JACEK KUDERA
Saarland University
jacek.kudera@uni-saarland.de
ORCID 0000-0003-3678-1067

INFLUENCE OF SENSORINEURAL HEARING LOSS ON FAMILIAR SPEAKER RECOGNITION: A FORENSIC PERSPECTIVE ON AGEING

Abstract

The main goal of this research was to discover the influence of high frequency sensorineural hearing loss on familiar speaker recognition during earwitnessing line-ups. The secondary objectives were to estimate the influence of familiarity with voices of the suspects on performance in the auditory speaker recognition test, and to correlate the results with forensically important factors such as a confidence scale from the line-up markings. The recordings from the line-up sessions were low-pass filtered to ensure an equal degree of signal distortion for all subjects and imitate the moderate, severe and profound hearing loss conditions. The results show that the correlation between mimicked hearing impairment and ability to identify a familiar speaker is statistically significant. It was observed that higher degree of signal distortion caused lower accuracy of recognition. Interestingly, it was reported that higher levels of familiarity and exposure to speakers' voices had a negative effect on speaker identification.

Introduction

Age-related hearing loss is one of the most common chronic conditions affecting adults (Nelson, Hinojosa 2006; Van Eyken et al. 2007; Huang, Tang 2010; Roth et al. 2011). The alarming tendency of hearing impairment has been reported among younger populations as well (Daniel 2007). The consequences of hearing loss can significantly decrease communicative competence, which can lead to isolation and dementia (Gates, Mills 2005) and to other social dysfunctions (Monzani et al. 2008).

Not only can undiagnosed and underestimated hearing impairment influence everyday interactions, but it can also have an impact on various official situations such as forensic line-up marking and earwitnessing. Furthermore, profound high frequency hearing loss can lead to a serious risk when perception of alarm calls and sirens is not intact (Wong, Leung 2005). This paper presents the results of the investigation into the influence of age-related hearing loss – imitated by the application of several low-pass filters referring to moderate, severe and profound hearing loss thresholds – on auditory identification of a familiar speaker. The present work touches on the fields of forensic linguistic, acoustics and audiology. An inspiration for this study came from a similar methodology applied by McClelland (2008) and Foulkes, Barron (2000), in which the speaker recognition task was conducted within a close social network. The speaker familiarity and the imitation of three conditions of the hearing impairment distinguish the methodology of the proposed study from a typical voice line-up session. In forensic practice, a line-up is usually conducted with speakers who are unfamiliar with the listener but heard during exposure in a crime scene. Hence, the alternation from a typical voice line-up lies in the familiarity of voices the participants are being exposed to (Zetterholm et al. 2012). This paper aims to verify to what extent presbycusis influences earwitnessing. The secondary aim of this work is to verify the effect of speaker familiarity on the correctness of speaker recognition under mimicked hearing loss conditions. Additionally, the confidence of recognition is measured and correlated with the test performance across three degrees of hearing impairment.

The present work is organized as follows: the initial sections provide the rationale for investigating hearing impairment in a forensic context and outline the importance of audiological screening in a line-up procedure. The next part of the paper presents the aetiology of sensorineural hearing loss and statistical data on age-related hearing impairment covering different populations. Then, the remarks on epidemiology and the question of standardization of hearing loss are given. Subsequently, the profiles of participants and modifications of line-up recordings are provided. The latter parts explain the preconditions and procedure. The final passages consist of the discussion of results and general suggestions for the improvement of earwitnessing line-up sessions.

Earwitnessing and hearing loss

Medical reports suggest that age-related hearing loss in European societies starts at the age of 60 (Roth et al. 2011); however, the areal distribution of the starting age is not equal and population-based health surveys need to be compared in order to find region-dependent differences among profiles of patients with hearing difficulties. Since hearing loss can affect a high percentage of a population, it should be treated with special attention, due to its possible impact on speaker recognition (henceforth: SR) and earwitnessing line-ups. The procedure of earwitnessing is to some extent similar to eyewitness line-up (Olsson 2004); and typically does

not include hearing assessment and tests which evaluate cognitive functioning. Many line-up guidelines do not mention the necessity of audiological screening (e.g. de Jong-Lendle et al. 2015). Even in *Considerations of guidelines for earwitness lineups* (Hollien 1996), the hearing assessment of participants was a neglected topic. Later, in *On earwitness lineups* (Hollien 2012: 7), it is suggested that: "The witness also should exhibit hearing adequate for the listening task. Pretests may have to be administered in order to establish these competencies". Thus, how reliable is earwitnessing, when the general ability of speech perception is questionable? To what extent does age-related hearing loss influence speaker recognition in an earwitnessing task? And finally, is the earwitnessing procedure reliable without pretesting subjects' hearing thresholds? This work attempts to answer the abovementioned questions and to define the influence of simulated moderate (MHL), severe (SHL) and profound hearing loss (PHL) on familiar speaker recognition during earwitnessing line-ups. Previous research concerning the influence of hearing impairment on speaker recognition reported an increased perceived listening effort for older listeners with HL (McAuliffe et al. 2012). Several studies had already pointed out that speech and speaker recognition can be influenced not only by hearing impairment, but also by age-related decrease of cognitive functions, regardless of audiological profile and familiarity with the perceived voice (Moore et al. 2014; Füllgrabe et al. 2015). Accurate speaker identification in the SR tests decreases with a lowering range of perceivable bandwidth (Davies-Venn et al. 2015), which directly correlates with a line-up marking among the subjects with diagnosed high frequency hearing loss. If familiar speaker recognition can be challenged due to witness' HL, then the recognition of voices to which a witness had never been exposed should be questioned as well. Hence, it is very important for a forensic expert to know as much as possible about the effect of different levels of hearing impairment on speaker recognition ability. In fact, the relation of HL and earwitnessing might be even more significant for familiar speaker recognition task than typical unfamiliar speaker identification. Since one can no longer audibly recognize familiar people, it might be more relevant – and tragic in consequences – than failing to remember a voice of an unfamiliar speaker.

