Auditory training strategies to improve speech intelligibility in hearing-impaired listeners

Koprowska, Aleksandra Maria

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Koprowska, A. M. (2022). *Auditory training strategies to improve speech intelligibility in hearing-impaired listeners*. DTU Health Technology. Contributions to Hearing Research Vol. 51

CONTRIBUTIONS TO
HEARING RESEARCH

Volume 51

Aleksandra Koprowska

Auditory training strategies to improve speech intelligibility in hearing-impaired listeners

DTU Health Tech
Department of Health Technology

Auditory training strategies to improve speech intelligibility in hearing-impaired listeners

PhD thesis by
Aleksandra Koprowska

Preliminary version: January 18, 2022

Technical University of Denmark

2022

Preprint version for the assessment committee.
Pagination will differ in the final published version.
Cover picture: © Natalia Koprowska, 2022

This PhD dissertation is the result of a research project carried out at the Hearing Systems Section, Department of Health Technology, Technical University of Denmark (Kgs. Lyngby, Denmark). Part of the project was carried out at the Institute of Clinical Research at the University of Southern Denmark (Odense, Denmark) and at the Copenhagen Hearing and Balance Center (Copenhagen, Denmark).

The project was partly financed by WSAudiology (2/3) and by the Technical University of Denmark (1/3).

Supervisors

Assoc. Prof. Jeremy Marozeau
Prof. Torsten Dau
Hearing Systems Section
Department of Health Technology
Technical University of Denmark
Kgs. Lyngby, Denmark

Dr. Maja Serman
WSAudiology
Erlangen, Germany

Abstract

Hearing loss causes substantial challenges in understanding speech. Hearing aids often improve speech perception in quiet but difficulties in understanding speech in noise remain one of the most common problems reported by hearing-aid users. Therefore, other ways of improving the speech-in-noise perception are of interest to both research and the hearing-aid industry. Auditory training is one of possible strategies to complement the usage of hearing aids by enhancing skills relevant for speech perception in noisy environment. The experiments described in this thesis investigated the efficacy of two selected auditory training methods (based on speech and musical stimuli) in a group of experienced hearing-aid users. The first method, phoneme-in-noise training, aimed to retrain the usage of speech cues in the presence of background noise. The second method, training motor synchronization with the musical beat, was meant to enhance predictive mechanisms that contribute to speech perception. The training outcomes were assessed using metrics of speech intelligibility in noise (both methods) and listening effort (in the case of phoneme-in-noise training) and were related to the results obtained in an active control group.

In the first chapter, the influence of phoneme-in-noise training on speech intelligibility is investigated. The training effects were assessed using logatomes from Danish nonsense word corpus (DANOK) as well as a hearing in noise test (HINT). The training improved identification scores in noise for vowels and consonants located in the middle of the logatomes but not for the onset consonants. The observed improvements were still present three months after the training. No significant impact on speech recognition thresholds (SRTs) measured with the HINT was found.

The second chapter describes the effect of phoneme-in-noise training on listening effort required to understand the HINT sentences, which was measured using pupillometry. Changes in pupil dilations were recorded when the participants were listening to sentences presented at two signal-to-noise ratios (SNRs): corresponding to 50% (SNR50) and 80% (SNR80) speech intelligibility (based on the pre-intervention performance). For both test conditions, the analysis of the pupil response patterns suggested that listening effort increased in both participant groups, but the change was more pronounced in the group that completed the training. As opposed to SRTs (reported in the first chapter), the speech recognition scores measured at fixed SNR revealed improved speech intelligibility in noise in the more challenging condition (SNR50). The outcomes demonstrated that pupillometry was sensitive to the effect of training in both conditions, also when the HINT did not reveal any significant performance improvement (SNR80). The elevation of listening effort exerted during the HINT might have been driven by increased motivation (in both groups) as well as changes in the cognitive resources allocation mechanisms (only in the group that completed the training).

The third chapter presents the design and implementation of a novel music-based training approach to support speech understanding in noise in hearing-aid users. The training program aimed at improving motor synchronization with the beat of music and was administered at the participants' homes using a mobile application. Modest improvements in synchronized tapping consistency throughout the training were found, but they did not result in significantly better performance in the outcome measures of beat perception or speech intelligibility in noise. Longer duration and higher intensity of training as well as more individually tailored task demands might be needed to obtain significant benefits using this method.

The results presented in this thesis can inform future development and refinement of speech- and music-based auditory training approaches as well as the selection of methods used to assess the benefits of these interventions.

Resumé

Høretab medfører væsentlige udfordringer for taleforståelsen. Høreapparater forbedrer ofte taleopfattelsen i rolige omgivelser, men at forstå tale i støj er fortsat et de mest almindelige problemer, høreapparatbrugere rapporterer. Andre måder at forbedre taleopfattelse i støj er derfor interessant for både forskning og høreapparatindustri. Auditiv træning er en af de mulige strategier for at komplementere brug af høreapparater med udvikling af færdigheder, som er relevante for taleopfattelse i støjende miljøer. Eksperimenterne, der er beskrevet i denne afhandling, undersøgte effekten af to udvalgte auditive træningsmetoder (baseret på tale og musiske stimuli) hos en gruppe af erfarne høreapparatsbrugere. Den første metode – fonem-i-støj-træning – havde til formål at genoptræne brug af perceptuel information i tale i baggrundsstøj. Den anden metode – træning af motorisk synkronisering med rytmen i musik – forsøgte at styrke de forudsigelsesmekanismer, der bidrager til taleopfattelse. Træningsudbyttet blev evalueret gennem måling af taleforståelse i baggrundsstøj (begge træningsmetoder) og lytteanstrengelse (kun fonem-i-støj-træningen) sammenlignet med resultater fra en aktiv kontrolgruppe.

I det første afsnit undersøges effekten af fonem-i-støj-træning på taleforståelsen. Træningseffekter blev evalueret ved brug af logatomer fra det danske nonsens-ord-korpus (DANOK) samt sætningstest i baggrundsstøj (eng. *hearing in noise test*; HINT). Træningen forbedrede identifikationsscorerne for vokaler og konsonanter midt i logatomerne men ikke for konsonanter i begyndelsen af logatomerne. Forbedringerne var stadigvæk til stede tre måneder efter afslutningen af træningen. Der blev ikke fundet nogen signifikant effekt på talegenkendelsestærsklerne (eng. *speech recognition threshold*; SRT) i HINT.

Det andet afsnit beskriver effekten af fonem-i-støj-træningen på den lytteanstrengelse der er nødvendig for at forstå HINT-sætninger, hvilket blev målt ved hjælp af pupillometri. Ændringer i pupilstørrelse blev registreret mens forsøgsdeltagerne lyttede til sætninger, som blev præsenteret under to forskellige signal-støj-forhold (eng. *signal-to-noise ratio*; SNR): svarende til 50% (SNR50) og 80% (SNR80) taleforståelse (fastlagt af performance før behandlingen). I begge eksperimentelle forhold tydede analyse af pupilresponser på stigning i lytteanstrengelse i begge deltagergrupper, men forandringen var mere markant i gruppen der modtog træning. I modsætning til SRT-resultaterne (som blev rapporteret i det første afsnit), viste talegenkendelsesscorerne målt med fast SNR en forbedring af taleforståelse under vanskeligere forhold (SNR50). Resultaterne demonstrerede, at pupillometri var følsom overfor træningseffekter

i begge testforhold, også når HINT ikke viste nogen signifikant forbedring af performance (SNR80). Stigningen i lytteanstrengelse under HINT kan være en følge af øget motivation (hos begge grupper) samt ændringer i mekanismer for tildeling af kognitive ressourcer (kun hos træningsgruppen).

Det tredje afsnit præsenterer design og implementering af en ny musikbaseret træningsmetode til forbedring af evnen til af forstå tale i baggrundsstøj hos brugere af høreapparater. Forsøgsdeltagerne gennemførte et træningsprogram, som havde til formål at forbedre motorisk synkronisering med rytmen i musik derhjemme ved hjælp af en mobil-app. Beskeden forbedring i motorisk synkronisering blev fundet under træningen, men førte ikke til signifikant bedre præstation i hverken rytmopfattelse eller taleforståelse i støj. Der kræves muligvis et mere længerevarende og intenst træningsprogram samt en mere individuelt tilpasset sværhedsgrad af opgaverne for at opnå signifikante fordele ved brug af denne metode.

