2 Prior research on blind listeners' auditory abilities

Although the literature on blind individuals' (forensic) speaker identification abilities is quite scarce, a large variety of other auditory abilities has already been investigated thoroughly in blind listeners and sighted controls. In the following section, results of some relevant studies are reported and discussed. Since the identification of a speaker by his or her voice is a complex task, it is important to consider underlying as well as related auditory abilities, which could attribute to a good speaker identification performance, such as speech perception in noise, frequency discrimination or temporal auditory resolution (see section 2.1).

2.1 Auditory abilities

2.1.1 Speech discrimination

In a speech discrimination experiment with blind and sighted participants, Niemeyer and Starlinger found blind listeners at all semantic levels significantly superior to sighted controls. The superiority was most pronounced at the highest semantic level where sentence discrimination with and without competing background noise was investigated (Niemeyer and Starlinger 1981, p. 512). Other researchers could confirm the enhanced ability of blind individuals to discriminate speech sounds in the presence of noise (Rokem and Ahissar 2009, p. 846; Muchnik et al. 1991, p. 22). However, blind individuals were not always found to be superior to sighted controls regarding their speech discrimination ability in silence (cf. Muchnik et al. 1991). Blind children performed even worse than sighted peers in a speech discrimination task in noise (Stankov and Spilsbury 1978).

2.1.2 Auditory attention

In a dichotic listening experiment in which different syllables were presented simultaneously to the listener's left and right ear via headphones, blind listeners reported significantly more correct syllables than sighted controls. Furthermore, the blind gave significantly more correct answers when the listener's attention was directed to one particular ear (Hugdahl et al. 2004, p. 30-31). Blind individ-

A. Braun, *The Speaker Identification Ability of Blind and Sighted Listeners*,
DOI 10.1007/978-3-658-15198-0_2

uals outperformed sighted listeners also in an auditory vigilance task. In this task, signals with low signal-to-noise ratios had to be detected which occurred at irregular time intervals (Hohmann Benedetti and Loeb 1972). Furthermore, it was discovered that the auditory blink, i.e. a masking effect which occurs when two auditory stimuli are presented shortly after each other, appeared to be attenuated in congenitally blind individuals at brief inter-target intervals (Goddard et al. 2004).

Since auditory attention is supposed to play a key role in speaker identification, it is likely that individuals with a higher level of attention will pick up more speaker-specific cues at the encoding stage, i.e. when they listen to a previously unknown voice for the first time. This could lead to a better speaker identification performance later on.

2.1.3 Perception of acoustic details

Hirsch et al. (2011) used a gating paradigm with truncated vowels and observed that congenitally blind individuals were able to perceive rounded vowels in a speech signal earlier than sighted controls. Ménard et al. (2009) carried out a vowel discrimination experiment with different synthetic vowel continua and had blind and sighted subjects indicate whether the second presented vowel in a triad was identical to the first or the third presented vowel. Blind listeners' discrimination scores for the continua /e/ – /ɛ/ and /ɛ/ – /a/ were significantly better compared to the scores of sighted controls. The discrimination performance of blind participants was also better for all other investigated vowel continua (i.e. /i/ – /e/, /i/ – /y/ and /y/ – /u/); however, these differences failed to reach significance (Ménard et al. 2009, p. 1410). In a study in which the discrimination between similar consonants in a foreign language was investigated, blind participants again performed substantially better than sighted controls. The result was marginally significant (cf. Sáez Sáez 2012, p. 49-50). Furthermore, congenitally blind individuals outperformed matched sighted controls in an auditory vowel discrimination task as well as in an auditory emotion discrimination task (Klinge 2011, p. 78).

The latter studies provide strong evidence that blind listeners are able to perceive more subtle details of speech – an ability which could further enhance blind listeners' speaker identification abilities.

2.1.4 Temporal aspects

Another auditory ability which has been investigated in blind as well as sighted individuals is temporal auditory resolution. In the experiment, a short temporal gap had to be detected in one of two – otherwise equal – noise bursts. Blind participants were found to be superior to sighted participants with regard to this task (Muchnik et al. 1991, p. 22; also cf. Sepehrnejad et al. 2011). However, other researchers who investigated temporal auditory resolution and temporal auditory sensitivity in blind and sighted subjects did not find any significant differences between the two listener groups (Weaver and Stevens 2006, p. 3; Goddard et al. 2004, p. 243; Bross and Borenstein 1982, p. 963). It should be mentioned, though, that the number of blind participants in the last two studies was as low as four and five, respectively, which also could explain the non-significant results.