Hearing impairment in a forensic context

From a forensic perspective, not only unfamiliar but also familiar speaker recognition is important. The increasing percentage of people suffering from hearing loss and its clear influence on auditory speaker recognition should be a matter of concern for forensic experts and judges. Hearing loss is primarily divided into: conductive hearing loss, sensorineural hearing loss, and a combination of both (so-called mixed loss). Age-related hearing impairment – presbycusis (or presbyacusis) – is related to the functional loss of sensory as well as neural elements of auditory pathways (Frisina, Walton 2006), usually characterized by symmetrical, irreversible and progressive hearing impairment. It is worth mentioning that diminished sensitivity

to high frequencies is a normal age-related hearing process, which is why presby-cusis oftentimes remains undiagnosed. At the age of 60 it seems difficult for one to hear frequencies above 12 kHz (Greenberg, Ainsworth 2004); however, not much of linguistically relevant information is transferred in such a register of high fre-quencies. Typically, sensorineural high frequency HL is characterized by bilateral, symmetrical and slowly progressive loss (Ciorba et al. 2012). In addition, the pro-gressive character of HL results in a decrease of perceivable range of frequencies. Depending on affected bandwidth, several stages of HL have been distinguished. Conductive hearing loss differs from presbycusis in aetiology. It is caused by a tran-sition pathology localised in a pathway from the outer ear via the ossicular chain to a tympanic membrane. Conductive hearing loss, due to mechanical causation, is often unilateral and, in contrast to sensorineural HL, typically affects hearing in the range of lower frequencies. Mixed hearing loss is a combination of conductive HL and presbycusis. Concerning the degradation of perceivable acoustic stimuli in HL, the influence of hearing impairments on social functioning is inevitable. The extent of HL affecting a patient's quality of life can be estimated precisely by the specific instruments and social functioning scales, i.e. HHIE – Hearing Handicap Inventory for the Elderly; HHIA – Hearing Handicap Inventory for Adults (New-man et al. 1990, 1991); or IOI-HA – International Outcome Inventory – Hearing Aids (Cox et al. 2003). Not only the obvious emotional and social interactions are affected by hearing loss (Stark, Hickson 2004). Good physical health and quality of life was reported by 39% of patients with HL, compared to 68% of the popula-tion with wider hearing bandwidths (Ciorba et al. 2012). Not only the comfort of life with presbycusis is limited. The age-related hearing loss carries another risk in situations in which high frequency sirens, warning sounds and beepers cannot be localized or may not even be heard.

Patophysiology

Hearing loss can be caused by a variety of genetic, environmental and medical factors. Presbycusis, due to its pathogenesis, can be classified into four types: (1) mechanical or conductive; (2) metabolic, associated with strial atrophy; (3) sen-sory, characterized by hair- and supporting cell loss from the organ of Corti; and (4) neural presbycusis, which can be described as a loss of neurons or neuronal function from the spinal ganglia up to the auditory neuraxis to the temporal cortex (Gates et al. 1989). The histopathologic classification of presbycusis defines: (1) sensory HL characterized by loss of hair cells and a high frequency hearing deficit; (2) metabolic, characterized by loss of stria vascularis with a low fre-quency hearing deficit; and (3) neural characterized by loss of ganglion cells and a variable pattern of hearing loss (Blevins et al. 2015). Schunknecht defined four different types of presbycusis: sensory (outer hair-cell loss), neural (ganglion-cell loss), metabolic (strial atrophy) and cochlear conductive stiffness of the basilar membrane (Huang, Tang 2010). This division was extended with mixed and in-termediate HL (Gates, Mills 2005).

Epidemiology and demographic profile

Hearing loss, as a major public health problem, affects on average 10% of the population to a degree in which successful communication is impaired. This rate increases to 40% in people above 65 years of age (Huang, Tang 2010). Differences in presence and degree of hearing impairment are visible among societies (Agrawal et al. 2008). This observation confirms that genetic, environmental and lifestyle factors are important in the development of presbycusis. Apart from investigating obvious factors such as constant occupational high intensity sound exposure, heritability with clear familial aggregation or cardiovascular diseases; the aetiological research on presbycusis covers other causations, e.g. nicotine or alcohol consumption, heavy metals, diets, lifestyle, leisure activities or socioeconomic status (Cruickshanks et al. 1998; Popelka et al. 2000; Poortinga 2007; Fransen et al. 2008; Dawes et al. 2014). Apart from the above, the risk factors affecting the inner ear area also include barotrauma, ototoxic substances, acoustic neuroma, or meningitis (Weber et al. 2006).

Nearly two decades ago hearing loss in the United States affected 30% to 35% of adults aged 65 to 75. Significant differences were noticed within different ethnicity and sex clusters. According to the reports of National Centre for Health Statistics, unilateral hearing loss was estimated among 20.3% of subjects above 12 years of age; whereas 12.7% of the population was estimated to suffer from bilateral hearing loss (Lin et al. 2011). In the age group between 20 and 29, high frequency HL was present in 8.5% of subjects (Agrawal et al. 2008). Overall estimations suggest that communication dysfluencies caused by hearing-related problems are present in 40% of the population above the age of 65 (Gates, Mills 2005). In a diachronic perspective, population-based studies suggest a significant increase of high frequency (3000, 4000 and 6000 Hz) pure-tone hearing loss among subjects aged 20–29. A similar tendency was discovered for speech-frequency HL. Another noteworthy observation states that there was a "lag of 8 to more than 20 years between the time one is aware of hearing difficulties and when professional hearing help is sought" (Huang, Tang 2010 : 1185). Having mentioned the differences of hearing loss profiles with regard to the investigation area, as well as sex and age of participants (Demeester et al. 2009), the research questions and line-up design should be made specific, to reflect the circumstances common in Scandinavia. The report on age-related HL which covers three Nordic localities opens with the following estimation:

> (…) several million people in the Nordic countries currently suffer from hearing impairment. In the three Nordic countries studied here (Denmark, Finland, and Sweden) the combined population is approximately 19 million people, of whom about 7 million are over age of 50 (Nordic Council of Ministers and the Nordic Council, 2003). The prevalence of hearing impairment increases with age, and is expected to grow in the future as the life-expectancy of these populations increases (…) (Hietanen et al. 2005: 500)

The continuing trend of aging societies in Europe might further exclude the elderly from social functioning due to difficulties in communication caused by age-related hearing loss. However, not only the problem of aging societies is a factor of high

relevance. Nowadays, a degree of high intensity sound exposure and so-called sound pollution severely affect human perception of acoustic stimulus, which in a longitudinal perspective can lead to a decrease of hearing thresholds. Social interactions require effectiveness in recognition of sound patterns, whereas in many cases of age-related HL the range of unperceivable frequencies overlaps with a bandwidth in which human spoken interaction occurs.

Hearing loss standardization

WHO and EU categorize hearing loss in slightly different clusters, thus when referring to a particular type of HL, specific boundary values should be mentioned. The WHO standard is defined by < 26 dB loss as normal hearing; mild loss is diagnosed as $26 \leq$ dB HL < 40; moderate loss: $41 \leq$ dB HL < 60; severe loss: $61 \leq$ dB HL < 80 and profound hearing loss above 80 dB HL. The level of 41 dB HL for the better ear has been classified as disabling. On the other hand, EU standards differ in classification and mention normal hearing threshold up to 21 dB HL; mild loss: $21 \leq$ dB HL < 39; moderate loss: $40 \leq$ dB HL < 69; severe loss: $70 \leq$ dB HL < 94 and profound one above 94 dB (Roth et al. 2011).

Material and method

The familiar speaker recognition task with filtered voice samples was given to eleven participants in order to examine their ability to recognize known speakers under the conditions of simulated profound, severe and moderate hearing loss. The familiarity with speakers introduced an alteration to a classical line-up paradigm. Since it is known that the shift of sound perception threshold typical of presbycusis closely correlates with the decline of speech understanding (Divenyi et al. 2005); the three different low-pass filters were applied to the sound stimuli in order to mimic different degrees of HL and verify the ability to recognize a familiar speaker across three stages of hearing impairment. An interview and a hearing assessment were conducted as part of a pretest to qualify the participants for the SR task. The influence of voice familiarity on the correctness of the line-up markings was examined in the third stage of the study. Apart from speaker familiarity, age and dialect seem significant in speaker recognition (Zetterholm et al. 2012); hence the procedure of selecting participants was conducted taking the above factors into account.

Subjects

In total 11 subjects (1 man and 10 women), aged 23–26, took part in the experiment. None of the participants were diagnosed with phonagnosia. Several previous studies pointed out that sex differences do not correlate with SR ability, which provides a justification for the asymmetrical male/female proportion (e.g. Yarmey, Matthys 1992; Hollien, Schwartz 2000; Hollien 2002). Since the HL thresholds were tested

for the Scandinavian regions and the spoken samples were in Danish, one of the requirements for participation in the study was a native command of the Danish language. Timing of an utterance along with quality of recordings should also be considered in line-up preparation (Hollien 2002). The influence of latency, witnesses' memory, talent or musical training might influence the results of earwitnessing (Hollien 2012). None of the listeners had musical training, nor were diagnosed with memory-related dysfunctions and language deficits.

In order to control for the speakers' mutual familiarity and exposure to heard voices, a questionnaire scaling the frequency of spoken face-to-face interactions with their classmates was handed to the subjects after the line-up markings. This allowed us to avoid a priming effect caused by the names on the questionnaire seen before exposure to the voices. To control for the familiarity and other extralinguistic characteristics, the pool of suspects was selected from the close network of classmates based on several factors such as: clear articulation, dialect, speech rate, pitch and smoking habit. None of the members of the social group were smokers, thus associative priming was also excluded from a set of possible test distractors. As previously stated, the preselection of speakers distinguishes the experimental procedure from a classical line-up paradigm.

Among the abovementioned variables, fundamental frequency has a significant impact on witnesses' ability to recognize a speaker. The mean F_0 values of four speakers reached for Speaker 1: $F_0 = 244$ Hz, SD = 33; Speaker 2: $F_0 = 245$ Hz, SD = 35; Speaker 3: $F_0 = 210$ Hz, SD = 32; Speaker 4: $F_0 = 135$ Hz, SD = 31. One of the speakers exhibited lower F_0 in comparison to the others. As reported by Sørensen (2012), "less common" voices, based on the F_0 measure, should be easier to recognize than the rest of the speakers whose fundamental frequency gravitates towards average values. The recordings were then selected to imitate the line-up situation with respect to these limitations.

Recordings

The utterances were recorded with a Tascam (DR-05) handheld recorder in an acoustically uncontrolled environment in order to imitate a real situation of suspect recognition from spoken evidence reflecting the circumstances of a crime scene exposure. The signal-to-noise (SNR) index of the samples, estimated by a cepstral-based algorithm, varied from 14 dB to 21 dB, which still ensured quality recordings. It guaranteed the absence of distractors in the perceived signal with natural background noise, quite typical for spoken evidence collected from various surveillance systems. In total, four voices were selected. The "suspects" were asked to state a simple asking-for-a-cigarette question. Since pace of speech and the ability to perceive two tones (gap detection) are also important variables in SR tasks for participants who are hard of hearing (Pichora-Fuller et al. 2006), the speech tempo in phones per second was similar in all recordings which excluded the speaker identification based on characteristic speech rate. The drop of hearing thresholds, typical of sensorineural HL, was simulated by applying a set of low-pass filters. Recordings were modified in Praat (Boersma, Weenink 2018), and STx (Balazs 2003) software packages. This method allowed us to

control for the degree of signal distortion and ensured the exact conditions of three different stages of the hearing loss for all listeners.