Resultaterne præsenteret i denne afhandling kan benyttes til videreudvikling af tale- og musik-baserede auditive træningsprogrammer samt til valg af målemetoder til at evaluere udbyttet af disse interventioner.

Streszczenie

Niedosłuch jest przyczyną znaczących trudności w zrozumieniu mowy. Aparty słuchowe często poprawiają percepcję mowy w ciszy, jednak trudności ze zrozumieniem mowy w szumie pozostają jednym z najczęściej wymienianych problemów u użytkowników tych urządzeń. Z tej przyczyny, inne sposoby poprawy zroumiałości mowy w hałasie u osób z niedosłuchem spotykają się ze sporym zainteresowaniem zarówno ze strony naukowców, jak i producentów aparatów słuchowych. Trening słuchowy jest strategią mogącą dopełnić korzyści płynące z użytkowania aparatów słuchowych poprzez poprawę umiejętności niezbędnych dla percepcji mowy w hałasie. Eksperymenty przedstawione w niniejszej pracy zbadały skuteczność dwu wybranych metod treningu słuchowego (opartych o mowę bądź bodźce muzyczne) u grupy doświadczonych użytkowników aparatów słuchowych. Pierwsza z nich, trening fonemów w szumie, miała na celu przywrócenie zdolności wykorzystywania informacji zawartych w sygnale mowy niezbędnych do kategoryzacji fonemów. Drugą metodą był trening synchronizacji ruchowej z pulsem muzycznych, którego skutkiem miało być udoskonalenie mechanizmów predykcji biorących udział w percepcji mowy. Rezultaty treningu zostały ocenione za pomocą pomiaru zrozumiałosci mowy w szumie (w przypadku obu metod) oraz wysiłku słuchowego (w przypadku treningu fonemów w szumie) w odniesieniu do wyników aktywnej group kontrolnej.

W rozdziale pierwszym został opisany wpływ treningu fonemów w szumie na zrozumiałość mowy. Efekty treningu zostały zmierzone przy użyciu logatomów z duńskiego korpusu wyrazów bez znaczenia (duń. *dansk nonsend ord korpus*; DANOK), a także testu zrozumiałości zdań w szumie (ang. *hearing in noise test*; HINT). Trening poprawił zdolność identyfikacji fonemów w szumie w przypadku samogłosek oraz spółgłosek zlokalizowanych w środku logatomów, jednakże nie w przypadku początkowych spółgłosek. Zaobserwowana poprawa była wciąż obecna po trzech miesiącach od zakończenia treningu. Nie wykryto wpływu treningu na próg zrozumiałości mowy w szumie (ang. *speech recognition threshold*; SRT) w teście zdaniowym (HINT).

Drugi rozdział przedstawia efekt treningu fonemów w szumie na wysiłek słuchowy niezbędny do zrozumienia testu zdaniowego (HINT), który został zmierzony za pomocą pupilometrii. Zmiany wielkości źrenicy zostały zarejestrowane, podczas gdy uczestnicy eksperymentu słuchali zdań zaprezentowanych w warunkach stosunku sygnału do szumu (ang. *signal-to-noise ratio*; SNR) odpowiadajacym zrozumiałości mowy na poziomie 50% (SNR50) oraz 80% (SNR80),

określonych na podstawie wyników sprzed interwencji. W obu warunkach eksperymentalnych analiza zmian wielkości źrenicy zasugerowała wzrost wysiłku słuchowego u obu grup słuchaczy, jednak zmiana była bardziej znacząca u osób, które ukończyły trening. W przeciwieństwie do progów zrozumiałości mowy (przedstawionych w pierwszym rodziale), wartości procentowe zrozumiałości mowy zmierzone przy stałym stosunku sygnału do szumu wskazały na poprawę jedynie w trudniejszym z dwu warunków eksperymentalnych (SNR50). Wyniki zademonstrowały, że pupilometria jest metodą pomiaru wrażliwą na skutki treningu nawet w sytuacji, gdy test zdaniowy (HINT) nie wykazał znaczącej popawy zrozumiałości mowy (SNR80). Wzrost wysiłku słuchowego mógł być związany ze zwiększoną motywacją do wykonania zadania (u obu grup) oraz zmianą w mechanizmach alokacji zasobów kognitywnych (u osób, u których przeprowadzono trening).

W trzecim rozdziale zaprezentowano projekt oraz implementację nowej metody muzycznego treningu słuchowego majacego na celu poprawę zrozumiałości mowy w szumie u użytkowników aparatów słuchowych. Trening synchronizacji ruchowej z pulsem utworu muzycznego został przeprowadzony w warunkach domowych za pomocą aplikacji mobilnej. W trakcie treningu nastąpiła niewielka poprawa konsystencji synchronizacji ruchowej, jednak nie zaskutkowała ona znaczącym polepszeniem percepcji pulsu muzycznego oraz zozumiałości mowy w szumie. Wydłużony i bardziej intensywny traning o indywidualnie dostosowanym stopiniu trudności może być niezbędny, aby otrzymać znaczące korzyści przy użyciu tej metody.

Rezultaty badań przedstawionych w tej pracy mogą zostać wykorzystane przy rozwoju przyszłych strategii treningu słuchowego wykorzystującego dźwięki mowy lub bodźce muzyczne, a także pomóc w selekcji metod pomiaru korzyści płynących z tych interwencji.

Acknowledgments

My PhD studies have been a long and challenging journey, which allowed me to learn a lot, not only about auditory training and science in general but also about myself. Finalizing this venture would not be possible without the substantial support of people around me, to whom I would like to address these acknowledgments.

First of all, I would like to thank my supervisors for sharing their experience and knowledge with me and also for a great dose of patience throughout the whole process: Maja Serman, for inspiring this project and your enthusiasm along the way, Jeremy Marozeau for your trust and advice, and Torsten Dau, for your feedback and understanding. It has been a rocky journey and I appreciate that all three of you supported me until the end. I would also like to thank Federica Bianchi and Sébastien Santurette, who trusted that I would be the right person to carry out this project and were in my supervision team at the beginning of this adventure. Even though our paths parted so quickly, I have no doubt that I would have gained a lot from working with you. Maybe we will have a chance to meet again in our professional lives.

I want to thank my officemates and colleagues from the House of Acoustics (Hearing Systems and Acoustic Technology). Specifically, I would like to acknowledge Rikke Skovhøj Sørensen for her enormous help with the experiments and for always being willing to proofread my writing in Danish (including the abstract of this thesis). Next, I would like to thank Florine Lena Bachmann for the fruitful cooperation on the joint part of our respective projects. It has been great to have your company. I greatly appreciate all the pupillometry-related discussions with Dorothea, Helia and Mihaela. Thank you for sharing with me your experience on this topic. I am grateful to Raul Sanchez-Lopez for valuable comments on my thesis and for always being such a kind and supportive colleague. I sincerely thank Michał Feręczkowski for his feedback on the thesis abstract, all sorts of good tips along the way and for helping me transport my equipment back and forth between Lyngby and Odense.

I would also like to acknowledge the role of our ’sisterly’ institution – Copenhagen Hearing and Balance Center. I would not have been able to recruit as many participants to my studies without this connection. I would also like to appreciate colleagues from WSAudiology: Filip Marchman Rønne for giving the green light to organize some help with finding volunteers for my studies in these difficult pandemic times, the audiologists Tina Vidanovska and Thit Eiberg Bak for their assistance in this process and Kaja Kallisch for sharing the experience from her project on auditory training. Special words of gratitude should be addressed to all the volunteers who invested so much time and effort in participating in my experiments. Their commitment was crucial to this project.

During my PhD studies, I had an opportunity to spend some time in the Institute of Clinical Research at the University of Southern Denmark (SDU) in Odense. I want to thank Professor Tobias Neher for giving me an opportunity to visit his lab. I would also like to acknowledge Rasmus Skipper from Odense University Hospital for his help with recruiting the participants and Anne Roslyng-Jensen for her assistance with the measurements. Big thanks to all the colleagues from SDU for making my stay with your research group such a nice experience, I really had a great time!

