100 ms from sound onset), but it may last for several hundred milliseconds as demonstrated first by Näätänen and colleagues (16). The generator process of the early portion of this attentional effect has been localized to the auditory cortex by scalp current density analysis of ERP responses (17,18,19) and by modeling the sources of magnetoencephalographic (MEG) counterparts of ERPs (20,21,22,23,24). These findings suggest that selection of attended sounds on the basis of their location or pitch takes

place in the auditory cortex. ERP studies also indicate involvement of the prefrontal brain areas in selective listening, as lesions in the dorsolateral prefrontal cortex diminish selective attention effects on auditory ERPs (25). Moreover, the early ERP attention effect generated in the auditory cortex is followed by a later attention-related negative ERP deflection (14,26). Scalp current density analysis suggests that this later attention effect originates from the frontal lobes (16).

According to Näätänen's (27-29) attentional-trace theory, selection of attended sounds for further processing on the basis of their location or pitch is based on matching auditory inputs with the attentional trace, which is a facilitation pattern in the auditory cortex representing the attended sounds. This theory is supported by studies showing that the negative-polarity displacement of the ERP to the attended sounds in relation to the ERP to the same sounds when they are unattended during attention to other sounds is caused by a large and long-duration processing negativity (PN) ERP component elicited by the attended sounds, which match perfectly with the attentional trace (31-34). Furthermore, these studies have shown that even to-be-ignored sounds elicit some PN that is larger in amplitude and longer in duration the more these sounds resemble the attended sounds, i.e., the better they match with the attentional trace (31-34). According to Näätänen (27-29), the early component of PN overlaps with the exogenous N1 components and therefore enhances the amplitude of the N1 deflection. While the early PN component is generated in the auditory cortex by the selective matching process, the later (>300 ms from sound onset) and more frontally distributed PN component (14,17,35) might be generated by a prefrontal process maintaining the attentional trace by top-down modulation of some neuron populations in the auditory cortex representing the attended sounds. In addition to active maintenance, the attentional trace appears to depend on physical reinforcement given by the attended sounds, since to-be-attended sounds delivered at very long intervals do not elicit any PN (30).

The attentional-trace theory contradicts the original proposal of Hillyard *et al.* (13) that the enhanced (more negative) N1 deflection elicited by the attended sounds compared to the unattended sounds would simply indicate amplified processing of the attended sounds or attenuated processing of the unattended sounds. Yet, Hillyard *et al.*'s proposal is supported by results showing that when attended and unattended tones are delivered to the opposite ears at a rate of several tones per second, the attention effect at the N1 latency is generated in the same auditory-cortex region as the supratemporal subcomponent of the N1 (24). Moreover, under such conditions, this N1 effect may be preceded by an even earlier attention effect, namely, a positive-polarity enhancement of the ERP to the attended tones as early as 20-50 ms after sound onset (24,36). However, the existence of such an early filtering process capable of selecting attended sounds among other sounds in conditions with large physical separations between the attended and unattended sounds does not rule out the existence of a later selection process based on an

Selective listening to speech

attentional trace and capable of fine-tuned discriminations between the attended and unattended sounds differing only slightly, e.g., in pitch (30,32,37).

Moreover, Näätänen's (27-29) suggestion that selective listening is based on attentional selection of sounds at a relatively late processing stage in the auditory cortex, rather than, for example, gating of sensory inputs ascending from the thalamus to the auditory cortex, is supported by functional magnetic resonance imaging (fMRI) findings indicating that the effects of selective attention occur predominantly in non-primary areas of the auditory cortex (38-40). Separability of the attention-related activations from the exogenous stimulus-dependent activations in the auditory cortex is also supported by fMRI results showing different dependency of attention-related and stimulus-dependent activations on the presentation rate of sounds (41).

Näätänen's (27-29) proposal that prefrontal cortical areas have a central role in the control and maintenance of auditory selective attention has been also supported by a number of fMRI and positron emission tomography (PET) studies (42-44). For example, in their PET study, Alho and colleagues (44) used a paradigm similar to that applied in many ERP and MEG studies (13,23,24,36). They presented tones at a fast rate (6 per second) to each ear and the participants' task was to attend either to the left-ear tones or to the right-ear tones in order to find from the attended tone sequence occasional target tones of 500 Hz occurring among the frequent non-target tones of 400 Hz. In a control condition, the left-ear and right-ear tone sequences were to be ignored and visually presented letters were to be attended. Comparison of brain activity in the selective listening and visual-attention conditions showed that selective listening to tones at a designated location is associated with bilateral activations in the auditory and prefrontal cortices, as well as with a unilateral activation in the right parietal cortex. The latter finding is consistent with results showing that the right parietal cortex has an important role in directing attention spatially (45). For example, patients with right parietal lesions often neglect sounds occurring in the left hemisphere (46).

According to Näätänen's original theory (27), the attentional trace represents different features of the attended sounds. However, it appears that the sound location and pitch are processed separately in attentional selection. For example, varying the pitch of tones over a wide range during selective listening to tones at a designated location or delivering tones randomly to the left or right ear during selective listening to tones of a designated pitch does not affect the amplitude or latency of the PN elicited by the attended tones (47). Moreover, when the listeners selectively attend to tones occurring among to-be-ignored tones differing from the attended ones in location, pitch, or both, long-duration PNs lasting for several hundred milliseconds are elicited both by the attended tones, as well as by those to-be-ignored tones that share either the pitch or location with the attended tones (47-49). However, the PNs to these to-be-ignored tones are smaller in amplitude than the PNs to the attended tones.