The voices of four selected "suspects" were presented in the form of the line-up which covered a familiar pool of the classroom network. Previous research pointed out that a size of sample also influences test results in reverse size-performance relation (see Rose 2002). In order to imitate circumstances of a crime exposure, the time of suspect-witness interaction was highly limited and background noise was inevitable. Regarding the time of exposure, some researchers argue for choosing rather short utterances for the line-up sessions (de Jong-Lendle et al. 2015). The kernel of the line-up samples consisted of a short Danish raising intonation phrase: *Undskyld, må jeg få en cigarette*? 'Excuse me, can I get a cigarette?'

Procedure

The experiment session consisted of three parts. The first two parts of the pretest (i.e. an interview and a hearing assessment) aimed to evaluate the hearing ability of the subjects and qualify them for the line-up sessions. Before the audiometric screening, the interview was conducted with the participants. Then, the earwitnessing line-up markings were gathered. After the exposure, the confidence of recognition was rated on a scale from 1 to 5. At last, the questionnaires were handed to the subjects to estimate a degree of familiarity with "suspects" and frequency of face-to-face interactions.

Preconditions

All subjects were interviewed for surgical interventions, cardiovascular factors including diabetes, balance problems and history of chronic middle ear inflammation to verify their ability to participate in testing. Any surgical interventions or conductive HL could have been disqualifying from further procedure. If no obstacles were found, the hearing assessment was conducted as the next step. The subjects underwent the pure tone air-conduction audiometry (250–8000 Hz) screening in an anechoic chamber at Aarhus University, according to clinical standard ISO 8253–1:2010. A manual console for diagnostic screening Oscilla SM950 was used for the hearing assessment. The audiograms were created according to the ASHA regulations (ASHA Guidelines 1997). After the complete hearing assessment, subjects attempted the SR task presented in the form of an earwitnessing line-up.

The participants were exposed to the standardized assessment procedure and all 11 subjects were qualified for the following stage. No hearing impairment was diagnosed. The assessed hearing was binaural without a significant ear privilege. The registered thresholds were within the normal hearing standards, marking the range of 20 dB HL for right and left channel. HL >20 dB for the better ear was an audiological criterion for exclusion, if measured for at least two frequencies of the speech recognition threshold (i.e. 500, 1000 and 2000 Hz). No significant differences were noticed in different age or sex clusters.

Familiar speaker recognition

The instructions were read by the examiner and simultaneously projected on a computer screen in front of the subjects. The trial version of the SR test preceded the line-up. After the trial, the subjects were able to raise questions to clarify the procedure. Then, the participants were asked to identify the speaker after three repetitions of each voice sample in every session. A set of low-passed samples was presented to the listeners in each of the three hearing loss conditions: session 1 – profound HL; session 2 – severe HL; and session 3 – moderate HL. Following the suggestions for conducting line-ups (Hollien 1996), subjects were not observed. Firstly, the recordings with profound HL filters and the highest signal distortion levels were projected to participants (session 1). This decision was motivated by the assumption that more accurate answers will be given after exposure to moderate HL samples (session 3) rather than after the most signal-distorting condition. Then, the subjects were asked to mark the samples corresponding to severe HL and moderate HL conditions. The signal was presented binaurally via the AKG K275 headset. The samples in each condition were randomized in order to exclude sequence memorizing, which could have influenced the decision during the sessions that followed. All the recordings were in Danish – the mother tongue of the subjects. Samples were randomized within the three sets of HL thresholds. The confidence of choices was marked on the 5-point Likert scale, in which 1 corresponded to the least confident choice and 5 meant high confidence of speaker recognition. No feedback information relative to performance was given during or after the procedure.

The recordings of "suspects" of similar age and socioeconomic status (Hollien 1996) were made five months before the actual experiment. As Hollien (2002) suggests, the effects of noncontemporary speech, related to limitations of one's memory, can challenge the auditory speaker identification. Not only sound perception and cognitive limitations of a listener should be taken into consideration in SR tests, but also the possible changes of speaker's voice quality over a certain period of time. Therefore, the impact of various internal and external factors which contribute to challenges of a reliable line-up setup such as intoxication, general health, recent abuse of vocal folds, etc. can modify one's speech quality and should not be neglected in the earwitnessing procedure, either.

Questionnaire

After the familiar speaker recognition test, a questionnaire with names of recorded "suspects" was handed to the participants. The task was to mark a degree of exposure to speech of their classmates in face-to-face conversation. A 5-grade scale was applied, where 1 meant no exposure to his/her voice, 2 – little exposure, 3 – some exposure, 4 – considerable exposure, and 5 – significant exposure. The participants were asked the following question: how often do you talk face-to-face with (*name*)?

Having tested the influence of simulated hearing impairment on performance in familiar SR task, a secondary assumption verified in this study was that the lack of familiarity with someone's voice makes accurate identification more difficult

(Hollien 1996). Accordingly, in the next stage the hypothesis that speakers who are most familiar to witnesses are almost always identified in contrast to listeners who are less familiar with the speakers (Hollien 2002) was verified.

Results

In order to answer the research questions, Friedman's test and Spearman's rank correlation tests were conducted with a level of significance α = 0.05. Results within the range $0.05 < p < 0.1$ were interpreted as significant on the level of statistical tendency.