Finally, I would like to mention my family and friends, who were very important to me during these intense years. Many thanks to Queralt, Jola, Eliana, Łukasz, Paweł, Florian, Kamil and Demi for sharing the kitchen as well as our everyday joys and worries. Big hugs to my little choir that filled my Sundays with music. Finally, I cannot imagine finalizing this thesis without the support of my family: my talented sister Natalia, who drew the beautiful cover picture, and my parents, who have always been there for me.

Related publications

Journal papers

- Koprowska A., Marozeau J., Dau T. & Serman M. (*in preparation*). "The influence of phoneme-based auditory training on speech intelligibility in hearing-aid users."
- Koprowska A., Wendt D., Serman M., Dau T. & Marozeau J. (*in preparation*). "The influence of auditory training on listening effort in hearing-aid users - insights from a pupillometry study."
- Koprowska A., Serman M., Dau T. & Marozeau J. (*in preparation*). "The development of a home-based rhythmic training method to improve speech perception in hearing-impaired listeners."

Contents

1

General introduction

The majority of everyday social interactions requires the capability to hear and understand speech. The process of verbal communication rarely takes place in a perfectly quiet environment. Situations in which the speech uttered by the interlocutor reaches the listener's ear simultaneously with background noises are more prevalent in daily life. An example of a particularly demanding communication setting is when the conversation takes place in the presence of multiple other talkers – a situation often referred to as a 'cocktail party' scenario (Cherry, 1953).

Understanding speech under such demanding circumstances relies on the robust encoding of the basic acoustic features. For example, pitch information can be an important cue for the segregation of the target talker from the noisy background based on the fundamental frequency of the voice (Coffey et al., 2017a). A healthy auditory system exhibits tremendous fidelity in encoding the acoustic signals, even in noisy situations. While normal-hearing listeners do not usually struggle with selectively attending to the talker of interest in complex conversational settings, for persons with age-related hearing loss (*presbyacusis*) such scenarios can present insurmountable difficulties.

The primary cause of presbyacusis is age-induced changes in the auditory periphery, including loss of hair cells and synaptic connections. Perceptual consequences of presbyacusis include the loss of sensitivity to low-level sounds, especially at high frequencies. As a result, some of the information contained in the speech signal – especially low-energy components in the higher-frequency part of the spectrum, fall out of the audible range of a person with a hearing loss. Another consequence of presbyacusis is a reduced temporal and spectral resolution (Arlinger and Dryselius, 1990), which leads to a poorer encoding of the fine-grained sound features, even when the signal is amplified to compensate for audibility deficits.

Additionally, age-related hearing loss is associated with a decline in cognitive function (Wayne and Johnsrude, 2015), which contributes to difficulties in understanding speech in noise. The complex combination of auditory and cognitive factors makes age-related hearing loss a health condition that cannot be addressed with a simple treatment method. The deprivation of the sense of hearing adversely affects everyday communication and has severe negative consequences for the individual well-being. Hearing impairment has been associated with an increased level of fatigue (Alhanbali et al., 2017), social isolation (Mick et al., 2014) and depression (Cosh et al., 2019) as well as overall reduced quality of life (Ciorba et al., 2012). It is estimated that the total number of people who need clinical intervention due to hearing loss will reach 700 million by the year 2050 (WHO, 2021). With an increasing number of people who require audiological treatment, there is a continuous need for providing efficient solutions that will mitigate the negative impact of hearing loss on individuals' lives.

1.1 Aural rehabilitation for adults with hearing loss

1.1.1 Hearing aids

Currently, the most common treatment for hearing loss in adults is the provision of hearing aids. Hearing aids are medical devices that compensate for audibility deficits by providing a frequency-specific and level-dependent gain based on the individual audiometric profile. Additionally, some advanced features available in modern hearing aids, like noise reduction algorithms or directional microphones, facilitate selective listening in the presence of interfering background noise. A recent market survey indicates that hearing aids improve listening comfort across a variety of situations and have an overall positive impact on the quality of life (Picou, 2020). However, around a third of hearing-aid owners still report that they are not satisfied with the performance of their devices when trying to follow a conversation in noise (Picou, 2020).

Limited benefits of hearing aids on speech intelligibility in the presence of background noise partially stem from the fact that the amplification cannot restore the temporal and spectral resolution typical for an unimpaired auditory system. Moreover, hearing-aid users might not be able to effectively utilize all the speech information, even if this is fully audible. Hearing loss progresses gradually and it usually takes several years before the person affected by it seeks professional help. During that time, the listener adapts to the limited acoustic information and develops abnormal perceptual strategies that are observed even in experienced hearing-aid users (Calandruccio and Doherty, 2008; Varnet et al., 2019). Furthermore, if the listeners suffer from some degree of cognitive decline, the hearing aids cannot compensate for this crucial limiting factor that affects speech understanding.

1.1.2 Auditory training

Auditory training is a potential complementary strategy to alleviate the speech-comprehension problems in hearing-aid users by retraining perceptual and cognitive skills relevant for speech perception in the presence of background noise. Auditory training is defined as active listening to auditory stimuli with the goal to improve auditory skills (Henshaw and Ferguson, 2013a; Stropahl et al., 2019). Even though this definition emphasizes the perceptual character of such an intervention, some training protocols may also involve tasks oriented towards cognition (e.g., Sweetow and Sabes, 2006) or sensory-motor interaction (e.g., Whitton et al., 2017).

Formal auditory training allows the listener to practice focused listening in a controlled, risk-free environment where one does not have to fear embarrassment as a consequence of failing to understand the message (Boothroyd, 2010). Moreover, it creates an opportunity to provide feedback on a trial-by-trial basis to guide listeners towards seeking new strategies in the case of an error and reward them for success. Feedback on a larger timescale allows the listeners to track their progress, stimulating thereby interest and motivation. Auditory training allows for multiple task and stimulus repetitions necessary to consolidate the acquired skills. Since individual perceptual differences may affect the effectiveness of the training, performance-based adaptation of the level of difficulty ensures that the user is constantly challenged while maintaining a success rate that is high enough to prevent discouragement. Overall, auditory training can create a learning-stimulating environment, which cannot be achieved during listening through hearing aids in everyday situations.

Despite the potential advantages, auditory training is not commonly involved in the hearing rehabilitation for the individuals with age-related hearing loss. It is estimated that less than 10% hearing professionals offer auditory training to their patients (Li-Korotky, 2012).

1.2 An overview of auditory training methods

Numerous approaches to auditory training suitable for hearing-aid users have been proposed, and at the moment, there seems to be no consensus on which of them is the most effective one. The existing training paradigms employ, i.a., speech-recognition tasks utilizing speech material with varying degrees of complexity (ranging from phonemes to sentences or even audiovisual speech), combinations of auditory and cognitive tasks, music instruction, a blend of auditory tasks and counseling, etc. A thorough review of all these methods would be rather time-costly. Therefore, only speech-based and music-based approaches will be discussed in detail below since these two classes of training were addressed by the experimental work presented in this thesis.

1.2.1 Auditory training using speech material

Phoneme-based training

Auditory training using material without semantic meaning requires the listener to utilize spectrotemporal low-level speech cues, following the notion of retraining the usage of these cues in the signal delivered by the hearing aids. There are several training methods that adapted this approach to auditory training.

The program SchooLo (Serman, 2012) trains phoneme discrimination (both consonants and vowels) with or without background noise using a five-alternative choice task. The training protocol allows for varying the task difficulty by manipulating the signal-to-noise ratio as well as the talker's gender and speaking rate. The emphasis is put on multiple repetitions of the stimuli. This training method has been proven suitable for training in the laboratory settings with cochlear-implant listeners (Schumann et al., 2015), normal-hearing listeners (Schumann et al., 2016) and hearing-aid users (Schumann et al., 2018).