This is probably because the attended tones, defined by both location and pitch, elicit overlapping location-specific and pitch-specific PNs and even an additional PN generated by integration of location and pitch information (48,49). Thus, selection by sound location and selection by pitch appear to be based on independent, exhaustive processes, i.e., finding a mismatch between the auditory input and the attentional trace on one dimension, e.g., location, does not lead to termination of processing of the sound on other dimensions, e.g., pitch. Furthermore, the scalp distribution of the PN for the to-be-ignored tones sharing the location with the attended tones and the scalp distribution of the PN for the to-be-ignored tones sharing the pitch with the attended tones differ from each other (48,49). This suggests that at least partly different areas in the auditory cortex are involved in generating these effects. These findings are in line with studies in non-human primates indicating that spatial and non-spatial auditory features are processed in different functional areas of the auditory cortex (50,51,52).

Evidence for separate processing of spatial and non-spatial auditory features in the human auditory cortex has been also found in fMRI studies. Alain and colleagues (53) investigated brain activity during sound localization and identification using a delayed matching-to-sample task, which required the participants to hold either the sound location or pitch in their working memory for a comparison with the location or pitch of a subsequent sound, respectively. Stronger activity during the location task than during the pitch task was observed in several brain regions including the posterior middle temporal gyrus and the superior and inferior parietal cortices in both hemispheres, as well as the right superior and left middle frontal gyri (Figure 1, orange squares). The pitch task, in turn, was associated with a stronger activity than the location task in the right inferior frontal cortex and in the auditory cortices bilaterally, the latter activations extending from the primary auditory cortices to the anterior auditory association cortices on the supratemporal plane (Figure 2, orange squares).

Separate spatial and non-spatial auditory processing was also suggested by the fMRI study of Maeder and colleagues (54). They investigated brain activity during sound localization and during recognition of complex environmental sounds. The inferior parietal lobule and parts of the premotor cortex were more active bilaterally during sound localization than during sound recognition, as well as the right prefrontal cortex, including the inferior frontal gyrus (Figure 1, yellow squares). The recognition task, in turn, was associated with a higher activity than the localization task bilaterally in the anterior part of the middle temporal gyrus and in the inferoposterior part of the precuneus, as well as in the posterior part of the left inferior frontal gyrus and in the left parahippocampal gyrus (Figure 2, yellow squares).

4. BRAIN ACTIVITY ASSOCIATED WITH SPEECH LISTENING

Numerous PET and fMRI studies have focused on brain activity during listening to continuous speech (55-

SPATIAL AUDITORY PROCESSING

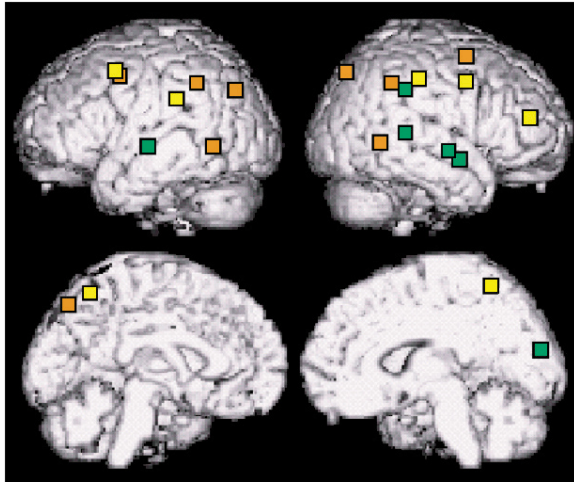


Figure 1. Cortical activations associated with spatial processing. Colored squares indicate projections of the loci of activation maxima onto the lateral or medial surface of each hemisphere. Orange indicates local maxima of the activation difference between spatial and non-spatial auditory working memory tasks in the study by Alain *et al.* (53). Yellow indicates local maxima from the comparison between sound-localization and sound-recognition tasks in the study by Maeder *et al.* (54), and dark green refers to the activation difference between selective listening to either speech delivered to the left ear or speech delivered to the right ear (data averaged across the two conditions) and reading (cf. Figure 2) in the study by Alho *et al.* (83). For the study by Maeder *et al.* (54), only maxima obtained at least in 12 out of the 18 participants are shown.

9). These studies found activation in the left inferior prefrontal cortex (55-59), at or near Broca's area (60). These left prefrontal areas are activated also during reading tasks and they are thought to be involved in semantic or syntactic processing, or both (56,57,61-66). Another common finding during listening to speech is enhanced activity in the left superior temporal cortex, especially in the areas posterior to the primary auditory cortex, i.e., in Wernicke's area (67), which has a central role in phonological processing, as indicated by speech-perception deficits in patients with lesions in this area (67-69), as well as by PET and fMRI studies demonstrating that even meaningless syllables activate this brain region (70-72).

However, temporal-cortex areas lateral and anterior to the primary auditory cortex are also involved in speech processing. For example, in their PET experiment, Scott and colleagues (73) studied hemodynamic brain activations during listening to natural speech (Sp), noise-vocoded speech (VCo), spectrally rotated speech (RSp), and rotated noise-vocoded speech (RVCo). VCo is produced by replacing the speech signal in different frequency ranges with noise that follows the energy changes in speech. It sounds like a harsh whisper and is comprehensible after some training. In RSp, the spectral structure of ordinary speech is rotated. Its temporal and

spectral complexity is similar to ordinary speech and it contains some phonetic structures (e.g., fricatives) and it may become partly intelligible after extensive training. In RVCo, in turn, the spectral structure of VCo is inverted making speech totally unintelligible even after extensive training. It was found that a part of the left superior temporal gyrus lateral and anterior to the primary auditory cortex along with the left posterior superior temporal sulcus respond to the presence of phonetic information, i.e., these areas were more active in the Sp, VCo and RSp conditions than in the RVCo condition. More anterior regions in the left superior temporal sulcus were activated only during listening to intelligible signals, i.e., they were more active in the Sp and VCo conditions than in the RSp and RVCo conditions. The right ventrolateral superior temporal gyrus, in turn, was more active during listening to signals with dynamic pitch variation (Sp and RSp) than during listening to signals consisting of noise (VCo and RVCo).

