Abstract

Hearing loss (HL) has multifaceted negative consequences for individuals of all age groups. Despite individual fitting based on clinical assessment, consequent usage of hearing aids (HAs) as a remedy is often discouraged due to unsatisfactory HA performance. Consequently, the methodological complexity in the development of HA algorithms has been increased by employing virtual acoustic environments which enable the simulation of indoor scenarios with plausible room acoustics. Inspired by the research question of how to make such environments accessible to HA users while maintaining complete signal control, a novel concept addressing combined perception via HAs and residual hearing is proposed. The specific system implementations employ a master HA and research HAs for aided signal provision, and loudspeaker-based spatial audio methods for external sound field reproduction. Systematic objective evaluations led to recommendations of configurations for reliable system operation, accounting for perceptual aspects. The results from perceptual evaluations involving adults with normal hearing revealed that the characteristics of the used research HAs primarily affect sound localisation performance, while allowing comparable egocentric auditory distance estimates as observed when using loudspeaker-based reproduction. To demonstrate the applicability of the system, school-age children with HL fitted with research HAs were tested for speech-in-noise perception in a virtual classroom and achieved comparable speech reception thresholds as a comparison group using commercial HAs, which supports the validity of the HA simulation. The inability to perform spatial unmasking of speech compared to their peers with normal hearing implies that reverberation times of 0.4 s already have extensive disruptive effects on spatial processing in children with HL. Collectively, the results from evaluation and application indicate that the proposed systems satisfy core criteria towards their use in HA research.

Kurzfassung

Hörverlust (HV) hat vielfältige negative Auswirkungen auf Menschen aller Altersgruppen. Trotz individueller Anpassung auf Basis von klinischen Verfahren werden Hörgeräte (HG) als Gegenmaßnahme wegen unzureichender Leistung von Betroffenen oft nicht akzeptiert. Infolgedessen wurde die methodische Komplexität in der Entwicklung von HG-Algorithmen durch die Verwendung von virtuellen akustischen Umgebungen, welche die Simulation von Innenraumszenarien mit plausibler Raumakustik ermöglichen, erhöht. Um HG-Träger*innen solche Umgebungen zugänglich zu machen, wird ein neuartiges Konzept zur Berücksichtigung der kombinierten Wahrnehmung über HG und das Resthörvermögen vorgeschlagen. Die Systemimplementierungen verwenden eine Softwareplattform für HG-Algorithmen in Kombination mit Forschungs-HG für die Wiedergabe der HG-Signale, und lautsprecherbasierte räumliche Wiedergabemethoden für die externe Schallfeldwiedergabe. Für einen zuverlässigen Systembetrieb werden unter Berücksichtigung perzeptiver Anforderungen Empfehlungen zu Systemkonfigurationen abgegeben. Untersuchungen mit normalhörenden Erwachsenen zeigen, dass die Eigenschaften der Forschungs-HG primär die Schalllokalisation beeinflussen, jedoch vergleichbare egozentrische auditive Entfernungsschätzungen wie bei lautsprecherbasierter Wiedergabe ermöglichen. Zur Demonstration der Systemanwendbarkeit wurde die Sprachwahrnehmung von Schulkindern mit HV bei Verwendung der Forschungs-HG in einem virtuellen Klassenzimmer in Präsenz von Störsprecherinnen getestet. Ähnliche Sprachverständlichkeitsschwellen wie jene einer Vergleichsgruppe mit kommerziellen HG bekräftigen die Plausibilität der HG-Simulation. Die Unfähigkeit der räumlichen Demaskierung von Sprache deutet darauf hin, dass bereits Nachhallzeiten ab 0,4 s die räumliche Verarbeitung von Kindern mit HV erheblich beeinträchtigen. Insgesamt zeigen die Evaluierungs- und Anwendungsergebnisse, dass die vorgeschlagenen Systeme wesentliche Kriterien für ihren Einsatz in der HG-Forschung erfüllen.

1

Introduction

Listening, understanding and communicating are essential skills for social interaction in everyday life. Hearing loss (HL) may have severe consequences on these skills and affect individuals by hindering language development and learning success already at an early age (Stevenson et al., 2010; Richardson, Long, & Woodley, 2004). According to the World Health Organization (2021), about 5 % of the world population, amounting to 432 million adults and 34 million children, suffer from disabling HL – a number that is estimated to grow to 700 million people by 2050. Depending on the individual predisposition, affected children are likely to withdraw from socialisation processes and sometimes suffer from compromised mental and general health (Levy-Shiff & Hoffman, 1985; Hogan et al., 2011; Stevenson et al., 2011; Patel et al., 2021). Similar negative consequences may also apply to adults and elderly people, especially those who refuse to acknowledge their HL and appropriate treatment, leading to an increased risk of isolation (Mick, Kawachi, & Lin, 2014; Shukla et al., 2020), depression (Boi et al., 2012; Mener et al., 2013), as well as memory problems and dementia (Wingfield, Tun, & McCoy, 2005; Moorman, Greenfield, & Lee, 2021).