Another training method targeting consonant-in-noise perception only was presented by Stecker et al. (2006). In contrast to SchooLo, the method applies a trial-by-trial adjustment of the signal-to-noise ratio based on the user's performance. The training improved consonant identification scores in new and experienced hearing-aid users and the effect generalized to untrained talkers. A modified version, with an optimized performance-based adaptation rule, was later developed by Woods et al. (2015b). The study also reported substantial improvements in the identification of consonants in noise (especially those, for which hearing aids provided smaller benefits) but no significant effect on sentence recognition in noise was found.

While the two previously described approaches use recordings of real talkers, a slightly different approach is offered in the program Phonomena (Moore et al., 2005). The program trains discrimination of phonemic contrasts using artificially generated phoneme continua, where the dissimilarity between pairs of phonemes (for example /p/-/b/) is gradually changed. The effectiveness of this program was assessed in individuals with mild-to-moderate hearing loss who were not using amplification in daily life (Ferguson et al., 2014). The improved discrimination abilities were not accompanied by an improvement in off-task outcome measures of speech intelligibility. However, significant improvements of divided attention, working memory and self-perceived listening abilities were found.

In summary, the approaches based on nonsense speech material consistently provide substantial on-task improvements. There is, however, limited evidence on generalization to untrained sentence material in hearing-aid users, primarily due to the fact that none of these methods has been yet evaluated in a clinical trial targeting this particular population of hearing-impaired listeners. An interesting finding is that the training of relatively low-level perceptual skills appears to have some impact on cognitive functions, as demonstrated by Ferguson et al. (2014).

Word-based training

Another approach to auditory training is to target the recognition of meaningful speech units. A word-based training encourages the listeners to use their lexical knowledge without being provided with the context of the sentence. Considering that a relatively small subset of words constitutes the majority of the words used in spoken language (French et al., 1930), such training should be able to cover most of the lexical material used in daily conversation.

A computerized frequent-word auditory training for older adults with hearing impairment was developed by Humes et al. (2009). The selected set of stimuli (600 different words), which corresponds to 80-90% of most frequently occurring words in American English, spoken by multiple talkers and presented in background noise, is trained using a closed-set multiple-choice task with a trial-to-trial auditory and orthographic feedback.

The frequent-word training administered in a group hearing-aid users yielded significant improvement in recognition of the trained set of words, which was maintained for at least eight and a half months (Humes et al., 2019). However, no effect was found on untrained sentence material or self-report outcome measures.

Combined speech comprehension and cognitive training

The Listening and Communication Enhancement program (LACE; Sweetow and Sabes, 2006) is an example of a combined auditory-cognitive approach. LACE includes the recognition of speech under difficult conditions (time-compressed speech, speech in the presence of babble and competing talkers), tasks targeting auditory working memory, speed of processing and usage of contextual cues to fill in the 'missing words', as well as practical everyday communication tips. Saunders et al. (2016) evaluated the effectiveness of LACE in a clinical trial. The outcome measures were designed to represent the same challenging speech conditions that were targeted in the training. No significant effect on untrained tasks was found.

1.2.2 Musical training

Being involved in long term musical activities seems to have beneficial consequences for auditory perception and cognitive functions. Musicians outperform non-musicians in several auditory tasks, for example, pitch discrimination (Micheyl et al., 2006) or temporal gap detection (Mishra et al., 2014). They also achieved higher scores in the tests of executive functions than non-musicians (Zuk et al., 2014). Differences between musicians and non-musicians have also been observed with neurophysiological measures. For example, musicians have more robust neural representations of syllables presented in the presence of background noise than non-musicians (Parbery-Clark et al., 2009a).

An enhanced neural encoding of speech features and improved cognitive functions might contribute to better speech intelligibility in the presence of background noise. Some of the behavioral studies comparing speech recognition in noise in musicians and non-musicians indeed revealed a benefit of musicianship (Başkent and Gaudrain, 2016; Clayton et al., 2016; Parbery-Clark et al., 2009b; Swaminathan et al., 2015;) which was, however, absent in other studies (Boebinger et al., 2015; Madsen et al., 2017; Madsen et al., 2019; Ruggles et al., 2014). Those mixed results suggest that the musician advantage on speech-in-noise perception might be limited to particular conditions (see for review: Coffey et al., 2017b).

The possibility of the transfer of a musical training effect to the speech domain and it's potential role in aural rehabilitation is a topic of an ongoing discussion (McKay, 2021). One source of skepticism towards this approach is the correlational character of most documented associations between musical training and speech-in-noise understanding. Schellenberg (2015) argued that the observed trends might be driven by pre-existing differences rather than the training itself. Another argument against using musical training in rehabilitation is that any musician benefit (assumed that it is caused by the music-making and not genetic factors) on speech perception is achieved through life-long practice, often starting early in childhood. It remains unknown if musical training commenced at an older age can lead to comparable outcomes. Some recent studies suggest, however, that no life-long expertise is needed to obtain at least some degree of speech perception enhancement due to musical training in elderly individuals with no previous musical experience (e.g., Zendel et al., 2019). These results encourage further exploration of music-related options for aural rehabilitation.

The effect of instrumental practice and choir singing on speech perception in elderly listeners

Zendel et al. (2019) demonstrated that six months of piano lessons led to a significant increase in the percentage of correctly recognized words presented in a challenging noise condition in older individuals with normal hearing and mild hearing loss. The behavioral improvement was accompanied by an increased brain activity associated with an enhanced sensory-motor coupling. No such effects were observed in either of the two control groups: the active one (that regularly played video games for six months) and the passive one.

Dubinsky et al. (2019) examined the impact of ten-week long choir singing in a group of older individuals with normal hearing and mild hearing loss. Improvements in the strength of neural representation of fundamental frequency, pitch discrimination ability and speech-in-noise perception were found, suggesting an enhanced pitch processing as the factor underlying the benefits in the speech domain. The study's main limitation was that it involved only a no-contact control group that completed two inter-spanned testing sessions. Considering that choir singing involves a lot of interpersonal interactions, it cannot be ruled out that the observed improvements were, at least to some extent, driven by the beneficial effect of an increased social exposure rather than the musical activity itself.

The effect of audio-motor instruction on speech perception in elderly hearing-aid users

A music-inspired training method for older hearing-aid users was proposed by Whitton et al. (2017). The training program was designed to imitate the situation when musicians need to adjust the fine details of their instrument's sound (e.g., pitch) using subtle movements (e.g., changing the finger's position on a string). This action often requires selective listening to one's instrument while other ensemble members play simultaneously. The training participants were asked to continuously monitor and adjust the parameters of the sound (level, frequency and modulation rate) to a reference by moving the fingertip on the tablet surface while suppressing the background noise. The analogy between the training task and music playing was meant to elicit similar learning mechanisms, including the audio-motor interaction. The training improved sentence and digit recognition scores in conditions in which the benefit of hearing aids was limited. However, the effect disappeared without continued practice, as revealed by the measurements collected six weeks later.

Rhythmic training as an option for auditory rehabilitation

There are various options for music-based rehabilitative interventions which remain unexplored. Another potential alternative, audio-motor rhythmic training, has been suggested as a mechanism for improving speech processing by shaping the mechanisms of forming temporal predictions (Schön and Tillmann, 2015). Recently, it has been proposed that predictions based on temporal characteristics of the speech signal can enhance speech intelligibility (Yates, 2016). Training rhythmic skills emerges as another possibility for a music-based training intervention for older hearing-impaired listeners.

1.3 Methodological aspect of research on auditory training

The 'golden standard' to provide high level of evidence in research on auditory training is a double-blinded randomized controlled trial (McKay, 2021). In such a study design, the training outcomes are assessed with reference to a control group that receives a non-training intervention, which should provide a similar expectation of benefit. The randomization procedure should result in the two groups being equal in terms of characteristics that could determine the training's success (e.g., age, degree of hearing loss, cognitive ability). A double-blinded randomized controlled trial design rules out possible confounds caused by test-retest performance improvements, group differences as well as participants' and experimenters' bias. Numerous studies up to date have fallen short of this requirement (see for review: Henshaw and Ferguson, 2013a; Stropahl et al., 2019). Lack of sufficient high-quality evidence for the efficacy of auditory training has been identified as a major obstacle to its implementation in clinical practice (Henshaw and Ferguson, 2013a).