In their fMRI study, Davis and Johnsrude (74) investigated brain activity associated with speech comprehension. They distorted spoken sentences in three different ways and varied the amount of distortion. In one condition, the speech was split into short segments and every other segment was replaced with signal-correlated noise that had the same spectral profile and amplitude envelope as speech but was unintelligible. The amount of intelligibility was increased by decreasing the duration of noise segments (separated by 200-ms speech segments) from 500 ms down to 100 ms. In another condition, they presented noise-vocoded speech (see above) and improved intelligibility by increasing the number of frequency bands used in vocoding. In a third condition, speech was presented over a background of speech-spectrum noise and intelligibility was improved by attenuating the noise. As participants listened to these different speech signals, brain activity along the superior and middle temporal gyri of the left hemisphere and in a smaller homologous area in the right hemisphere increased with improving intelligibility of speech. Correlations between intelligibility and brain activity were also found in the left inferior frontal gyrus and in the left hippocampus.

The aforementioned superior temporal sulcus areas appear to overlap with those that Belin *et al.* (75) found to be responsive to human voice (Figure 2, blue squares). In their fMRI study passive listening (no task instruction was given to the participants) to vocal sounds including both speech and non-speech (e.g., coughs, laughs, and cries) was associated with a higher activity along the superior temporal sulcus in relation to listening to other complex sounds. These voice-specific activations were observed bilaterally but they appeared to be stronger in the right hemisphere. The important role of the right superior temporal sulcus region in processing human voice was supported by the results of von Kriegstein and colleagues (76). In their fMRI study the listeners were presented with sentences varying in verbal content and speaker identity. In one condition the listeners were to attend to the verbal content and to find a particular sentence irrespective of who spoke the sentence, whereas in the other condition they were to recognize a particular speaker irrespective of the

VOICE/SOUND RECOGNITION

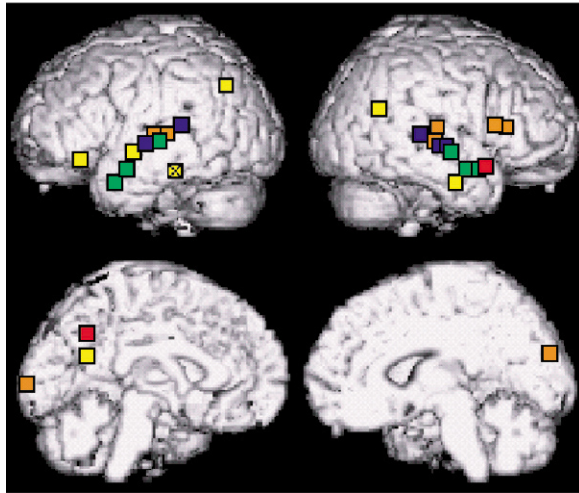


Figure 2. Cortical activations associated with voice and sound recognition. Colored squares indicate projections of the activation maximum locations onto the lateral or medial surface of each hemisphere. Orange indicates local maxima of the activation difference between non-spatial and spatial auditory working memory tasks in the study by Alain *et al.* (53). Yellow indicates results of comparison between environmental sound recognition and sound localization tasks in the study by Maeder *et al.* (54). Blue indicates local maxima revealed by comparison of passive listening to human voice with listening to other complex sounds in the study by Belin *et al.* (75). Red indicates local maxima of the activation difference between attention to the voice and attention to the content during listening to sentences in the study by von Kriegstein *et al.* (76). Light green squares indicate local maxima of the activation difference during selective listening to either a male or female voice (data averaged across the two conditions) and reading (cf. Figure 1) in the study by Alho *et al.* (84). For Maeder *et al.* (54), only maxima obtained at least in 12 out of the 18 participants are shown and the crossed yellow square indicates a maximum in the left ventromedial parahippocampus projected onto the lateral surface of the hemisphere.

meaning of the sentence. The right anterior superior temporal sulcus showed higher activity during the voice-recognition task than during the verbal-content task (Figure 2, red squares). The left posterior middle temporal gyrus, in turn, was more active during the content task than during the voice-recognition task, whereas the middle regions of the superior temporal and middle temporal gyri were activated by both tasks in comparison with a control task where the listeners were to recognize a particular speech-envelope noise pattern among other noise patterns. In a further study, von Kriegstein and Giraud (77) found that the right anterior superior temporal sulcus was equally activated both during recognition of a particular familiar voice among other familiar voices and during recognition of a particular unfamiliar voice among other unfamiliar voices. However, the right posterior superior temporal sulcus displayed stronger activity when the voices were unfamiliar than when they were familiar.