It is known that some blind individuals listen to audio books or the audio output from their screen readers[5] at a highly accelerated playback speed (cf. Röder 2004). Several recent studies show that blind individuals outperform sighted controls in the comprehension of ultra-fast speech whereby some of the blind were even able to understand compressed speech up to 22 syllables per second (Moos and Trouvain 2007; also see Dietrich et al. 2013; cf. Gordon-Salant and Friedman 2011). These results suggest that speech processing in general can be substantially enhanced in blind individuals – most likely due to a training effect.

2.1.5 Pattern recognition

In a pattern recognition experiment in which the auditory substitution of vision was investigated, early blind listeners performed significantly better than blindfolded sighted controls (Arno et al. 2001). In this experiment, participants had to scan visually presented patterns on a screen with an optical device such as a head-worn camera or an optical pen. The captured pixels were recoded acoustically as sinusoidal tones and mapped onto an artificial *acoustic retina* (the x-axis being represented by increasing frequency, the y-axis by harmonicity. Brightness was coded as loudness). Like a real retina, the acoustic retina had a fovea in which the resolution was higher compared to the periphery of the visual/acoustic field. With the acoustic information they received, participants were asked to replicate the visual pattern on the screen with a set of aluminum strips and dots.

Practice alone cannot account for the better performance of blind participants in this study since all blind and sighted listeners had received an equal amount of training before the experiment was carried out.

5 A screen reader is a computer program which converts the screen content into spoken language so that blind and visually impaired users can get access to it (cf. Accessible Tech 2014).

2.1.6 Perception of pitch and loudness

Juurmaa (1967) tested the pitch discrimination abilities of blind and sighted listeners and found the former superior to the latter. He noted, however, that he did not observe any clear associations between test performance and either onset or duration of blindness (Juurmaa 1967, p. 111). More recent studies confirm Juurmaa's result: In a pitch discrimination experiment involving early blind, late blind and sighted listeners, Gougoux et al. (2004) found early blind participants superior to the other two groups (cf. also Rokem and Ahissar 2009, p. 846; Starlinger and Niemeyer 1981, p. 506). In a pitch discrimination task which was used as a separator between two parts of a different experiment, blind and sighted listeners performed equally well. However, it should be noted that the chosen pitch difference in this task was always 50 Hz – a difference which can be detected easily in a frequency range of 900-1050 Hz (cf. Röder and Rösler 2003, p. 32).

Yates et al. (1972) investigated blind and sighted listeners' ability to perceive differences in loudness between two successively presented pure tones and were unable to find any significant performance differences between both listener groups. Starlinger and Niemeyer (1981) did not observe any significant group differences either when testing the difference limen for intensity in blind and sighted listeners; although there was a non-significant trend in favor of the blind group (Starlinger and Niemeyer 1981, p. 506). Juurmaa (1967) found blind participants significantly inferior to sighted controls in a loudness discrimination task.

2.1.7 Absolute threshold measurements

Hohmann Benedetti and Loeb (1972) observed that the group of blind listeners in their experiment had a mean absolute hearing threshold which was 6.8 dB higher than the mean threshold of the sighted group. This result clearly contradicts the sensory compensation hypothesis. However, it is possible that the poorer result of the blind group in this study is – at least to some extent – due to age-related hearing loss since the blind participants ranged from 25-58 years of age while the age range of the sighted participants was considerably smaller, i.e. 18-21 years of age (cf. Hohmann Benedetti and Loeb 1972, p. 11). The authors also tested a second group of blind participants ranging from 19-45 years of age. The mean absolute hearing threshold of this group was comparable to that of the sighted control group. Another – perhaps more plausible – explanation for the higher hearing threshold which was found for the blind participants in the first experiment has to be taken into account. Since the authors did not provide any details on the etiology of blindness of their participants, it cannot be excluded that some of the blind listeners suffered from syndromes which affected not only their vision but also their hearing. See, for example, usher syndrome or CHARGE syndrome (cf. Newton and Moss 2001, p. 28).

Acoustic reflex[6] thresholds were found to be similar in blind and sighted subjects (Starlinger and Niemeyer 1981, p. 507).