Characterizing the benefits of training requires the selection of appropriate outcome measures. The standard method of characterizing the benefits of auditory training is to administer a speech-in-noise test before and after the training and to compare results across these two points in time and experimental groups. The speech material used in different studies varies in complexity and amount of contextual cues, ranging from digits-in-noise and words-in-noise to low- and high-context sentences. There is also a substantial diversity in the types of background noises, including, for example, stationary speech-shaped noise, multi-talker babble and competing talkers. While a specific research question sometimes determines the selection of outcome measures, the lack of standardization constitutes a substantial challenge in the interpretation of the results and comparisons across different studies (Henshaw and Ferguson, 2013a).

Some researchers attempted to characterize the intervention outcomes using other methods, such as electrophysiological measurements (e.g., Anderson et al., 2013; Dubinsky et al., 2019), self-report questionnaires (Ferguson et al., 2014), pupillometry (Kuchinsky et al., 2014) and cognitive tests (e.g., Ferguson et al., 2014). Electrophysiological assessment and cognitive tests allow insights into the mechanisms that underlie the potential improvement in speech recognition. On the other hand, self-report questionnaires provide extremely valuable information on how relevant the training effects are from the patient's perspective.

Recently, there has been growing interest in using listening effort measures to assess the benefits of aural rehabilitation. Methods of quantifying listening effort include various self-report, behavioral and physiological measures (McGarrigle et al., 2014), which have been shown to provide complementary information to the results of speech-in-noise tests (e.g., Humes, 1999; Wendt et al., 2017). So far, only a few studies have attempted to characterize the impact of auditory training on listening effort (Henshaw and Ferguson, 2013b; Kuchinsky et al., 2014). Therefore, it is still unknown which of the available methods would be most suitable in the context of assessing the benefits of auditory training.

1.4 Overview of the thesis

This thesis investigates the efficacy of auditory training in improving speech perception in elderly hearing-impaired listeners who already use hearing aids. Two training approaches are considered in the dissertation: a speech-based program (phoneme-in-noise identification training) and a music-based program (rhythmic training).

In *chapter 2*, the influence of phoneme-in-noise identification training on speech intelligibility in elderly hearing-aid users is explored. The generalization of the training to an untrained sentence material is assessed with an active control group as a reference. The results of a follow-up test after three months are included to provide a better understanding of the retention of the training-related changes.

Chapter 3 focuses on how the phoneme-in-noise training affected listening effort in the same population of hearing-aid users. Task-evoked pupil responses recorded during the sentence intelligibility test are used to characterize the participants' cognitive load before and after the intervention. Possible training-induced changes in effort-allocation mechanisms are discussed based on the trends observed in the data.

Chapter 4 describes the design and implementation of a training program based on a rhythmic task. The results of an evaluation study in a group of elderly hearing-impaired listeners are presented. The impact of the training on speech intelligibility is discussed, and suggestions for further improvement of the method are provided.

Chapter 5 summarizes the main findings of the thesis and discusses the limitations and opportunities of the two training approaches. The perspectives for future research on speech- and music-based training programs as well as their potential clinical applications are considered in the context of the new findings provided in this thesis.

2

The influence of phoneme-based auditory training on speech intelligibility in hearing-aid users[a]

Abstract

Auditory deficits linked to sensorineural hearing loss commonly cause difficulties in understanding speech, especially in the presence of background noise. The benefits of hearing aids on speech intelligibility in challenging listening scenarios remain limited. Inspired by earlier investigations on auditory training, the present study investigated the effects of training of phoneme discrimination in background noise in hearing-aid users. Specifically, the efficacy of phoneme discrimination training on phoneme perception as well as on sentence intelligibility in untrained conditions were explored. Twenty 63-to-79 years old individuals with a mild-to-moderate sensorineural hearing loss and at least one year of experience using hearing aids received either a two-week training program or a control intervention. Tests of phoneme identification, using logatomes from Danish Nonsense Word Corpus (DANOK), and sentence intelligibility, using the Danish hearing in noise (HINT) test, were administered before and after the respective interventions as well as, for the training group only, at a follow-up visit after three months. It was found that the training provided an increase in phoneme identification scores for vowels and post-vowel consonants, but not for onset consonants. The post-training performance did not change until the follow-up visit, suggesting that the training-related improvement was retained over three months. No generalization effect to the untrained (sentence intelligibility) task was revealed.

---

[a] This chapter is based on Koprowska A., Marozeau J., Dau T. & Serman M. (*in preparation*).

Overall, the results of this study demonstrated a training-induced refinement of auditory perception at a phoneme level but did not provide evidence for the generalization to an untrained sentence intelligibility task.

2.1 Introduction

The primary role of hearing aids is to compensate for the loss of audibility by amplifying the acoustic signal based on individual hearing thresholds. However, several studies reported a limited benefit of amplification on speech intelligibility (e.g., Abavisani and Allen, 2017; Scheidiger et al., 2017; Woods et al., 2015a). One of the reasons for these findings is that the frequency- and level-dependent gain generally do not compensate for other deficits associated with a hearing loss, such as, e.g., a loss of auditory spectral and temporal resolution, which reduce the fidelity of the representation of fine acoustic features in the stimuli. Robust 'internal' representations of these 'low-level' features may be especially relevant for understanding speech in adverse conditions, e.g., in the presence of background noise (Coffey et al., 2017a). Additionally, the provided amplification of the acoustic signal through hearing aids may affect how listeners map the sensory input to corresponding internal linguistic representations.

Furthermore, years of listening with reduced sensitivity to sound before receiving the first hearing prosthesis lead to changes in how individuals utilize the information contained in the acoustic signal. People with hearing loss, even when provided with amplification, tend to apply different cue- and spectral weighting strategies than people without hearing deficits (Varnet et al., 2019; Calandruccio and Doherty, 2008). In both studies, these abnormal listening strategies were still present even after substantial experience with hearing aids. Wearing hearing aids in daily life was not found to be effective enough to relearn the usage of certain cues that are crucial for speech perception.

This lack of 'implicit learning' may be explained in light of the Reverse Hierarchy Theory (Ahissar et al., 2008). According to this theory, the perception in each sensory (e.g. auditory) domain is based on a hierarchy of low-to-high level representations, where the low-level representations encode basic stimulus parameters (e.g., spectral and temporal features) whereas the high-level representations correspond to phonological entities that carry an abstract/semantic meaning (e.g., words). The theory assumes that in the majority of daily-life scenarios only the high-level representations are easily accessible to the listener. Under certain circumstances, when high-level inputs are not informative enough (e.g. when the speech signal is highly degraded due to background noise), a backward search to more informative low-level inputs might be activated. This process, however, requires time and stimulus repetition. Furthermore, improved sensitivity to low-level stimulus parameters can only be achieved by a task- or cue-specific practice regime rather than by naïve perception.

Based on these considerations, the hypothesis of the present study was that hearing-aid users need practice that is targeted towards low-level representation of speech because re-learning at these processing stages does not occur spontaneously. For a given hearing-aid processing scheme, an appropriately designed training program could re-train the usage of certain speech cues. Such training could, for example, focus on phonological units, which do not carry any semantic meaning but can be assigned to a certain perceptual category, i.e. specific phonemes. The effects of perceptual training on phoneme perception have previously been investigated in several listener populations differing in their hearing abilities. A training-induced improvement of phoneme perception measures has been demonstrated in normal-hearing listeners (Schumann et al., 2016), cochlear implant (CI) patients (Schumann et al., 2015) as well as non-aided individuals with a mild-to-moderate hearing loss (Ferguson et al., 2014). Studies with listeners who used amplification in daily life yielded similar outcomes. Improvements of syllable identification scores (Stecker et al., 2006), consonant identification thresholds (Woods et al., 2015b) and phoneme discrimination thresholds (Henshaw and Ferguson, 2013b) after the training have also been reported in experienced hearing-aid users.

Successful on-task learning is a promising finding; however, it does not guarantee that the training will have implications for speech comprehension in everyday situations. This would require that the learning effects generalize to untrained tasks and conditions. Demonstrating the generalization effect is crucial to consider the training intervention for potential use in clinical settings. Currently, the mechanisms underlying the generalization and transfer of learning are not fully understood. Reverse Hierarchy Theory proposes that such transfer to other tasks or conditions would only occur if the trained cue is relevant in the novel context and can be used in a similar way (Ahissar et al., 2008). Thus, if the ultimate goal of the training is to improve speech intelligibility in the presence of background noise then training phoneme representations in noise might yield better chances of generalization than training them in quiet.