5. BRAIN ACTIVITY ASSOCIATED WITH SELECTIVE LISTENING TO SPEECH

One might assume that selective listening to continuous speech would enhance brain activity in the auditory and prefrontal cortices like selective listening to non-speech sounds (42-44) and especially the activity in the left-hemisphere areas involved in speech processing, since attention to discrete syllables and words has been found to activate predominantly the left auditory cortex (28,29,64,78,79). However, brain activity associated with selective attention to continuous speech might differ from brain activity during selective attention to discrete auditory stimuli, such as tones, syllables, and words (42-44,64,78-80), because human listeners are highly experienced in attending to continuous speech. Such experience might explain the very early onset of the attention effects observed in the few ERP studies on selective listening to continuous speech (81,82). In these studies, ERPs elicited by words in the attended and unattended speech messages delivered to the opposite ears differed from each other as early as 40-50 ms from word onset. However, the early onset of these attention effects might also be related to the high rate of sounds in continuous speech, since ERP and MEG studies on selective attention to tones have shown that the latencies of the attention effects depend on the rate of the attended sounds (23,24,29,36). There may also be differences in the effects of selective attention to continuous speech on brain activity depending on whether the attended speech differs in location or in pitch (e.g., male vs. female voice) from concurrent irrelevant speech due to the segregated processing of spatial and non-spatial auditory information reviewed above. However, in both kinds of selective listening conditions, the listener may also use semantic and prosodic (e.g., rhythm and intonation) features of the attended speech in order to separate it from the concurrent irrelevant message(s). Finally, when selective attention to speech is based on speaker identity, one might expect enhanced attention-related activity especially in the superior temporal sulcus of the right hemisphere which is also activated during voice recognition (75-77).

In two recent PET studies, Alho and colleagues (83,84) investigated effects of attention on brain activity during selective listening to speech. In their first study (83), two messages spoken by the same female voice were delivered to opposite ears together with a third written message running on a screen in front of the participant, whose task was to attend to one message at a time in order to be able to recall it afterwards and to ignore the other two messages. During selective listening to the message delivered to the right ear, activity increased bilaterally in the auditory cortical areas of the superior temporal gyrus and sulcus in comparison with the reading task. However, no enhanced activity was observed in the left auditory cortex during selective listening to the left-ear message in relation to the reading task, this comparison revealing enhanced activity only in the right superior temporal gyrus and sulcus. The prominent activation in the right superior temporal areas in both listening conditions was reasoned to be caused either by enhanced processing of the prosodic

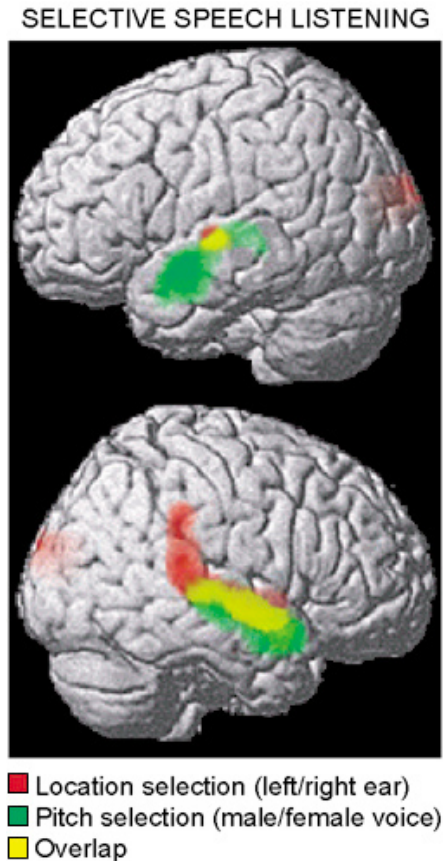


Figure 3. Brain areas showing enhanced activity during selective listening to speech on the basis of either its location (left or right ear) or its pitch (male or female voice). Red indicates brain areas with significantly enhanced activity ($p < .05$, corrected for multiple comparisons) during selective listening to the left-ear or right-ear message (data averaged across the two conditions) in relation to a reading task with two irrelevant speech messages. Green indicates areas with significantly enhanced activity during selective listening to the male or female voice (data averaged across the two conditions) in relation to a reading task with two irrelevant speech messages. Yellow indicates an overlap of these attention effects. Data from Alho *et al.* (83,84).

features in the attended speech, such as the rhythm and intonation, as prosodic features are processed predominantly in the right auditory cortex (85-88), or by enhanced activity in the right superior temporal sulcus areas involved in voice processing (75-77). The lack of increased activity in the left hemisphere during selective listening to the left-ear message than during reading, in turn, was thought to be due to activation in the left auditory cortex, including Wernicke's area, even during reading (69,70). This interpretation was supported by comparisons of the two selective listening conditions with an additional control condition in which the participants were to discriminate meaningless letter strings running on the screen and to ignore the two concurrent speech messages. These comparisons revealed bilateral activity in the

superior temporal cortices. More prominent left-hemisphere activity during listening to the right-ear message than during listening to the left-ear message was probably due to the fact that during spatial auditory attention tasks the attention effects are sometimes observed to be larger in the hemisphere contralateral to the attended direction (23,36,44).

The second PET study on speech listening by Alho *et al.* (84) was similar to the first one except that the two concurrent speech messages were now spoken by different voices (male and female) and they were delivered binaurally and therefore perceived as originating from the same location. Comparison of brain activity during selective listening to either speech message with brain activity in a reading condition where the two spoken messages were to be ignored showed bilateral activations along the superior temporal sulci. These activations were thought to be associated with enhanced processing of the attended voice (75-77). However, it was acknowledged that the activation in the left hemisphere might be partly caused by enhanced activity in the anterior areas of the left superior temporal sulcus involved in speech comprehension (73,74). Furthermore, the activity in the right hemisphere might be partly due to enhanced processing of the prosody in the attended speech (85-88).