Degrees of HL and performance

In the first step, the hypothesis of the influence of hearing loss on correctness of speaker recognition was tested by the means of a Friedman's test. As pointed in Table 1, the results are statistically significant. Therefore, the next post-hoc Bonferroni/Dunn's test for multiple comparisons was conducted. At Speaker 3 and Speaker 4, two statistically significant results were noted – between profound HL and severe HL, as well as between profound HL and moderate HL. The performance from session 1, which consisted of recordings mimicking profound hearing loss, was significantly lower. No differences at moderate and severe HL with voices of Speaker 2 and Speaker 3 were noticed. The difference between profound and moderate HL was only close to the level of statistical significance.

Speaker	Hearing Loss	Correct	Incorrect	χ^2	p
Speaker 2	Profound	0.27	0.73		
	Severe	0.64	0.36	8.40	0.015
	Moderate	0.73	0.27		
Speaker 3	Profound	0.27a	0.73		
	Severe	0.73b	0.27	19.51	<0.001
	Moderate	0.91b	0.09		
Speaker 1	Profound	0.19a	0.81		
	Severe	0.91b	0.09	14.25	0.001
	Moderate	0.82ab	0.18		
Speaker 4	Profound	0.82a	0.18		
	Severe	0.91b	0.09	21.41	<0.001
	Moderate	0.91b	0.09		

Table 1. Dunn's test. Performance in speaker recognition depending on the level of HL (*a*, *b* indexes indicate the group differences; p < 0.05; χ^2 – Friedman's test results; *p* – statistical significance). The order of speakers corresponds to the exposure in the first line-up.

Voice familiarity and recognition scores

In the second step, the correlation between correct speaker recognition and the degree of familiarity was verified. The Spearman's rank correlation test was conducted. As presented in Table 2, only one relation was statistically significant. The level of correct recognition of the voice of Speaker 3 was negatively correlated in profound HL condition. Surprisingly, the higher the degree of familiarity marked on the 5-point scale was, the less accurate the recognition.

			Familiarity – Speaker 2	Familiarity – Speaker 3	Familiarity – Speaker 1	Familiarity – Speaker 4
Profound HL	Speaker 2	ρ Spearman's	0.199			
		Significance	0.557			
	Speaker 3	ρ Spearman's		-0.676		
		Significance		0.022		
	Speaker 1	ρ Spearman's			0.245	
		Significance			0.467	
	Speaker 4	ρ Spearman's				-0.156
		Significance				0.648
Severe HL	Speaker 2	ρ Spearman's	-0.061			
		Significance	0.858			
	Speaker 3	ρ Spearman's		-0.534		
		Significance		0.091		
	Speaker 1	ρ Spearman's			-0.164	
		Significance			0.629	
	Speaker 4	ρ Spearman's				-0.104
		Significance				0.760
Moderate HL	Speaker 2	ρ Spearman's	-0.199			
		Significance	0.557			
	Speaker 3	ρ Spearman's		0.055		
		Significance		0.872		
	Speaker 1	ρ Spearman's			-0.245	
		Significance			0.467	
	Speaker 4	ρ Spearman's				-0.104
		Significance				0.760

Table 2. Correlation of speaker identification and degree of familiarity/voice exposure

Confidence of recognition

Confidence of recognition is an important factor which demonstrates witnesses' trustworthiness. It is oftentimes taken into consideration by judges and forensic professionals. General tendencies show that eyewitnesses and earwitnesses are both overconfident in their ability to remember and report correctly (Leippe et al. 2009).

In order to verify the correlation between confidence and the level of HL, the Friedman's test was conducted. As pointed out in Table 3, the results reached the threshold of statistical significance, thus post-hoc Bonferroni/Dunn's test for multiple comparisons was conducted. At profound and moderate HL conditions with Speakers 2, 3 and 4, statistically significant results were measured. In line with the expectations, the confidence level was the highest at moderate HL. The confidence accuracy level marked for the responses given under severe HL was not different from the other clusters even on the level of statistical tendency. The differences between profound vs. severe and moderate HL reached the threshold of significance in the case of the voice of Speaker 1. Overall, the confidence of recognition was the lowest under the profound HL condition. No significant differences were measured at severe and moderate HL clusters even on the level of statistical tendency.

Speaker	Hearing loss	M	SD	χ^2	p
2	Profound	1.45a	0.69		
	Severe	2.36ab	1.03	12.80	0.002
	Moderate	3.36b	1.43		
3	Profound	1.18a	0.40		
	Severe	3.27ab	1.90	16.19	<0.001
	Moderate	4.27b	1.27		
1	Profound	1.27a	0.65		
	Severe	3.18b	1.17	11.73	0.003
	Moderate	3.36b	1.63		
4	Profound	2.63a	1.36		
	Severe	3.54ab	1.44	11.44	0.003
	Moderate	4.27b	1.27		

Table 3. Dunn's test. Confidence of recognition and degree of HL
(*a, b* indexes indicate the group differences; *M* – mean; *SD* – standard deviation; χ^2 – Friedman's test result; *p* – statistical significance)

Confidence of recognition and voice exposure

In the last step of statistical testing, the relation between self-evaluated confidence of recognition and speakers' familiarity/voice exposure was investigated. Again, the Spearman's rank correlation tests were conducted. As stated in Table 4, both statistically significant correlations were noticed in the moderate HL cluster. The level of confidence of recognition was negatively correlated with speakers' familiarity at Speaker 1 and Speaker 4; therefore, the more confident participants were about an identified speaker, the less familiar a speaker was to subjects. The observed correlation was very strong.

Interestingly, two more correlations were observed on the level of statistical tendency. In profound HL condition, the degree of exposure to the voice of Speaker 1 was positively correlated with the confidence of recognition. In this case, the higher the familiarity scores, the higher the confidence of correct identification. The measured relation was also strong. Conversely, within the severe HL cluster, the markings of Speaker 4 showed that the correlation between confidence of recognition and voice exposure was negative. This effect was also strong. Other dependencies were not close to the significance threshold.