2.2 Simple versus complex auditory functions

While the previously reported findings regarding the auditory abilities of blind compared to sighted listeners seem inconsistent at first glance (section 2.1.), a closer examination reveals a pattern. It appears that the degree of complexity of a given auditory task plays a key role in whether the blind are likely to outperform sighted controls or not. Some researchers argue that the superiority of blind listeners only manifests itself in complex auditory tasks which involve higher level auditory functions (Hugdahl et al. 2004, p. 31; cf. also Stankov and Spilsbury 1978, p. 492; Niemeyer and Starlinger 1981, p. 513; Nadig 2009). This would explain why blind individuals are able to outperform sighted controls, for instance, in dichotic listening tasks, speech discrimination in noise and sound localization experiments, but not in experiments in which simple auditory functions are tested (e.g. acoustic reflex thresholds or absolute hearing thresholds). However, this explanation alone cannot account for all observed inconsistencies in auditory research on blind and sighted listeners (e.g. why blind participants outperformed sighted controls in many pitch discrimination tasks but not in loudness discrimination tasks). Röder and Neville (2003) point out that the selection criteria for blind participants as well as different experimental methods may have a considerable impact on the test results. “Blind people constitute a very heterogeneous population with individuals differing in the etiology, degree, onset and duration of blindness as well as their rehabilitation history” (Röder and Neville 2003, p. 255; also cf. Ménard et al. 2009, p. 1407; Kupers and Ptito 2014, p. 41).

A further source of inconsistent results arises from studies in which auditory abilities of blind children are compared to those of age-matched sighted controls. Psychological research shows that blind children can lag behind their sighted peers in the development of certain cognitive abilities. (Hollins 1989 p. 167; cf. Fernández et al. 1988, p. 69; also cf. Röder et al. 2002, p. 935 for a review).

Another aspect which might lead to inconsistent findings concerns the group of sighted listeners. In some studies, sighted controls were blindfolded

6 Acoustic reflex = involuntary contraction of the stapedius muscle in the middle ear in response to high-intensity sound stimuli. (cf. Dobie and Van Hemel 2004, p. 93).

(e.g. Rokem and Ahissar 2009; Wan et al. 2010; Kattner and Ellermeier 2014) while in others, they were not (Pasqualotto et al. 2013). The performance of blind, sighted and blindfolded sighted participants was only investigated in a few studies (e.g. Sáez Sáez 2012; Stevens and Weaver 2005). Blindfolding sighted participants has the advantage of neutralizing any visual cues; however, putting sighted controls in a – for them – very unnatural situation could also be seen as a disadvantage which might have a negative effect on the test results.

2.3 Auditory memory

Apart from the auditory abilities described above, listeners' auditory memory is assumed to play a key role in speaker identification.

According to the modal memory model which was proposed by Atkinson and Shiffrin in 1968, human memory consists of three different parts: a *sensory register* where sensory information resides for a very short period of time, a *short-term store* where (unrehearsed) information can be held for up to 30 seconds and a more or less permanent *long-term store*. It is assumed that selected information is transferred (i.e. "copied") from short-term to long-term store (Atkinson and Shiffrin 1968, p. 90-91). Although this memory model was highly influential at the time, it was later criticized for its oversimplification. According to Atkinson and Shiffrin's model, patients with a defective short-term store should also exhibit an impaired long-term store. Baddeley, however, states that patients exist where this is not the case (cf. Baddeley 2012, p. 5).

Newer memory models assume more than just one unitary short-term and one unitary long-term store and the models are less linear than the model proposed by Atkinson and Shiffrin. Baddeley differentiates between *short-term memory*, i.e. the simple temporary storage of information, and *working memory*, i.e. a combination of storage and manipulation (Baddeley 2012, p. 4). According to the theory of Baddeley and Hitch, working-memory itself consists of a central executive which coordinates and controls three subsystems: the *phonological loop* which contains auditory information, the *visuo-spatial sketch pad* which contains visual and spatial information and the (later added) *episodic buffer* which has a larger storage than the first two subsystems and helps connecting working memory and long-term memory (Baddeley and Hitch 1974; Baddeley 2000; also cf. Goldstein 2011, p. 143).

With regard to long-term memory, Squire and Zola (1996) distinguish *declarative (explicit)* and *non-declarative (implicit)* memory. Declarative memory can be further subdivided into *episodic memory* (the memory for events) and *semantic memory* (the memory for facts). Non-declarative memory consists of *procedural memory* (skills and habits), *priming*, *classical conditioning* and *non-associative learning.* According to Tulving 1983, information which is typically relevant in legal testimony of witnesses is stored in episodic (long-term) memory (cf. Tulving 1983, p. 35). However, also non-declarative memory can play a role (cf. Smith and Kosslyn 2007, p. 235). The question whether blind and sighted listeners differ in their short-term, long-term or working memory skills is addressed in the following section.