With regard to segregated spatial and non-spatial auditory processing discussed above, it is of interest to compare the results from the two selective speech-listening studies of Alho and colleagues (83,84). Figure 3 superimposes the brain areas showing significantly enhanced activity during selective listening to speech delivered to the left ear or to speech delivered to the right and the areas showing significantly enhanced activity during selective listening to the male or female voice. As seen in Figure 3, in the right hemisphere, the activation associated with spatial (left vs. right ear) selective listening to speech appeared to extend to more posterior temporo-parietal areas than activity associated with non-spatial (male vs. female voice) selective listening to speech. The activity associated with non-spatial selective listening, in turn, tended to occupy in both hemispheres superior temporal sulcus areas anterior and inferior to those activated by spatial selective listening to speech.

Thus, these results (83,84) suggest that somewhat different cortical areas may be involved in spatial and non-spatial selective listening to speech. These results are in line with the studies by Alain *et al.* (53) and Maeder *et al.* (54) showing segregation of auditory-cortex areas involved in spatial and non-spatial auditory working memory and discrimination. As seen in Figure 1, the attention-related activations observed by Alho *et al.* (83) during spatial selective listening to speech have their maxima in the right hemisphere at or in the vicinity of the posterior temporal and parietal activations related to spatial auditory processing according to the results of Alain *et al.* (53) and Maeder *et al.* (54). As seen in Figure 2, the attention-related activations observed by Alho *et al.* (84) during selective listening to either a male or female speaker, in turn, have their maxima in the anterior superior temporal

sulcus bilaterally. The right hemisphere maximum is in the vicinity of the anterior temporal activations related to recognition of complex sounds according to Maeder *et al.* (54) and to voice recognition according to von Kriegstein *et al.* (76). Moreover, the activation peaks in the middle part of the left and right superior temporal cortex observed by Alho *et al.* (84) are in the vicinity of activations associated with voice processing according to Belin *et al.* (75), with working memory for pitch according to Alain *et al.* (53), and with sound recognition according to Maeder *et al.* (54).

Finally, although previous studies showed enhanced activity in the prefrontal cortex during selective listening to non-speech sounds (42-44), no prefrontal activity was observed in the selective speech-listening studies by Alho and colleagues (83,84). This lack of prefrontal activity might be due to the fact that the prefrontal areas activated by selective listening to speech were also activated in the comparison conditions by reading or letter-string discrimination. However, this appears not to be the case, since in the second selective speech listening study by Alho and colleagues (84), comparison of brain activity during speech listening or reading with that recorded during an additional resting condition without any spoken or written messages showed no prefrontal activations. Therefore it appears that speech listening and reading depend less on prefrontal control functions than do other types of tasks demanding selective attention.

6. SUMMARY AND CONCLUSIONS

In summary, brain imaging studies suggest that partly different cortical areas are involved in spatial and non-spatial auditory processing (23,24,75,76) including attentional selection of speech for further processing on the basis of its locus of origin or speaker identity (83,84). As seen in Figures 1 and 2, most of these studies found activity associated with spatial auditory processing in posterior temporal and temporo-parietal regions especially in the right hemisphere, while activations related to non-spatial voice or sound recognition are observed bilaterally along the superior temporal gyrus and sulcus. As seen in these figures, the distances between the activation peaks found in different studies on spatial auditory processing (Figure 1) are relatively long, as well as the distances between activation peaks found in voice/sound recognition studies (Figure 2). However, it should be borne in mind that these activation peaks do not indicate the total activated cortical areas. For example, although the activation peaks observed in the superior temporal cortex by Alho and colleagues (83,84) are in different locations for spatial and non-spatial selective listening (see the green squares in figures 1 and 2) there is an overlap between the superior temporal areas showing significant activations in these studies (Figure 3). Furthermore, in PET and fMRI studies reviewed here, the imaging data were averaged across the participants after normalization of individual brains in a common anatomical space, i.e., the Talairach space (89). However, there is substantial between-participant variation in the coordinates of different brain structures even after such normalization procedures (90). Therefore, the loci of activation peaks seen in imaging data averaged across individual brains do

not necessarily correspond to the loci of activation peaks in different individuals. Yet, the activation peaks in the data averaged across participants still suggest the brain areas where individual activations overlap. Finally, the spatial resolution in the PET data of Alho *et al.* (83,84) is only 1-2 cm which complicates anatomical interpretation of their results.

Bearing the aforementioned methodological limitations in mind, the results from PET studies on selective listening to natural speech by Alho and colleagues (83,84) still suggest that selective attention to speech enhances processing of speech in the auditory cortical areas along the superior temporal gyrus and sulcus. This activity extends in the right hemisphere to the parietal cortex, if the relevant and irrelevant spoken messages are separated spatially, and to the anterior superior temporal sulcus areas binaurally, if the relevant speech differs from the irrelevant speech only in terms of the different voices of the speakers. These results are in line with studies showing that different cortical areas are involved in spatial and non-spatial auditory processing (48-53). Thus the right temporo-parietal areas appear to have an important role in the spatial selection of an attended speech message for further processing, whereas the left and right anterior superior temporal sulcus areas are presumably involved in the selection of an attended speech message on the basis of the speaker's voice. However, there is a substantial overlap between the areas in the middle parts of the superior temporal cortices showing enhanced activity during spatial and non-spatial selective listening to speech. These overlapping activations might be caused by effects of attention on initial auditory processing observed also for non-speech sounds (40-44), as well as by effects of attention on left-temporal activity associated with speech comprehension (73-74), right-temporal activity related to processing of prosody (85-88), and bilateral temporal activity associated with voice recognition (75-77). Finally, the studies by Alho *et al.* (83,84) revealed no prefrontal activity during selective attention to speech, while such activity is observed, for example, during selective listening to non-speech sounds (42-44). The lack of such prefrontal activations during selective listening to speech might be due to the fact that speech listening even in a distracting environment is an over-learned skill rehearsed in everyday life. Therefore, perhaps selective listening to speech has become less dependent on the prefrontal control functions than other kinds of attentional tasks.