While assessing the generalization of the training, it is important to control for other factors than the training itself that can cause the improvement of the participants' performance. These factors include the effect of test-retest (procedural learning), the effect of social interaction and the effect of the expectation of improvement. Only a study design involving a group that receives a 'placebo' intervention (an active control group) allows controlling for all these factors, provided that both the active control group and the training group have equal expectations (Boot et al., 2013). In earlier studies on auditory training, this criterion was not always fulfilled as designs with a passive control group or a delayed training group were more common (Stropahl et al., 2019). The more recent studies by Whitton et al. (2017), Saunders et al. (2016) and Humes et al. (2019) included active controls while assessing the efficacy of training programs for hearing-aid users. However, none of these studies focused specifically on phoneme training.

The present study evaluated the efficacy of phoneme-in-noise training in experienced hearing-aid users. First, we hypothesized that the proposed training will improve the phoneme identification skills in noise in a group of experienced hearing-aid users. This would indicate that, for a given population, the intervention was successful in refining a specific skill in a specific context (background noise). Secondly, we investigated whether the training resulted in an improved performance using untrained speech material. Such a finding would indicate that the trained skill is useful in a novel context and would provide the basis for considering its applicability as a clinical intervention. To provide solid evidence for the presence or absence of generalization, we decided to involve an active control group, which has not been done in previous phoneme-training studies. The third objective of this study was to investigate if the potential effects of training persist over time. Such information would be particularly useful from a clinical perspective to know, for example, if the training should be repeated in regular time intervals to maximize its benefits (Stropahl et al., 2019).

2.2 Methods

2.2.1 Participants

Participants were recruited from the databases of DTU Hearing Systems in Lyngby, Audiological Clinic of Bispebjerg Hospital in Copenhagen and WSAudiology in Lynge, Denmark. The inclusion criteria were: 1) age 60-80 years, 2) Danish as a first language, 3) symmetric mild to moderate hearing loss – defined as better-ear pure-tone average (PTA) of thresholds for 0.5, 1, 2 and 4 kHz within the range 26-55 hearing level (dB HL), 4) symmetric hearing loss, i.e. an interaural difference between the PTAs no greater than 10 dB (Noble and Gatehouse, 2004), 5) at least one year of experience using hearing aids, 6) no cognitive disorders. Individuals identified as suitable candidates based on the information available in the databases were contacted according to the routine procedures of the respective institutions. Out of 24 individuals who expressed interest in the study and were assessed for eligibility, two did not meet the inclusion criteria and two discontinued the intervention due to COVID19-related restrictions. A total number of 20 participants completed the study. The detailed information about the participants can be found in Table 2.1.

2.2.2 Study design and procedure

The experiment was designed as a parallel-group randomized controlled trial. The study design is illustrated in Figure 2.1. Participants who were assigned to the intervention group (n = 10) received the phoneme-training program. This group will be referred to as the “training group”. Participants allocated to the “control group” (n = 10) were enrolled in another procedure (described in detail further below). For practical reasons, the phoneme-training program was ready to use earlier than the program for the control group. Therefore, the allocation to the groups depended on the order of the enrollment in the study: the first ten participants were assigned to the training group, and participants 11 – 20 were part of the control group. In the invitation letter, all participants received the same information about the objective of the investigation and were not informed about the type of intervention they received. Both groups completed the same number of visits in the research facility and had approximately the same amount of contact with the experimenters. After completing the study, individuals who had been assigned to the control group were informed about it

Table 2.1: Characteristics of the participants

No.	Age	Sex	DS score	PTA better ear [dB HL]	WRS [%]	HA brand	HA experience [years]
Training group							
1	75	M	18	52.50	80	Siemens	15
2	66	M	30	45.00	96	Siemens	5
3	79	M	25	50.00	84	Oticon	20
4	66	F	23	47.50	96	Beltone	8
5	73	F	-	43.75	88	Oticon	9
6	73	F	-	45.00	96	Widex	11
7	63	M	27	52.50	92	Siemens	6
8	79	M	20	32.50	92	Widex	3
9	70	M	-	36.25	92	Widex	12
10	73	M	37	40.00	92	GN	7
Mean	**71.7**		**24.88**	**44.50**	**90.8**		**9.6**
Control group							
11	70	M	24	35.00	80	Siemens	6
12	68	F	41	28.75	96	Phonak	5
13	79	F	-	37.50	96	Widex	7
14	80	M	39	33.75	100	Phonak	5
15	64	F	-	41.25	100	Widex	15
16	67	M	37	46.25	96	Widex	1.5
17	70	F	33	51.25	92	Widex	20
18	68	M	26	27.50	88	Widex	10
19	63	F	41	37.50	92	Widex	8
20	65	M	35	55.00	96	Widex	15
Mean	**69.4**		**34.5**	**39.38**	**93.6**		**9.25**

and received an offer to complete the phoneme training program.

The recruitment process, data collection and analysis were managed by one investigator. Thus, it was not possible to achieve double-blinding in this experiment. Nevertheless, a potential bias in scoring participants' responses in the sentence test was eliminated since a research audiologist who performed scoring did not know which group the subjects belonged to.

After an information meeting and signing a consent form, participants underwent an eligibility assessment. The condition of the ear canal and the tympanic membrane was checked with an otoscope. If an audiogram less than one year old was not available in the database, both air conduction (frequencies:

0.125, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8 kHz) and bone conduction (frequencies: 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4 kHz) thresholds were measured. Next, the participants were screened for mild cognitive impairment using a Danish version of the Montreal Cognitive Assessment questionnaire (Nasreddine et al., 2005). The research audiologist who administered the questionnaire spoke Danish as a native language and had completed a required certification. A cutoff value of 23 out of 30 points was used as an inclusion criterion (Carson et al., 2018).

The primary outcome measures were phoneme identification scores obtained using the Danish Nonsense Word Corpus (DANOK; Nielsen and Dau, 2019) and speech recognition thresholds (SRTs) measured using the Danish hearing in noise test (HINT; Nielsen and Dau, 2011). Both were administered prior to and after receiving the allocated intervention.

Additionally, the participants completed a working memory test (auditory digit span test; Wechsler, 2008) and a monosyllabic word recognition test (WRS – word recognition score) before the intervention. Due to technical problems, the auditory digit span scores were not measured for three participants in the training group and two participants in the control group. The six visits comprising one of the allocated interventions were distributed equally over two weeks with two- to three-day intervals. The pre-training assessment took place no more than one week before the beginning of the intervention and the post-training assessment took place no more than one week after the intervention was completed. This schedule was observed with minor exceptions enforced by the limited lab capacity, the participants' availability and the epidemiological situation. The participants from the training group were invited to an additional follow-up visit three months after the post-training assessment but three of the appointment were canceled due to COVID19-related lockdown.

The experiment was approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). All participants received a nominal reimbursement.