2.3.1 Short-term memory and working memory

Juurmaa (1967) found blind listeners slightly superior to sighted controls in a short-term memory experiment in which pairs of words had to be remembered. Interestingly, the superiority was more pronounced when the word pairs were not related in any meaningful way. Juurmaa concludes that "…wholly mechanic, immediate memory based on the sense of hearing is better developed in the blind than in the partially sighted and the seeing" (Juurmaa 1967, p. 110). There was a trend for early blind individuals to perform better than late blind individuals and a slight positive correlation was found between task performance and the number of years spent in blindness.

In memory experiments carried out by Rokem and Ahissar (2009), congenitally blind individuals showed a significantly greater forward digit span as well as a significantly greater verbal span for pseudo-words than sighted listeners. However, no significant differences occurred between both listener groups when the digits had to be recalled backwards (digit span backwards). The authors conclude that "…while blind individuals could hold more items in their short-term memory, they had no such benefit when asked to manipulate these elements" (Rokem and Ahissar 2009, p. 845).

Congenitally blind 10-year-old children outperformed matched sighted controls on three different short-term memory tasks (digit span forward, remembering 15 words, learning names) as well as on both given working memory tasks. In the first one, participants had to recall a series of digits in reversed order, in the second one, they had to listen to two sentences, recall the last words of both sentences and tell whether the sentences contain true or false statements (Withagen et al. 2013, p. 2164). With the help of an extensive test battery designed to compare several auditory and cognitive abilities of blind and sighted children, Stankov and Spilsbury discovered that a better memory for tones was largely responsible for the advantage of the blind group (cf. Stankov and Spilsbury 1978, p. 500). However, not all studies found enhanced short-term memory abilities in blind compared to sighted children. In Fernández et al. (1988, p. 71), child listeners had to remember aurally presented letters while listening to competing speech sounds. Blind and sighted children performed equally well in this task.

Raz et al. (2007) found congenitally blind participants superior to sighted controls in an item memory task in which subjects were asked to recall as many words from a list as possible. Furthermore, the blind outperformed the sighted in a serial memory task in which the words of a list as well as their exact serial position had to be remembered (Raz et al. 2007, p. 1129). In an experiment on auditory working memory, it was found that irrelevant sounds (i.e. noise or speech) interfere with serial word recall in sighted but not in blind individuals (Kattner and Ellermeier 2014, p. 2212).

Cattaneo and Vecchi conclude that "…it is likely that the short-term memory advantage of blind individuals results from better stimulus encoding, rather than from superiority at subsequent processing stages" (Cattaneo and Vecci 2011, p. 28; also see Rokem and Ahissar 2009).

If the latter is true, blind individuals could have a further advantage in speaker identification tasks since their enhanced short-term memory would allow them to encode more speaker-specific cues when they are exposed to a previously unknown speaker.

2.3.2 Long-term memory

A number of studies provide evidence that also long-term memory is enhanced in blind individuals compared to sighted controls. Röder and Rösler (2003) investigated the use of two different encoding strategies in a long-term memory experiment with congenitally blind, late blind and sighted listeners. For semantic encoding, listeners had to name aurally presented environmental sounds. For physical encoding, listeners had to rate the acoustic quality of the heard sounds on a 5-point scale ranging from *harsh* to *soft*. All stimuli were different, but some were conceptually highly similar, e.g. the bark of a dog and the bark of another dog (Röder and Rösler 2003, p. 29-30). In a second session, listeners from both groups were provided with a larger set of environmental sounds and had to indicate which of the sounds they had already heard in the first part of the experiment (Röder and Rösler 2003, p. 30). The results show that congenitally blind listeners performed better than late blind and sighted listeners; however, only the difference between the congenitally blind and the sighted group reached statistical significance. When matched for age, the difference between the late blind and the sighted group also became significant (Röder and Rösler 2003, p. 33). Within all listener groups, semantic encoding yielded significantly better results than physical encoding. A more detailed analysis revealed that congenitally blind listeners had a significantly lower false memory rate than sighted listeners when a physical encoding strategy was used (Röder and Rösler 2003,

p. 31-32). The authors argue that – apart from a better stimulus encoding mechanism – also an improved retrieval monitoring might have contributed to the superiority of blind participants (cf. Röder and Rösler 2003, p. 36). Cobb et al. (1979) carried out two similar experiments in which congenitally blind and sighted participants had to recognize tactile objects and non-speech environmental sounds which they were presented with one week earlier. In both experiments, blind and sighted listeners performed equally well. Pasqualotto et al. (2013) set up a memory experiment with congenitally blind, late blind and sighted listeners in order to investigate the possibility of enhanced auditory memory retrieval in blind participants experimentally. All participants were presented with several words which were semantically related to a "lure" word that was not included in the list (Pasqualotto et al. 2013, p. 162). The results show that congenitally blind participants could not only recall more words but also had lower false memory rates with regard to the lure word than late blind and sighted participants (Pasqualotto et al. 2013, p. 164). High hit rates in combination with low false memory rates are also important in speaker identification tasks.