7. ACKNOWLEDGMENTS

Preparation of this review was supported by the Academy of Finland (grant #201660).

8. REFERENCES

1. Cherry EC: Some experiments on the recognition of speech with one and with two ears. *J Acoust Soc Am* 25, 975-979 (1953)
2. Spieth W, JF Curtis & JC Webster: Responding to one of two simultaneous messages. *J Acoust Soc Am* 26, 391-396 (1954)

3. Treisman A: Verbal cues, language and meaning in selective attention. *Am J Psychol* 77, 206-219 (1964)
4. Pashler HE: The psychology of attention. The MIT Press, Boston (1997)
5. Alain C & S R Arnott: Selectively attending to auditory objects. *Front Biosci* 5, 202-212 (2000)
6. Broadbent DE: Perception and communication. Pergamon, NY (1958)
7. Moray NP: Attention in dichotic listening: Affective cues and the influence of instructions. *Q J Exp Psychol* 9, 56-60 (1959)
8. Wood NL & Cowan N: The cocktail party phenomenon revisited: attention and memory in the classic selective listening procedure of Cherry (1953). *J Exp Psychol Gen* 124, 243-262, (1995)
9. Treisman AM: Contextual cues in selective listening. *Q J Exp Psychol* 12, 242-248 (1960)
10. Lewis JL: Semantic processing of unattended messages using dichotic listening *J Exp Psychol* 85, 225-228 (1970)
11. Kahneman D & A Treisman: Changing views of attention and automaticity. In: Varieties of attention. Eds: Parasuraman R & Davies DR, Academic Press, London (1984)
12. Treisman A, R Squire & J Green: Semantic processing in dichotic listening? A replication. *Mem Cogn* 2, 641-646. (1974)
13. Hillyard SA, RF Hink, VL Schwent & TW Picton: Electrical signs of selective attention in the human brain. *Science* 182, 177-180 (1973)
14. Hansen JC & SA Hillyard: Endogenous brain potentials associated with selective auditory attention. *Electroencephal clin Neurophysiol* 49, 277-290 (1980)
15. Hansen JC & SA Hillyard: Selective attention to multidimensional auditory stimuli. *J Exp Psychol Hum Perc Perf* 9, 1-19 (1983)
16. Näätänen R, AWK Gaillard & S Mäntysalo: Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol* 42, 313-329 (1978)
17. Giard MH, F Perrin, J Pernier & F Peronnet: Several attention-related waveforms in auditory areas: A topographic study. *Electroencephal clin Neurophysiol* 69, 371-384 (1988)
18. Alaini M, MH Giard, JF Echallier, J Pernier: Selective auditory attention effects in tonotopically organized cortical areas: A topographic ERP study. *Hum Brain Mapp* 2, 159-169 (1994)
19. Giard MH, A Fort, Y Mouchetaint-Rostain & J Pernier: Neurophysiological mechanisms of auditory selective attention in humans. *Front Biosci* 5, 84-94 (2000)
20. Kaufman, L & SJ Williamson: Recent developments in neuromagnetism. In: Evoked Potentials III. Eds: Barber C & Blum T, Butterworths, Boston (1987).
21. Hari, R, M Hämäläinen, E Kaukoranta, J Mäkelä, SL Joutsiniemi & J Tiihonen: Selective listening modifies activity of the human auditory cortex. *Exp Brain Res* 74, 463-470 (1989)
22. Arthur, DL, PS Lewis, PA Medvick & A Flynn: A neuromagnetic study of selective auditory attention. *Electroencephal clin Neurophysiol* 78, 348-360 (1991)
23. Rif J, R Hari, MS Hämäläinen & M Sams: Auditory attention affects two different areas in the human supratemporal cortex, *Electroencephal clin Neurophysiol* 79, 464-472 (1991)
24. Woldorff MG, CG Gallen, SA Hampson, SA Hillyard, C Pantev, D Sobel & FE Bloom: Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc Natl Acad Sci USA* 90, 8722-8726 (1993)
25. Knight RT, SA Hillyard, D Woods & HJ Neville: The effects of frontal cortex lesions on event-related potentials during auditory selective attention. *Electroencephal clin Neurophysiol* 52, 571-582 (1981)
26. Näätänen R, & PT Michie: Early selective-attention effect on the evoked potential. A critical review and reinterpretation. *Biol Psychol* 8, 81-136 (1979)
27. Näätänen, R: Processing negativity: An evoked-potential reflection of selective attention. *Psychol Bull* 92, 605-640 (1982)
28. Näätänen, R: The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behav Brain Sci* 13, 201-288 (1990)
29. Näätänen R: Attention and brain function. Lawrence Erlbaum Associates, Hillsdale, NJ (1992)
30. Alho K, J Lavikainen, K Reinikainen, M Sams & R Näätänen: Event-related brain potentials in selective listening to frequent and rare stimuli. *Psychophysiol* 27, 73-86 (1990)
31. Alho K., M Sams, P Paavilainen & R Näätänen: Small pitch separation and the selective-attention effect on the ERP. *Psychophysiol* 23, 189-197 (1986)
32. Alho K., N Donauer, P Paavilainen, K Reinikainen, M Sams & R Näätänen: Stimulus selection during auditory spatial attention as expressed by event-related potentials. *Biol Psychol* 24, 153-162 (1987)
33. Alho K., K Töttölä, K Reinikainen, M Sams & R Näätänen: Brain mechanisms of selective listening reflected by event-related potentials. *Electroencephal clinical Neurophysiol* 68, 458-470 (1987)
34. Michie PT., N Solowij, JM Crawford & L C Glue: The effects of between-source discriminability on attended and unattended auditory ERPs. *Psychophysiol* 30, 205-220 (1993)
35. Alho K., DL Woods & A Algazi: Processing of auditory stimuli during auditory and visual attention as revealed by event-related potentials. *Psychophysiol* 31, 469-479 (1994)
36. Woldorff M & SA Hillyard: Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephal clin Neurophysiol* 79, 170-191 (1991)
37. Alho K: Selective attention in auditory processing as revealed by event-related brain potentials. *Psychophysiol* 29, 247-263 (1992)
38. Grady CL, JW Van Meter, JM Maisog, P Pietrini, J Krasuski & JP Rauschecker Attention-related modulation of activity in primary and secondary auditory cortex. *NeuroReport* 8, 2511-2516 (1997)
39. Jäncke L, S Mirzazade & NJ Shah: Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci Lett* 266, 125-128 (1999)
40. Petkov CI, X Kang, K Alho, O Bertrand, EW Yund & DL Woods: Attentional modulation of human auditory cortex. *Nat Neurosci* 7, 658-663 (2004)
41. Rinne T, J Pekkola, A Degerman, T Autti, IP Jääskeläinen, M Sams & K Alho: Modulation of auditory