			Familiarity – speaker 2	Familiarity – speaker 3	Familiarity – speaker 1	Familiarity – speaker 4
Profound HL	Speaker 2	ρ Spearman's	-0.153			
		Significance	0.652			
	Speaker 3	ρ Spearman's		-0.370		
		Significance		0.263		
	Speaker 1	ρ Spearman's			0.554	
		Significance			0.077	
	Speaker 4	ρ Spearman's				-0.243
		Significance				0.472
Severe HL	Speaker 2	ρ Spearman's	-0.158			
		Significance	0.642			
	Speaker 3	ρ Spearman's		-0.448		
		Significance		0.167		
	Speaker 1	ρ Spearman's			0.039	
		Significance			0.910	
	Speaker 4	ρ Spearman's				-0.557
		Significance				0.075

Table 4. Confidence of recognition and degree of familiarity/voice exposure
(continued on next page)

			Familiarity – speaker 2	Familiarity – speaker 3	Familiarity – speaker 1	Familiarity – speaker 4
Moderate HL	Speaker 2	ρ Spearman's	-0.058			
		Significance	0.865			
	Speaker 3	ρ Spearman's		0.076		
		Significance		0.825		
	Speaker 1	ρ Spearman's			-0.755	
		Significance			0.007	
	Speaker 4	ρ Spearman's				-0.738
		Significance				0.009

Table 4. Confidence of recognition and degree of familiarity/voice exposure (continued)

Discussion and suggestions for improvement

As concluded on the basis of the dataset gathered in this study, acoustically imitated conditions of hearing impairment have a significant influence on familiar speaker recognition. Apart from the degree of familiarity of voice which participants are exposed to the length as well as quality of samples undoubtedly influence speaker recognition performance for both humans and computers (Nolan 1991; Sigmund 2003; Hautamäki et al. 2010; Schroeder 2013; Hollien et al. 2016). Regardless of the limited duration of samples and signal distortion, in this study the correct recognition of familiar speakers appeared possible. Although much longer evidence is often used in forensic procedures, in this study speakers were able to identify a familiar speaker after exposure to very short and filtered spoken samples. The degree of familiarity with voices from the line-up exhibited, counter-intuitively, a negative correlation with performance in SR test. In line with the expectations, it was observed that higher confidence scale ranks were given at the lower degree of signal distortion. As predicted, the highest levels of confidence of correct recognition were noticed at moderate HL condition which corresponded to the highest hearing thresholds. Interestingly, it was also reported that listeners' markings on the scale of confidence of recognition are not in line with actual performance in the SR test. These results show that not only a degree of HL can interfere with correct speaker recognition, but also a degree of exposure to a speaker's voice has an impact on the line-up markings. The results of this study suggest that factors such as the amount of previous exposure to suspects' voice should be taken into consideration in a line-up procedure. On the basis of presented outcomes, it can be also implied

that the earwitnessing task during a line-up procedure should be accompanied by a mandatory audiometric screening. Ideally, the instructions for a hearing assessment with the examples of an audiological interview should be sent to legal institutions and law enforcement personnel conducting line-up sessions. Hearing can be affected not only by the natural ageing process but, among other factors, by intensive short-term sound exposure; thus sending short information to all participants with simple instructions to avoid high intensity sounds immediately before testing might be beneficial.

This study, by simulating the conditions of three degrees of hearing impairment, has pointed out that mimicked hearing loss can influence the scores in familiar speaker recognition tasks. Interestingly, it was reported that markings on a confidence scale do not necessarily overlap with actual performance. The above results can be treated as prediction of difficulties which clinical subjects or often undiagnosed elderly can be confronted with during speaker recognition tasks and line-up sessions. Other cognition-related aspects of the ageing process should also be taken into consideration in validating line-up results. The importance of participants being able to decode speech is crucial in earwitnessing procedures. It seems worth noting that these predictions might not entirely reflect the performance of clinical subjects, due to other social and cognitive-related factors influencing speaker recognition. As previous reports suggest (e.g. Nolan 1980; Sørensen 2012), atypical pitch contour influences recognition of speaker in line-up tests. A range of F0, audibly perceived as uncommon, can also cause that a speaker is distinguished during a line-up. As reported in this study, the level of the signal distortion and filtered thresholds have an influence on earwitnessing line-up markings and familiar speaker recognition. Even though future research would undoubtedly benefit from enlarging the pool of tested subjects, results presented in this study reached the thresholds of statistical significance. In the follow-up, validating the techniques of signal amplification for SR tasks seems to be an interesting topic as well.

One of the other possibilities of improving speech perception in official situations is the application of an induction hearing loop. Many regulations of public buildings design are already introduced (e.g. BS8300 regulation in the UK or EN 60118–4 standard) in order to ensure high quality sound for users of hearing aids. The users of assistive listening devices reported that speech intelligibility and general satisfaction with sound perception is efficiently improved in rooms equipped with an induction loop (Kociński, Ozimek 2015). Participants of the line-up sessions who use hearing aids might then simply be asked to switch the hearing aid to T-mode for the induction loop signal receiver. Another challenge lies in the large number of electronic devices, police stations are often equipped with. Because they emit a magnetic field, such installations should be considered during the magnetic loop calibration as potential signal distractors.

The question of age-related degenerative processes, in the situation of many societies becoming older, seems of high relevance. According to the WHO estimation, by 2025 the increasing tendency of the number of patients suffering from presbycusis will have reached more than 500 million individuals (Sprinzl, Riechelmann 2010),

many of whom could be exposed to line-ups during earwitnessing and speaker recognition tasks. Therefore, the challenges of the future should be considered today, especially in terms of the irreversible character of presbycusis, which could tragically influence real verdicts.