Perleth and Effinger tested incidental memory in blind, partially sighted and sighted participants with the help of a 30-minute mystery audio drama. The audio drama was immediately followed by a questionnaire in which the listeners were asked about specific details of the story. Some of the participants were retested after about eight weeks (Perleth and Effinger 2001, p. 131). The results indicate that blind listeners were able to remember significantly more details of the story than the partially sighted and the sighted group. This was true for both of the tested time intervals; however, the number of subtests in which the blind were found superior to the sighted was smaller after eight weeks (cf. Perleth and Effinger 2001, p. 143+145).

In summary, the vast majority of memory studies indicate that blind individuals outperform sighted controls in short-term, long-term and working memory tasks. This result will have to be considered when the hypotheses for the speaker identification experiment of the present study are formulated (see Chapter 5).

2.4 Physiological and brain imaging studies

Over the last few decades, a number of neuroimaging studies have been published on reorganizational processes following visual deprivation. It was found that occipital brain regions – which are mainly associated with visual processing in sighted people – can also be activated when blind individuals process auditory, tactile or olfactory stimuli (recent reviews: Kupers and Ptito 2014;

Ricciardi et al. 2014; also see Röder et al. 2002; Ortiz et al. 2010). The assumption that occipital cortex activation in blind individuals is merely an epiphenomenon (i.e. activation without a functional role) is considered rather unlikely (cf. Burton 2003). In line with the latter are, for instance, results from TMS (transcranial magnetic stimulation) studies. In those studies, TMS was applied over occipital brain areas while blind and sighted participants performed a Braille reading task (Cohen et al. 1997) or a verbal processing task (Amedi et al. 2004). It was found that TMS reduced the task performance of blind but not sighted participants. These results provide strong evidence for cross-modal plasticity in blind individuals: an annexation of (otherwise idle) visual brain areas by other sensory modalities could explain – at least partially – the superiority of blind listeners over sighted controls that was observed in a variety of non-visual tasks. Gougoux et al. (2009) found that auditory areas show significantly less activation in blind compared to sighted listeners when vocal and non-vocal stimuli are processed. Comparing the corpus callosum (which connects both brain hemispheres) of congenitally blind and sighted individuals revealed that the isthmus of the corpus callosum, which contains fibers that carry auditory information, is significantly enlarged in congenitally blind individuals (cf. Tomaiuolo et al. 2014).

A study on resting-state functional connectivity shows that blind individuals appear to have weaker functional connectivity within the extra-striate visual cortex as well as between visual and non-visual sensory networks than sighted controls; however, functional connections between the visual cortex and cognitive control networks (e.g. memory, attention, task-switching) are much stronger in the blind than in the sighted (Burton et al. 2014). In a similar study, it was found that although functional connectivity within the occipital cortex is reduced, connections between the occipital cortex and frontal language cortices are stronger in early blind participants compared to sighted controls (Liu et al. 2007). Elbert et al. (2002) showed in an MEG[7]-study that the tonotopic map in the auditory cortex of blind individuals is 1.84 times larger compared to the tonotopic map of sighted individuals (Elbert et al. 2002, p. 9942). In a study on spectral and temporal neural encoding of speech and clicks at the subcortical level, congenitally blind compared to sighted individuals showed frequently shorter latencies and higher amplitudes of auditory brainstem responses (ABR) to (artificial) speech stimuli. No such pattern was found for click ABR. A detailed analysis revealed that the blind gained better results in source as well as filter classes of speech ABR. "It is possible that these [congenitally blind] sub-

7 "Magnetoencephalography (MEG) is an imaging technique used to measure the magnetic fields produced by electrical activity in the brain via extremely sensitive devices known as SQUIDs" (FMRIB 2015).

jects had enhanced neural representation of vocal cord vibrations, better neural synchronization, and faster response to neural encoding of the onset and offset parts of speech stimuli at the brainstem level" (Jafari and Malayeri 2014, p. 407).