2.2.3 Training

Training program

The training intervention was based on the SchooLo phoneme training program (Serman, 2012). SchooLo is a PC-based tool intended both for clinical studies and for usage at home by hearing-impaired individuals, in its original

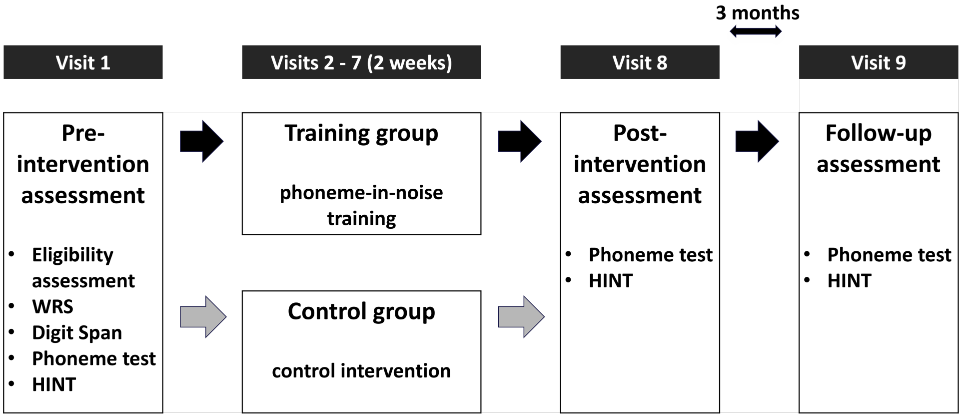

Figure 2.1: Study design. Participants completed eight (control group) to nine (training group) visits. The eligibility assessment and tests of word recognition in quiet (WRS) and auditory working memory (digit span) were performed during the first visit. Outcome measures of phoneme identification and sentence intelligibility were obtained upon the first (pre-intervention), the eighth (post-intervention) and the ninth (follow-up) visit. During two weeks between the first and the eighth visit, the participants received the allocated intervention.

version aimed at the German-speaking population. It uses speech material that consists of consonant-vowel-consonant (CVC) or vowel-consonant-vowel (VCV) tokens (e.g., /m**i**mm/, /e**b**e/, /a**ff**a/ etc.), where the middle phoneme is always the target vowel or consonant, which can appear in various coarticulation environments. SchooLo allows for using a variety of talkers with or without background noise, multiple repetitions of the stimulus, providing both visual and auditory feedback as well as adjusting the level of difficulty to avoid tiredness and maintain the user's focus and motivation. The program has been successfully used with normal-hearing listeners (Schumann et al., 2016) and CI patients (Schumann et al., 2015). Hence, Schoolo was considered to fulfill the key criteria and was used, after some modifications, also in the present study . The modifications concerned replacement of the speech material and change in the rules for adjustments of the difficulty level, as will be explained below.

Training material

The speech material used in the present study was the Danish Nonsense Word Corpus (DANOK; Nielsen and Dau, 2019). The corpus consists of logatomes that account for a large number of possible combinations of phonemes existing in the Danish language. It comprises recordings of four talkers (two male and two female) as well as a stationary speech-shaped noise matching the long-term spectrum of each talker. The main advantages of using DANOK are its 'nonsense' character as well as the possibility to train a variety of coarticulation effects using multiple talkers and maskers.

In DANOK, all logatomes are built according to the template CVC/i/, where the position of the target depends on whether the logatome was intended for measuring consonant (C-word) or vowel (V-word) recognition. The final /i/ sound was added to all speech tokens to achieve the nonsense character. In C-words (e.g., /dagi/, /funi/), one of 15 consonants /p t k b d g m n l f v s r h j/ can appear in the first position (C1) and one of 12 consonants /p t k b d g m n l f v s/ in the post-vowel position (C2). One of three vowels /ɑ e u/ can appear in between C1 and C2 (where ɑ is pronounced as in the beginning of a Danish word *arbejde*). In the training material, 14 alternatives were included, with the omission of /h/, which in the validation study (Nielsen Dau, 2019) turned out to be particularly susceptible to floor effects. In V-words (e.g., /bani/, /våbi/), one of nine vowels /i e a ɑ y ø u o å/ can be used as a target V-sound and one of the three consonant sounds: /b, n, v/ can appear on the onset and medial position.

Six training lists, corresponding to six planned training sessions, were combined for each possible target position (C1, C2 and V), resulting in 18 training lists in total. The selection of the logatomes to be included in the training lists was aimed at maximizing the variety of stimuli while providing an equal number of repetitions of each trained phoneme. The requirements were that 1) each target phoneme should appear twice within every training list, 2) the number of possible consonant-vowel (CV) or vowel-consonant (VC) combinations for each target phoneme should appear the same number of times throughout the whole training, and 3) none of the target phonemes should be repeated within the same logatome throughout the whole training (e.g., /badi/ could appear once in C1 list and once in C2 list but not twice in C1 or C2 list).

The recordings of two talkers - one male (M1) and one female (F2) - and corresponding noise files were used in the training. Within each training list, half of the logatomes were spoken by a male and half by a female talker.

Training procedure, interface and feedback

The training procedure was self-administered and self-paced. Participants communicated with the program with the help of a touch screen. The participant started the training block by pressing a Start button. Each trial consisted of four phases:

1. Stimulus presentation
2. Response alternatives displayed (*Repeat* button enabled)
3. Participant's response
4. Feedback (*Repeat* and *Continue* buttons enabled)

The items of the training list were presented in randomized order. After the audio file was played, five response alternatives were shown on the screen: the correct response and four competing logatomes that differed from it with one letter at the position of the target phoneme. The interface is shown in Figure 2.2 in panel A. The competing logatomes were individualized per subject and selected based on the four most frequent confusions registered in the phoneme test during the pre-training assessment. Simultaneously with the response alternatives, a *Repeat* button was enabled, and the participant could listen to the target logatome multiple times before selecting the response.

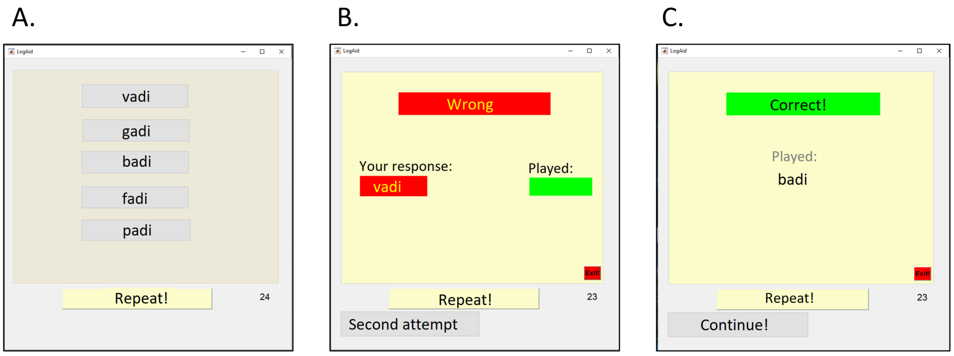

Figure 2.2: Training interface. A. After the first stimulus presentation, five response alternatives were displayed. The *Repeat* button could be used to replay the stimulus multiple times. B. Feedback message after an incorrect response (e.g., /vadi/ instead of /badi/) The *Repeat* button could be used to replay the auditory feedback (both the logatome incorrectly selected by the user and the correct one) multiple times. C. Feedback message after a correct response. The *Repeat* button could be used to replay the auditory feedback (the correctly selected logatome) multiple times.

For each item of the training list, three attempts were allowed. The first attempt was the initial presentation of the logatome. If the listener provided a correct response, the program proceeded to the next list item. In case of an incorrect response, the program proceeded to a second attempt with increased salience of speech cues. In the case of another incorrect response, the program proceeded to a third attempt and the salience of speech cues was further increased.

Schoolo offers flexibility in how better salience of speech cues in the consecutive attempts can be achieved. In the previous study with normal-hearing listeners (Schumann et al., 2016), the signal-to-noise ratio (SNR) was increased in 3-dB steps (-3, 0 and +3 dB). In the experiment with CI listeners (Schumann et al., 2015), the speech was presented in quiet and the changes concerned the gender of the talker (from female to male), the speech rate (from normal to clear speech) and the test material (from CVC token to a word with embedded target phoneme).

In the present study, the noise level during the second attempt was decreased by 2 dB compared to the first attempt. During the third attempt, the logatome was presented in quiet. If the user's third response was again incorrect, the correct response was displayed, and the program proceeded to the first attempt of the following list item.

Each participant's response was followed by feedback. First, a message *Correct* or *Wrong* was displayed. In case of a correct response, the target logatome was presented on the screen and via the loudspeaker (Figure 2.2 panel C). If the response was incorrect, the first logatome that was selected by the user was presented, both written and played from the loudspeaker (Figure 2.2 panel B). Then, the target audio file was presented again for comparison but was not revealed yet to the user (unless the wrong response was given at the third attempt). After the feedback was presented, the *Repeat* button was activated again. The participant could use it to replay the auditory feedback several times before pressing the *Continue* button.

The procedure was repeated until all logatomes from the training list were played. Then, a congratulation message was displayed on the screen and the training block ended.