Selectively listening to a target talker while ignoring masking and distracting noise sources – commonly referred to as the “cocktail party effect” (Cherry, 1953) – is among the most important abilities for successful communication. The outcomes of behavioral studies in the field of speech-in-noise (SiN) perception corroborate inferior unmasking performance of target speech in individuals with HL across age groups, compared to age-matched individuals with normal hearing (NH) (Cameron, Glyde, & Dillon, 2011; Bronkhorst, 2000). From a historical

Related publication:
Pausch et al., 2018b.

perspective, cupping the hand behind the ear can be considered maybe the first natural and still extensively used technique to increase speech intelligibility in such situations. Besides acting as a visual sign for the conversational partner to speak up, the cupped hand creates acoustic resonances between 1 and 3 kHz that are beneficial for speech understanding, and enhances directivity to increase the signal-to-noise ratio (SNR) between the target talker and detrimental noise sources (Uchanski & Sarli, 2019). Possibly inspired by the shape of the ears in animals, people also experimented with mechanical hearing aids (HAs) in the form of hearing trumpets, allowing to further increase directivity and tailor the resonance frequencies. Such devices had a multifaceted appearance, including chairs with two attached horns, representing probably one of the first approaches to provide binaural amplification (Stephens & Goodwin, 1984; Hawley, Litovsky, & Culling, 2004). Electric HAs were implemented by the end of the 19th century, initiated by the invention of the telephone, and since then continuously miniaturised by incorporating advancing technologies such as integrated circuit designs (Mills, 2011). Unobtrusive and miniaturised designs contributed significantly to a higher acceptance rate of HAs in view of the stigma of HL (David & Werner, 2016). While the parameters of the first analogue devices were adjusted using potentiometers, rapid progress in digital technology soon enabled to program analogue devices digitally, and to develop completely digitised devices with on-board digital signal processors (DSPs) (Kates, 2008). Nowadays, assistive listening devices, such as HAs, represent an effective technical intervention tool capable of partially restoring impaired hearing and helping to overcome the aforementioned consequences of HL (Chisolm et al., 2007; Tomblin et al., 2014). Various sophisticated HA algorithms have been specifically designed to address the deficiencies in affected individuals (Hamacher et al., 2005; Kates, 2008). More recently, access to the devices has become even more convenient by the availability of over-the-counter HAs (Warren & Grassley, 2017) and self-fitting approaches (Keidser & Convery, 2016). Premium devices sometimes feature additional acoustic scene classification algorithms (Yellamsetty et al., 2021), utilising machine learning approaches such as deep learning (Vivek, Vidhya, & Madhanmohan, 2020).

Despite the major advances in technology, extensive surveys reported that HA users still complain about poor performance of their devices, particularly in noisy situations like demanding indoor communication settings with unfavourable room acoustics (Hougaard & Ruf, 2011) – although there seems to be a trend towards higher satisfaction rates (Bisgaard & Ruf, 2017). Dissatisfaction may consequently lead to irregular device usage (Bertoli et al., 2009) and, if at all, only marginal improvement of the original problems. To some extent, the experienced performance deficiencies can be attributed to clinical assessment and fitting routines for the parametrisation of HAs, which may result in settings that are less effective in complex real-world situations (Cord et al., 2007; Timmer, Hickson, & Launer, 2018). Evaluating HA algorithms using oversimplified test scenarios also favours a gap between laboratory performance and reality. Significant advances in the field of acoustic virtual reality (Vorländer, 2020) have triggered research groups both in academia and industrial companies to utilise virtual acoustic environments (VAEs) for the design and improvement of HAs under more plausible conditions. The fact that reproduction of VAEs via headphones is not feasible and likely entails uncontrolled HA algorithm behaviour, let alone feedback issues, motivates the use of loudspeaker-based spatial audio reproduction (Minnaar, Favrot, & Buchholz, 2010; Grimm, Ewert, & Hohmann, 2015; Grimm, Kollmeier, & Hohmann, 2016; Oreinos & Buchholz, 2016). Investigations included, for example, assessing the real-world benefit of beamforming algorithms (Compton-Conley et al., 2004; Gnewikow et al., 2009), or aimed at perceptually validating SiN test results obtained in loudspeaker-based VAEs (Cubick & Dau, 2016). The availability of increased computational resources allowed to implement interactive low-latency listening scenarios using advanced room acoustic simulations in combination with highly efficient convolution algorithms (Noisternig et al., 2008; Pelzer et al., 2014; Mehra et al., 2015; Wefers, 2015; Schissler, Stirling, & Mehra, 2017).