- cortex activation by sound presentation rate and attention. *Hum Brain Mapp*, in press (2005)
42. Tzourio N, F El Massioui, F Crivello, M Joliot, B Renault & B Mazoyer: Functional anatomy of auditory attention studied with PET. *NeuroImage* 5, 63-77 (1997)
43. Zatorre RJ, TA Mondor & AC Evans: Auditory attention to space and frequency activates similar cerebral systems. *NeuroImage* 10, 544-554 (1999)
44. Alho K, SV Medvedev, SV Pakhomov, MS Roudas, M Tervaniemi, K Reinikainen & R Näätänen: Selective tuning of the left and right auditory cortices during spatially directed attention. *Cogn Brain Res* 7, 335-342 (1999)
45. Mesulam MM: Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol* 28, 597-613 (1990)
46. Heilman KM & E Valenstein: Auditory neglect in man. *Arch Neurol* 26, 32-35 (1972)
47. Alho K, M Sams, P Paavilainen, K Reinikainen & R Näätänen: Event-related brain potentials reflecting processing of relevant and irrelevant stimuli during selective listening. *Psychophysiol* 26, 514-528 (1989)
48. Woods DL, K Alho & A Algazi: Stages of auditory feature conjunction: An event-related brain potential study. *J Exp Psychol Hum Perc Perf* 20, 81-94 (1994)
49. Woods DL & C Alain: Conjoining three auditory features: An event-related brain potential study. *J Cogn Neurosci* 13, 492-509 (2001)
50. Hackett TA, I Stepniewska, JH Kaas: Prefrontal connections of the parabelt auditory cortex in macaque monkeys. *Brain Res* 817, 45-58 (1999)
51. Rauschecker JP & B Tian: Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc Natl Acad Sci USA* 97, 11800-11806 (2000)
52. Tian B, D Reser, A Durham, A Kustov & JP Rauschecker: Functional specialization in rhesus monkey auditory cortex. *Science* 292, 290-293 (2001)
53. Alain C, SR Arnott, S Hevenor, S Graham & CL Grady: "What" and "where" in the human auditory system. *Proc Natl Acad Sci USA* 98, 12301-12306 (2001)
54. Maeder PP, RA Meuli, M Adriani, A Bellmann, E Fornari, J P Thiran, A Pittet & S Clarke: Distinct pathways involved in sound recognition and localization: a human fMRI study. *NeuroImage* 14, 802-816 (2001)
55. Mazziotta JC, ME Phelps, RE Carson & DE Kuhl: Tomographic mapping of human cerebral metabolism: auditory stimulation. *Neurology* 32, 921-937 (1982)
56. Mazoyer BM, S Dehaene, N Tzourio, V Frak, N Murayama, L Cohen, O Levrier, G Salamon, A Syrota & J Mehler: The cortical representation of speech. *J Cogn Neurosci* 5, 467-479 (1993)
57. Schlosser MJ, N Aoyagi, RK Fulbright, J C Gore & G McCarthy: Functional MRI studies of auditory comprehension. *Hum Brain Mapp* 6, 1-13 (1998)
58. Caplan D, N Alpert & G. Waters: PET studies of syntactic processing with auditory sentence presentation. *NeuroImage* 9 343-351 (1999)
59. Nakai T, K Matsuo, C Kato, M Matsuzawa, T Okada, G H Glover, T Moriya & T Inui: A functional magnetic resonance imaging study of listening comprehension of languages in human at 3 Tesla - comprehension level and activation of the language areas. *Neurosci Lett* 263 33-36 (1999)
60. Broca P: Perte de la parole, ramollissement chronique et destruction partielle du lobe antérieur gauche du cerveau. *Bull Soc Anthropol* 2, 235-238 (1861)
61. Petersen SE, P T Fox, MI Posner, M Mintun & ME Raichle: Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585-589 (1988).
62. Stromswold K, D Caplan, N Alpert & S Rauch, Localization of syntactic comprehension by positron emission tomography. *Brain Lang* 52 452-473 (1996)
63. Just MA, PA Carpenter, TA Keller, WF Eddy & KR Thulborn. Brain activation modulated by sentence comprehension. *Science* 274 114-116 (1996)
64. Bavelier D, D Corina, P Jezzard, S Padmanabhan, VP Clark, A Karni, A Prinster, A Braun, A Lalwani, JP Rauschecker, R Turner & H Neville: Sentence reading: A functional MRI study at 4 Tesla. *J Cogn Neurosci* 9, 664-686 (1997)
65. Dapretto M & SY Bookheimer: Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron* 24, 427-432 (1999)
66. Poldrack RA, AD Wagner, MW Prull, JE Desmond, GH Glover & JD Gabrieli: Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage* 10, 15-35 (1999)
67. Wernicke C: Der Aphasische Symptomenkomplex: Eine Psychologische Studie auf Anatomischer Basis. Cohn und Wegert, Breslau (1874)
68. Mazzocchi F & LA Vignolo: Localisation of lesions in aphasia: Clinical-CT scan correlations in stroke patients. *Cortex* 15, 627-653 (1979)
69. Naeser MA, N Helm-Estabrooks, G Haas, S Auerbach & M Srinivasan: Relationship between lesion extent in 'Wernicke's area' on computed tomographic scan and predicting recovery of comprehension in Wernicke's aphasia. *Arch Neurol* 44, 73-82 (1987)
70. Binder JR, JA Frost, T A Hammeke, RW Cox, SM Rao & T. Prieto: Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci* 17 353-362 (1997)
71. Hugdahl K, K Brönnick, S Kyllingsbaek, I Law, A Gade & O B Paulson: Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: a 15O-PET study. *Neuropsychologia* 37, 431-440 (1999)
72. Zatorre RJ, AC Evans, E Meyer & A Gjedde: Lateralization of phonetic and pitch discrimination in speech processing. *Science* 256, 846-849 (1992)
73. Scott SK, CC Blank, S Rosen & RJ Wise, Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400-2406 (2000)
74. Davis MH & IS Johnsrude. Hierarchical processing in spoken language comprehension. *J Neurosci* 23, 3423-3431 (2003)
75. Belin P, RT Zatorre, P Lafaille, P Ahad & B Pike: Voice-selective areas in the human auditory cortex. *Nature* 403 309-312 (2000)
76. von Kriegstein K, E Eger, A Kleinschmidt & AL Giraud: Modulation of neural responses to speech by directing attention to voices or verbal content. *Cogn Brain Res* 17, 48-55 (2003)
77. von Kriegstein K & AL Giraud: Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *NeuroImage* 22, 948-955 (2004)