Further evidence suggesting that the superiority of blind listeners reveals itself already in early perceptual processing stages comes from a study on auditory perceptual consolidation. In an auditory backward masking task, blind and sighted listeners had to indicate whether two tone-pair stimuli (which were followed by a mask) were the same or different. The performance of blind individuals was unaffected by the mask at all tested mask delays except when the mask was presented simultaneously with the second tone-pair stimulus. Sighted listeners, however, needed a mask delay of 160 ms in order to perform equally well as blind listeners. Interestingly, no performance differences between the blind and the sighted occurred in a single tone auditory backward masking task (Stevens and Weaver 2005). The results are in line with the conclusions drawn from behavioral studies (see section 2.2.): the superiority of blind listeners manifests itself in complex rather than simple auditory tasks.

2.5 Speaker identification and discrimination abilities

So far, not many studies have focused on (forensic) speaker recognition[8] abilities of blind individuals, and the methodological approaches of the existing studies are different: some researchers investigated *speaker discrimination* abilities (i.e. participants were just asked to judge voices on a same/different basis) whereas other researchers tested blind and sighted listeners' *speaker identification* performance with the help of voice lineups, i.e. participants had to pick a previously heard target voice from a set of similar sounding voices. Kreiman and Papcun (1991) compared sighted listeners' results from speaker discrimination and speaker identification tasks and found that the overall test performance did not differ between the two tasks. A more detailed analysis, however, revealed that listeners' hit rate (i.e. correct identifications) and false alarm rate (i.e. incorrect identifications) both were significantly higher in the speaker discrimination task

8 Speaker recognition is a hyperonym of speaker identification and speaker verification (Becker 2012, p. 16). Definitions according to a forensic phonetician: Speaker identification: "…an utterance from an unknown speaker has to be attributed, or not, to one of a population of known speakers for whom reference samples are available." (Nolan 2009, p. 9). Speaker verification: "…an identity claim by an individual is accepted or rejected by comparing a sample of his speech against a stored reference sample spoken by the individual whose identity he is claiming, and making a decision on the basis of a predetermined similarity threshold" (Nolan 2009, p. 8). Note that engineers and phoneticians define of the aforementioned terms differently (see Becker 2012, p. 16-21 for a discussion; also cf. Nolan 2009, p. 8-10; Gfrörer 2014, p. Rn. 3).

compared to the speaker identification task (hit rate 25% higher; false alarm rate 14% higher). Both tasks yielded thus somewhat different results.

Further differences in studies which investigated blind listeners' speaker recognition abilities concern the selection criteria of blind participants (e.g. age, onset of blindness, etiology of blindness), the kind and quality of the used voice samples (e.g. read vs. spontaneous speech samples, sustained vowels vs. sentences, telephone recordings vs. high quality recordings) and the type of memory which has been tested (incidental vs. intentional memory; short-term vs. long-term memory). An overview of previous research on the speaker recognition abilities of blind compared to sighted listeners is given in Table 1 on page 45.

Note that all previous studies suffer from at least one of the following limitations which make their results less applicable to forensic phonetics: a) speaker discrimination ability instead of speaker identification ability was investigated, b) the first exposure to the target voice was immediately followed by the speaker identification test and c) the sample of (blind) participants was very small.

2.5.1 Behavioral studies on blind listeners' speaker recognition abilities

Bull et al. (1983) are assumed to be the first who carried out a voice lineup experiment with blind and sighted listeners. The voice lineups consisted of 5, 7 or 9 voices and listeners were informed that the voice of the respective target speaker was always present in the lineup (closed-set). In this forced-choice experiment, blind listeners performed significantly better than sighted controls, i.e. the blind were significantly more accurate in picking the respective target voice from the lineup. However, since a target voice was always included in the lineups, listeners' hit rates (i.e. correct identifications) and false alarm rates (i.e. how often a distractor speaker was mistaken for the target) could not be assessed separately from each other. Furthermore, the researchers were unable to find any significant correlations between blind listeners' speaker identification performance and the age at onset of blindness, the number of years living with blindness, the degree of blindness or listeners' IQ scores. A subgroup of blind participants who had received special musical training as piano tuners performed equally well as blind listeners without such training.

Elaad et al. (1998) presented (clarity enhanced) telephone quality voice lineups with voice samples of 2-6 different speakers to blind and sighted lay listeners as well as to three voice identification experts. Sixteen of the lineups were target-present lineups and one lineup was a target-absent lineup which did not include a sample of the target speaker's voice. All lineups were presented immediately after listeners had been exposed to the respective target voice in telephone quality. This experiment was more realistic with regard to forensic phonetics since listeners were cautioned that the target voice may or may not be present in the lineup (open test). The voice identification experts outperformed both groups of lay listeners (i.e. blind and sighted), which – compared to each other – performed equally well in the speaker identification task. Blind lay listeners were, however, less confident than sighted lay listeners in their decisions.