This thesis contributes to the continued bridging of the aforementioned gap by exploring VAEs adapted for HA applications, rendered and reproduced by state-of-the-art simulation methods and spatial audio reproduction technologies, respectively. Novel binaural and hybrid concepts address the main research question of how to properly integrate HAs in the virtual scene, while retaining full control over HA algorithm parametrisation and the signals involved. The proposed systems are comparatively evaluated based on measurements and perceptual experiments, to be subsequently applied in a clinical experiment. Moreover, the flexibility and effectiveness of VAEs shall be demonstrated to further promote their application in HA research.

Thesis outline

A brief summary of the thesis structure is provided below per chapter. Chapter 2 introduces the required basics and fundamental concepts. At the beginning of Chapter 3, general application areas of VAEs are presented. Adaptations and extensions necessary to enable application in HA research are elaborated based on the proposed concept. After a compact introduction to current spatial audio reproduction methods, two specific implementation variants are discussed. Both approaches allow for transparent simulation and reproduction of signals as combined via research HAs and loudspeaker-based methods, either utilising binaural technology with a real-time processing backbone, or panning-based approaches as part of a static hybrid auralisation system. In Chapter 4, the two system variants become subject to objective evaluations based on different types of spatial rendering functions. Drawing on the results from a benchmark analysis of an example virtual indoor scene using the combined binaural real-time auralisation system, recommendations on simulation configurations for operation on conventional desktop computers are derived. Measurements on combined end-to-end latency (EEL) further provide insights on the reactiveness of the implemented system. Subsequently, results on the ideal and achievable channel separation (CS) in loudspeaker-based binaural reproduction under consideration of detrimental reflections is investigated. The errors in recreated spatial transfer functions obtained from different spatial audio reproduction methods are analysed at ear level in the presence of detrimental reflections. Chapter 5 investigates selected spatial audio quality parameters in the scope of two perceptual experiments involving adults with NH. In Experiment 1, differences in the localisation performance of virtual sound sources (VSSs), reproduced via headphones and the individual reproduction paths of the binaural real-time auralisation system, are compared to the localisation of real sound sources (RSSs). Experiment 2 assesses egocentric auditory distance perception in simulated room acoustics depending on variations in the specific implementation of the static hybrid auralisation system. Chapter 6 demonstrates the application of the binaural real-time auralisation system in a clinical experiment. Children with HL are tested for SiN perception in the presence of distracting talkers, while immersed in simulated room acoustics. The results are compared with those from an age-matched control group of children with NH to identify performance differences and discuss their reasons, addressing aspects of adverse room acoustics. Finally, the key findings are summarised and an outlook with ideas for future work is provided.

2

Fundamentals

This chapter provides concise descriptions of the fundamental concepts used in this thesis. The first part focuses on coordinate system conventions, digital signal processing basics, the measurement of impulse responses (IRs) of linear time-invariant (LTI) systems, such as rooms, room acoustic parameters, and source directivity. The second part deals with the human auditory system and threshold measurements, selected psychoacoustic topics, as well as HL and its perceptual effects. HAs and HA algorithm rationales are introduced and discussed in the scope of their rehabilitative purposes. Subsequently, binaural listening concepts are presented on the basis of directional transfer functions in listeners with and without HAs, introducing the most important localisation cues and differences between transfer function characteristics. The chapter ends with a short introduction to SiN tests and related terms to quantify SiN perception performance.

2.1 Coordinate systems

In this thesis, two different coordinate systems are used: a right-handed fixed *global* coordinate system and a right-handed variable *local* coordinate system, each with three orthogonal axes x, y and z, and x', y' and z', respectively, see Figure 2.1a. The positions of rigid bodies are represented by their current translational offsets to the origin of the global coordinate system in three-dimensional space, and accordingly defined by a triplet of Cartesian coordinates $[x \quad y \quad z]^{\mathsf{T}}$. Independent from the position, each rigid body has a current orientation. Po-

---

Related publications:

Pausch et al., 2018b; Pausch and Fels, 2020; Pausch, Doma, and Fels, 2021.