78. Jäncke L, K Specht, JN Shah & K Hugdahl: Focused attention in a simple dichotic listening task: an fMRI experiment. *Cogn Brain Res* 16, 257-266 (2003)
79. Hugdahl K, T Thomsen, L Ersland, LM Rimol & J Niemi: The effects of attention on speech perception: an fMRI study. *Brain Lang* 85, 37-48 (2003)
80. O'Leary DS, NC Andreasen, RR Hurtig, RD Hichwa, L Watkins, LLB Ponto, M Rogers & PT Kirchner: A positron emission tomography study of binaurally and dichotically presented stimuli: Effects of level of language and directed attention. *Brain Lang* 53, 20-39 (1996)
81. Woods DL, SA Hillyard & JC Hansen: Event-related brain potentials reveal similar mechanisms during selective listening and shadowing. *J Exp Psychol Hum Perc Perf* 10, 761-777 (1984)
82. Teder W, T Kujala & R Näätänen: Selection of speech messages in free-field listening: First measurements of brain responses to real speech. *NeuroReport* 5, 307-309 (1993)
83. Alho K, V A Vorobyev, SV Medvedev, SV Pakhomov, MS Roudas, M Tervaniemi, T van Zuijen & R Näätänen: Hemispheric lateralization of cerebral blood-flow changes during selective listening to dichotically presented continuous speech. *Cogn Brain Res* 17, 201-211 (2003)
84. Alho K, VA Vorobyev, SV Medvedev, SV Pakhomov, MG Starchenko, M Tervaniemi & R Näätänen: Selective attention to human voice enhances brain activity bilaterally in the superior temporal sulcus. *Brain Res*, in press (2005).
85. Benson DF: Aphasia and the lateralization of language. *Cortex* 22, 71-86 (1986)
86. Heilman KM, R Scholes & RT Watson: Auditory affective agnosia. Disturbed comprehension of affective speech. *J Neurol Neurosurg Psychiat* 38, 69-72 (1975)
87. Ross ED: Right hemisphere's role in language, affective behavior and emotion. *Trends Neurosci* 7, 342-346 (1984)
88. Buchanan TW, K Lutz, S Mirzazade, K Specht, NJ Shah, K Zilles & L Jäncke: Recognition of emotional prosody and verbal components of spoken language: An fMRI study. *Cogn Brain Res* 9 227-238 (2000)
89. Talairach J, P Tournoux: Co-planar stereotaxic atlas of the human brain. Georg Thieme Verlag, Stuttgart, Germany (1988)
90. Kang X, O Bertrand, K Alho, EW Yund, TJ Herron & DL Woods: Local landmark-based mapping of human auditory cortex, *NeuroImage* 22 1657-1670 (2004)

Key Words: Selective Attention, Auditory, Speech, ERP, MEG, PET, fMRI, Review

Send correspondence to: Dr. Kimmo Alho, Department of Psychology, PO Box 9, FIN-00014 University of Helsinki, Finland, Tel: 358-9-19129407, Fax: 358-9-19129401, E-mail: kimmo.alho@helsinki.fi

<http://www.bioscience.org/current/vol12.htm>