Acknowledgements

References

Agrawal Y., Platz E.A., Niparko J.K. 2008. Prevalence of hearing loss and differences by demographic characteristics among US adults: Data from the National Health and Nutrition Examination Survey, 1999–2004. – *Archives of Internal Medicine* 168.14: 1522–1530.

ASHA Guidelines 1997 = American Speech-Language-Hearing Association. 1997. *Guidelines for Audiologic Screening*. [www.asha.org/policy].

Balazs P. 2003. Masking filter, phase vocoder and STx – an introduction. – *SAMPTA03-Sampling Theory and Applications*. [https://www.kfs.oeaw.ac.at/index.php?option= com_content&view=article&id=46:stools-stx-intelligent-sound- processing&catid= 179&Itemid=854&showall=&limitstart=2&lang=en].

Blevins N., Deschler D., Park L. 2015. Presbycusis. – *UpToDate*. Waltham (MA). [https://www.uptodate.com/contents/presbycusis/print].

Boersma P., Weenink D. 2018. Praat: Doing phonetics by computer. [Computer program, Version 6.0.40; http://www.praat.org/].

Ciorba A., Bianchini C., Pelucchi S., Pastore A. 2012. The impact of hearing loss on the quality of life of elderly adults. – *Clinical Interventions in Aging* 7: 159–163.

Cox R.M., Alexander G.C., Beyer C.M. 2003. Norms for the international outcome inventory for hearing aids. – *Journal of the American Academy of Audiology* 14.8: 403–413.

Cruickshanks K.J., Klein R., Klein B.E., Wiley T.L., Nondahl D.M., Tweed T.S. 1998. Cigarette smoking and hearing loss: The epidemiology of hearing loss study. – *JAMA* 279.21: 1715–1719.

Daniel E. (rev.). 2007. Noise and hearing loss. – *Journal of School Health* 77.5: 225–231.

Davies-Venn E., Nelson P., Souza P. 2015. Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearing. – *The Journal of the Acoustical Society of America* 138.1: 492–503.

Dawes P., Cruickshanks K.J., Moore D.R., Edmondson-Jones M., McCormack A., Fortnum H., Munro K.J. 2014. Cigarette smoking, passive smoking, alcohol consumption, and hearing loss. – *Journal of the Association for Research in Otolaryngology* 15.4: 663–674.

de Jong-Lendle G., Nolan F., McDougall K., Hudson T. 2015. Voice lineups: A practical guide. – The Scottish Consortium for ICPhS 2015 (ed.). *18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow. [http://www.internationalphoneticassociation.org/icphs- proceedings/ ICPhS2015/ Papers/ICPHS0598.pdf].

Demeester K., Van Wieringen A., Hendrickx J.J., Topsakal V., Fransen E., Van Laer L., Van de Heyning P. 2009. Audiometric shape and presbycusis. – *International Journal of Audiology* 48.4: 222–232.

Divenyi P.L., Stark P.B., Haupt K.M. 2005. Decline of speech understanding and auditory thresholds in the elderly. – *The Journal of the Acoustical Society of America* 118.2: 1089–1100.

Foulkes P., Barron A. 2000. Telephone speaker recognition amongst members of a close social network. – *Forensic Linguistics* 7: 180–198.

Fransen E., Topsakal V., Hendrickx J.J., Van Laer L., Huyghe J.R., Van Eyken E., Lemkens N., Hannula S., Mäki-Torkko E., Jensen M., Demeester K., Tropitzsch A., Bonaconsa A., Mazzoli M., Esposo A., Verbruggen K., Huygen P.L.M., Kunst S., Manninen M., Diaz-Lacava A., Steffens M., Wienker T.F., Pyykkö I., Cremers C.W.R.J., Kremer H., Dhooge I., Stephens D., Orzan E., Pfister M., Bille M., Parving A., Sorri M., Van de Heyning P., Van Camp G. 2008. Occupational noise, smoking, and a high body mass index are risk factors for age-related hearing impairment and moderate alcohol consumption is protective: A European population-based multicentre study. – *Journal of the Association for Research in Otolaryngology* 9.3: 264–276.

Frisina R.D., Walton J.P. 2006. Age-related structural and functional changes in the cochlear nucleus. – *Hearing Research* 216: 216–223.

Füllgrabe C., Moore B.C., Stone M.A. 2015. Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. – *Frontiers in Aging Neuroscience* 6: 1–25. [https://www.frontiersin.org/articles/10.3389/fnagi.2014.00347/full].

Gates G.A., Caspary D.M., Clark W., Pillsbury III H.C., Brown S.C., Dobie R.A. 1989. Presbycusis. – *Otolaryngology – Head and Neck Surgery* 100.4: 266–271.

Gates G.A., Mills J.H. 2005. Presbycusis. – *The Lancet* 366.9491: 1111–1120.

Greenberg S., Ainsworth W.A. 2004. Speech processing in the auditory system: An overview. – Greenberg S., Popper A. (eds.). *Speech processing in the auditory system*. New York: Springer.

Hautamäki V., Kinnunen T., Nosratighods M., Lee K.A., Ma B., Li H. 2010. Approaching human listener accuracy with modern speaker verification. – *INTERSPEECH-2010*: 1473–1476.

Hietanen A., Era P., Henrichsen J., Rosenhall U., Sorri M., Heikkinen E. 2005. Hearing among 75-year-old people in three Nordic localities: A comparative study. – *International Journal of Audiology* 44.9: 500–508.

Hollien H. 1996. Consideration of guidelines for earwitness lineups. – *Forensic Linguistics* 3: 14–23.

Hollien H. 2002. *Forensic Voice Identification*. London: Academic Press.