Results from a more recent study indicate that blind participants outperform sighted controls in a speaker identification experiment only under certain acoustic conditions. Multi-presentation voice lineups (in which the voice of the target speaker occurred several times among multiple voice samples from similar sounding speakers) were presented to blind and sighted listeners about one week after they had been exposed to a high quality recording of the target speaker's voice for the first time (familiarization). During the familiarization, listeners were asked to memorize as many speaker-specific cues of the target speaker's voice as possible. Voice lineups were presented about 7-9 days later in studio quality as well as in cell phone quality. Blind listeners outperformed sighted controls only under studio quality conditions. Under cell phone quality conditions, both listener groups performed equally well. The response criterion β[9] did not differ between blind and sighted listeners (Braun 2012).

Winograd et al. (1984) tested voice discrimination abilities of blind and sighted listeners and were unable to find any performance differences between both listener groups. In this experiment, blind and sighted participants had to listen to a study tape which contained messages from 20 speakers. Afterwards, listeners were exposed to a test tape with voice samples from 40 different speakers (including the 20 "old voices" from the study tape) and were asked to make old-new-decisions on the voice samples. All listeners were informed in advance that their voice recognition ability would be tested.

Röder, Wolber and Neville (unpublished) used an incidental memory paradigm in order to set up a voice discrimination experiment. Blind and sighted participants were asked to listen to 44 sentences, which were all spoken by different speakers (i.e. 22 male and 22 female), and had to indicate whether a particular voice would be easy or hard to remember. In the consecutive recognition phase, the previously heard stimuli were intermixed with 44 new voice samples and participants were asked to indicate which of the voices they had already heard before. Blind participants performed significantly better than sighted controls; the response criterion β did not differ between both listener groups. In order to investigate whether blind individuals recognize voices as well as sighted

9 The response criterion β can be seen as an indicator for the proportions of conservative and progressive raters among the listeners. Conservative raters focus on keeping the false alarm rate as low as possible and accept some false negatives (misses) whereas progressive raters focus on keeping the hit rate as high as possible and accept some false positives (false alarms) (cf. Künzel 1990, p. 26).

individuals recognize faces, an independent group of sighted participants performed the same experiment as stated above with the exception that all voice samples were replaced by pictures of the respective speakers. The face recognition performance of sighted individuals significantly exceeded blind listeners' speaker recognition performance[10] (published in excerpts in Röder and Neville 2003). Table 1 gives an overview of all cited studies:

2.5.2 Physiological studies on blind listeners' speaker recognition abilities

Apart from behavioral studies, several brain imaging studies have been carried out in order to investigate voice processing in blind and sighted individuals. Gougoux et al. (2009) presented blind and sighted listeners with vocal and non-vocal acoustic stimuli while they were undergoing a functional magnetic resonance imaging (fMRI) scan. When blood oxygenation level-dependent (BOLD) contrasts of all (vocal + non-vocal) stimuli were compared to baseline, blind listeners showed stronger activation in occipital regions than sighted controls. The opposite pattern was observed in auditory areas in which sighted listeners showed stronger activation than the blind. When the hemodynamic BOLD response to vocal stimuli was compared to the BOLD response elicited by non-vocal stimuli, all participants showed stronger activations for vocal stimuli in bilateral temporal regions – especially along the superior temporal sulcus (STS). A more detailed analysis revealed that congenitally blind compared to late blind and sighted participants had significantly stronger activations in the left STS. Furthermore, the congenitally blind group showed a trend for stronger activation (which was just short of statistical significance) in the bilateral fusiform areas. "This result is in good line with suggestions that voices are 'auditory faces'[11]…" (Gougoux et al. 2009, p. 2973). When the degree of BOLD activation in voice selective areas along the left posterior STS was correlated with participants' scores from an offline performed speaker discrimination experiment, a significant positive correlation was found for the blind, but not for the sighted group. No significant correlations were found between blind listeners' speaker discrimination scores and the onset or duration of blindness.

10 Note that this is a psychological experiment. In a forensic setting, turning a speaker recognition task into a face recognition task simply by using pictures of the respective speakers would be a very dangerous approach because individuals who sound similar do not necessarily look similar.

11 Note that although person recognition by voice and by face have some characteristics in common, there are also large differences between voice and face recognition (see Barsics 2014 for a recent review; cf. also Stevenage et al. 2011).

Study	Method	Time delay	Samples	Quality	Memory	Blind (birth)	Sighted	Result
Bull et al. 1983	voice lineup closed-set	5 seconds	same sentence	tape - tape	intentional	92(n.s.)	72	blind > sighted *sig.*
Elaad et al. 1998	voice lineup open test	subsequent presentation	same text	phone - (clarity enhanced) phone recording	intentional	15(n.s.)	18 3 experts	laymen: blind = sighted experts > blind, sighted *sig.*
Braun 2012	voice lineup multi-presentation	1 week	read speech	studio - studio	intentional	8(7)	22	blind > sighted *sig.*
	voice lineup multi-presentation	1 week	read speech	studio - cell phone	intentional	8(7)	31	blind = sighted
Winograd et al. 1984	discrimination study tape	subsequent presentation	read speech	tape - tape	intentional	12(10)	24	blind = sighted
Röder et al. (unpublished)*	discrimination study tape	subsequent presentation	same sentence	(n.s.)	incidental	18(18)	18	blind > sighted *sig.*

Table 1: Overview of details and different methodological approaches used by the aforementioned behavioral studies on blind listeners' speaker identification and speaker discrimination abilities. Time delay = time interval between the first exposure to the target voice (familiarization) and the recognition phase; Quality = quality of the sound samples used for the familiarization as well as the recognition test; Blind/sighted = number of blind and sighted participants, respectively. The number in brackets indicates the proportion of congenitally blind listeners. Studies on voice discrimination are shaded in gray. n.s. = not specified, sig. = significant difference. *For Röder, Wolber and Neville (unpublished) see excerpts in Röder and Neville 2003.

Hölig et al. (2014a) carried out an fMRI priming experiment and presented congenitally blind and sighted listeners with sets of two successive voice stimuli. The first and the second stimuli came from just one speaker or from two different speakers and listeners were asked to indicate whether the second voice sample was produced by an old or a young speaker. Congenitally blind individuals showed a significantly stronger mean activation in bilateral occipital regions than sighted controls while listening to the vocal stimuli. When hemodynamic responses to person-incongruent stimulus pairs were compared to hemodynamic responses to person-congruent trials, congenitally blind listeners had stronger activations than sighted listeners in the right anterior fusiform gyrus. At the same time, sighted listeners showed stronger activations than congenitally blind listeners in the right posterior STS. In an offline performed voice training phase prior to the main experiment described above, participants had to learn voice-name associations for all voice stimuli which were later used in the fMRI study. Congenitally blind participants were able to learn the voices much faster than sighted controls and also achieved significantly better results in an offline performed speaker identification test than the sighted. In a voice matching task (which had also been performed outside the scanner), congenitally blind and sighted listeners performed equally well.

In a similar follow-up study with late blind and sighted participants, it was found that also late blind participants show significantly stronger activations in the right anterior fusiform gyrus to person-incongruent stimuli compared to person-congruent stimuli. In the offline performed behavioral tests, results of late blind and sighted participants were generally similar. However, late blind listeners learned the voices significantly faster than sighted listeners (Hölig et al. 2014b).

The same priming paradigm employed in the two fMRI studies by Hölig et al. (2014a and 2014b) had already been used two years earlier in an EEG study by Föcker et al. (2012). Also here, congenitally blind listeners were able to learn the voices faster than sighted controls. Furthermore, the blind outperformed the sighted in both speaker identification tasks; however, no significant performance differences were observed in the voice matching task between blind and sighted listeners. In the main experiment in which event-related potentials (ERPs) were recorded, congenitally blind (but not sighted) participants showed a significantly enhanced negativity 100-160 ms after the stimulus onset of the second stimulus of person-incongruent trials compared to person-congruent trials.

2.6 Summary

Brain imaging studies show that the brains of (congenitally) blind individuals adapt to the lack of vision by undergoing substantial functional reorganizational changes.

Although blind listeners were found to be superior to sighted controls in some experiments on human speaker identification or speaker discrimination, other studies in the area did not report any significant performance differences between blind and sighted listeners (cf. sections 2.5.1. and 2.5.2.). It remains to be shown whether the results of blind listeners differ from those of the sighted in a speaker identification experiment which adheres to forensic phonetic guidelines for voice lineups.

http://www.springer.com/978-3-658-15197-3

The Speaker Identification Ability of Blind and Sighted Listeners
An Empirical Investigation
Braun, A.
2016, XXVI, 133 p. 19 illus., Softcover
ISBN: 978-3-658-15197-3