Hollien H. 2012. On earwitness lineups. – *Investigative Sciences Journal* 4.1: 1–17.

Hollien H., Didla G., Harnsberger J.D., Hollien K.A. 2016. The case for aural perceptual speaker identification. – *Forensic Science International* 269: 8–20.

Hollien H., Schwartz R. 2000. Aural-perceptual speaker identification: Problems with non-contemporary samples. – *Forensic Linguistics* 7: 199–211.

Huang Q., Tang J. 2010. Age-related hearing loss or presbycusis. – *European Archives of Oto-Rhino-Laryngology* 267.8: 1179–1191.

Kociński J., Ozimek E. 2015. Speech intelligibility in rooms with and without an induction loop for hearing aid users. – *Archives of Acoustics* 40.1: 51–58.

Leippe M.R., Eisenstadt D., Rauch S.M. 2009. Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. – *Law and Human Behavior* 33.3: 194–212.

Lin F.R., Niparko J.K., Ferrucci L. 2011. Hearing loss prevalence in the United States. – *Archives of Internal Medicine* 171.20: 1851–1853.

McAuliffe M.J., Wilding P.J., Rickard N.A., O'Beirne G.A. 2012. Effect of speaker age on speech recognition and perceived listening effort in older adults with hearing loss. – *Journal of Speech, Language, and Hearing Research* 55: 838–847.

McClelland E. 2008. Voice recognition within a closed set of family members. [Paper presented at the International Association for Forensic Phonetics and Acoustics Annual Conference, Lausanne, 20–23 July].

Monzani D., Galeazzi G.M., Genovese E., Marrara A., Martini A. 2008. Psychological profile and social behaviour of working adults with mild or moderate hearing loss. – *Acta Otorhinolaryngologica Italica* 28.2: 61–66.

Moore D.R., Edmondson-Jones M., Dawes P., Fortnum H., McCormack A., Pierzycki R.H., Munro K.J. 2014. Relation between speech-in-noise threshold, hearing loss and cognition from 40–69 years of age. – *PLoS ONE* 9.9: e107720. [doi: 10.1371/journal.pone.0107720].

Nelson E.G., Hinojosa R. 2006. Presbycusis: A human temporal bone study of individuals with downward sloping audiometric patterns of hearing loss and review of the literature. – *The Laryngoscope* 116 (S112): 1–12.

Newman C.W., Weinstein B.E., Jacobson G.P., Hug G.A. 1990. The hearing handicap inventory for adults: Psychometric adequacy and audiometric correlates. – *Ear and Hearing* 11.6: 430–433.

Newman C.W., Weinstein B.E., Jacobson G.P., Hug G.A. 1991. Test-retest reliability of the hearing handicap inventory for adults. – *Ear and Hearing* 12.5: 355–357.

Nolan F. 1991. Forensic phonetics. – *Journal of Linguistics* 27.2: 483–493.

Nolan F.J.D. 1980. *The phonetic bases of speaker recognition*. Cambridge: Cambridge University Press.

Olsson J. 2004. *Forensic linguistics: An introduction to language crime and the law*. London, New York: Continuum.

Pichora-Fuller M.K., Schneider B.A., Benson N.J., Hamstra S.J., Storzer E. 2006. Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry. – *The Journal of the Acoustical Society of America* 119.2: 1143–1155.

Poortinga W. 2007. The prevalence and clustering of four major lifestyle risk factors in an English adult population. – *Preventive Medicine* 44.2: 124–128.

Popelka M.M., Cruickshanks K.J., Wiley T.L., Tweed T.S., Klein B.E., Klein R., Nondahl D.M. 2000. Moderate alcohol consumption and hearing loss: A protective effect. – *Journal of the American Geriatrics Society* 48.10: 1273–1278.

Rose P. 2002. *Forensic speaker identification*. London, New York: Taylor & Francis.

Roth T.N., Hanebuth D., Probst R. (rev.). 2011. Prevalence of age-related hearing loss in Europe. – *European Archives of Oto-Rhino-Laryngology* 268.8: 1101–1107.

Schroeder M.R. 2013. *Computer speech: Recognition, compression, synthesis.* [vol. 35]. Berlin, Heidelberg: Springer Science & Business Media.

Sigmund M. 2003. *Voice recognition by computer*. Marburg: Tectum Verlag.

Sørensen M.H. 2012. Voice line-ups: Speakers' F0 values influence the reliability of voice recognitions. – *International Journal of Speech, Language & the Law* 19.2: 145–158.

Sprinzl G.M., Riechelmann H. 2010. Current trends in treating hearing loss in elderly people: A review of the technology and treatment options – a mini-review. *Gerontology* 56.3: 351–358.

Stark P., Hickson L. 2004. Outcomes of hearing aid fitting for older people with hearing impairment and their significant others. – *International Journal of Audiology* 43.7: 390–398.

Van Eyken E., Van Camp G., Van Laer L. 2007. The complexity of age-related hearing impairment: Contributing environmental and genetic factors. – *Audiology and Neurotology* 12.6: 345–358.

Weber P.C., Deschler D., Sokol N. 2006. Etiology of hearing loss in adults. – *Up-To-Date* 13: 3. [https://www.uptodate.com/contents/etiology-of-hearing-loss-in-adults].

Wong L.T., Leung L.K. 2005. Minimum fire alarm sound pressure level for elder care centres. – *Building and Environment* 40.1: 125–133.

Yarmey A.D., Matthys E. 1992. Voice identification of an abductor. – *Applied Cognitive Psychology* 6: 367–377.

Zetterholm E., Sarwar F., Thorvaldsson V., Allwood C.M. 2012. Earwitnesses: The effect of type of vocal differences on correct identification and confidence accuracy. – *International Journal of Speech, Language & the Law* 19.2: 219–237.