Stimulus presentation

The training stimuli were presented via a loudspeaker calibrated with speech-shaped noise at a sound pressure level (SPL) of 65 dB at a distance of one meter. At the beginning of the first session, the SNR for all participants was +4 dB for the consonant blocks (C1 and C2) and 0 dB for the V block. These were the same values as those used in the respective subsets of the phoneme test during the pre- and the post-intervention assessment (as described further below).

Two types of adjustments of the SNR values during the training were introduced in this study, which were not present in the original Schoolo. First, the SNR was adjusted according to the listener's performance. The result was evaluated every fifth trial. If a person responded correctly five times out of five trials, the SNR was lowered by 2 dB. If a person responded correctly four times, the SNR was decreased by 1 dB. If three responses were correct, the SNR remained unchanged. Likewise, when the number of correct responses within five trials was two, the SNR was increased by 1 dB. If there were fewer than two correct responses, the SNR was increased by 2 dB. The SNR upon the end of the block was set as an initial value for the next session (for the block dedicated to the same target position).

The second type of adjustment was applied only in the consonant blocks. The intention was to compensate for differences in identification thresholds between different consonants in speech-weighted noise. It has been shown that some of the consonants, especially those with high-frequency components, are much less affected by masking in speech-shaped noise due to their spectral content (Phatak and Allen, 2007; Woods et al., 2010). A set of consonants with lower identification thresholds was selected based on data from the DANOK evaluation study (Nielsen and Dau, 2019) and included: /f j k r s t/. If any of those consonants appeared on the target position, a correction of –4 dB SNR was added on top of the manipulations described in the previous paragraph. Both types of adjustment were used to define the SNR value corresponding to the first presentation of the training item (i.e. during the first attempt).

The noise started three seconds before the onset and finished one second after the offset of the logatome. The level of speech signal was roved between 65 and 71 dB SPL. In principle, the level of the noise was adjusted to achieve the desired SNR. However, this sometimes resulted in high noise levels, which could lead to tiredness and annoyance. Therefore, if the computed level of the noise exceeded 70 dB SPL, the noise was set to 65 dB SPL and the speech was adjusted to achieve the desired SNR.

Setup and procedures

Half of the participants completed the training sessions in a room designed according to IEC 268-13 standard (suitable for listening tests on loudspeakers) at the Technical University of Denmark, Lyngby. The other half trained in a measurement booth in the Audiology Clinic of Bispebjerg Hospital, Copenhagen. The participants were seated one meter in front of the loudspeaker and always performed training while wearing their own hearing aids. The experimenter was present in the room, available to provide help or to answer questions. Before the first visit, a more extensive introduction to the training was provided. Participants were encouraged to use the *Repeat* button as frequently as they wanted.

Each training session consisted of three mandatory blocks dedicated to three different target positions. The order of blocks was balanced across the sessions. Before each block, information about which target position is trained (C1, C2 or V) was provided. The participants were offered a short break after each training block.

The time of each visit was limited to 1,5 hour and all participants were able to complete three blocks within the given time. Eight participants completed all six planned sessions. Two other participants (No. 2 and 10) completed only five out of six planned sessions due to technical problems with their hearing aids. Because of a mistake of the experimenter, the initial SNR in the first vowel training session was set to a wrong value (4 dB instead of 0 dB) for participant no. 2. This affected the SNR adjustment during the whole training process and might have produced misleading results. Therefore, participant no. 2 was not included in the analysis of the training progress when the V target was concerned.

2.2.4 Control intervention

The control intervention was the behavioral part of a procedure that was developed to record electrophysiological responses to continuous speech (Bachmann and Hjortkjær, in preparation). During each visit, the participants listened to 50-seconds long excerpts of audiobooks: *Fluernes Herre* (*Lord of the Flies* by William Golding) and *Politikens Store Eventyrbog* (an anthology featuring fairy tales by Hans Christian Andersen, Brothers Grimm and Charles Perrault), read by a male and a female talker, respectively. After each excerpt, the participant was asked to answer a few multiple-choice questions about the content of the story. The duration was 1,5 hour per session. The stimuli were delivered via insert phones and the audiogram-based amplification using the 'Cambridge formula' (Moore and Glasberg, 1998) was provided, except for the fifth session where the speech was not amplified.

2.2.5 Tests administered prior to the intervention

Two tests administered only during the pre-intervention assessment measured monosyllabic word recognition scores in quiet and auditory working memory span. The respective purposes of these tests were to assess the participants' eligibility for the study in terms of a functional benefit of the hearing aids and to characterize their memory span, which is closely linked to the ability to understand speech in noise (Akeroyd, 2008).

Word recognition score (WRS)

The monosyllabic word recognition test was administered to characterize the listeners' ability to understand speech in quiet while wearing hearing aids. A list of 25 monosyllabic words from DANTALE I corpus (Elberling et al., 1989) was presented through a loudspeaker placed one meter in front of the participant. The number of correctly repeated words was registered and a monosyllabic word recognition score in % correct was computed. A WRS score below 75% would be considered as an indication of a low benefit of hearing aids (caused by poor fitting or other unknown issues) and would result in excluding the participant from the study.

Auditory digit span test

The digit span test (Wechsler, 2008) is one of the most common methods to measure auditory working memory capacity. It was included in the test battery to characterize the cognitive ability of the participants, which could be used in the analysis as a predictor of training benefits. In this study, a computerized version of this test was used. The participant was seated in front of a screen with headphones on. Written instructions were displayed on the screen before the procedure started.

During each trial, the participant was presented with a sequence of digits spoken in Danish by a male voice. After the whole sequence was presented, the user was prompted to type the sequence on a keyboard. After typing the digits, the participant had to confirm the responses and could proceed to the next trial. The length of the sequence could vary from two to eight digits and two sequences of the same length were presented directly one after another. The length of the sequence was increased every second trial. The procedure was terminated after the participant failed to provide the correct response for two consecutive trials of the same length.

The test was administered in two blocks. During the first block, the task was to recall digits in the same order as they were presented (‘Digit Span Forward’ task). In the second block, the task was to recall digits in a reversed order (‘Digit Span Backward’ task).

The following scoring method was used in each trial (Blackburn and Benton, 1957): two points were given if all items were recalled in the correct order; one point was given if all items of the sequence were recalled but in an incorrect order; otherwise, no points were given for the trial. The final score was computed as a sum of scores for the Forward and the Backward task.

The participants took off their hearing aids before putting on the headphones. All stimuli presented during the test were spectrally shaped according to the ’Cambridge formula’ (Moore and Glasberg, 1998) based on the audiometric thresholds at nine frequencies (0.125, 0.25, 0.5, 1, 2, 1.5, 2, 4, 6, 8 kHz), averaged across the ears. Before the test started, a sample sequence was presented and the participants could adjust the output volume to a comfortable level.

2.2.6 Outcome measures

The outcome metrics of speech intelligibility administered before and after the intervention were: phoneme identification scores - measured with DANOK - and speech recognition thresholds (SRTs) obtained in HINT.

Phoneme identification test

Logatomes from the DANOK corpus were used to assess the phoneme identification in the presence of background noise. An 84-item consonant list and a 54-item vowel list were compiled from the corpus. The number of repetitions of each phoneme and combination of phonemes was kept as close as possible to those used in Nielsen and Dau (2019), such that all of them were equally represented to the greatest possible degree. In the consonant list, each of the C1 targets appeared six times and each of the C2 targets - seven times. Each possible CV and VC combination with three vowels (/a u i/) was repeated twice for each C1 phoneme and two to three times for each C2 phoneme. In the vowel list, each target alternative appeared nine times throughout the test and the number of every possible combination with three consonants (/b n v/) was equal.

The logatomes in the test lists were spoken by a female talker F1, which was not used during the training. Logatomes were presented in stationary noise matching the frequency magnitude spectrum of the talker's voice. The noise started three seconds before and ended one second after the presentation of the logatome. The level of the target was fixed at 65 dB SPL. The test lists were administered at fixed SNRs, following the approach taken in the DANOK validation study (Nielsen and Dau, 2019). The consonant list was presented at +4 dB SNR and the vowel list was presented at 0 dB SNR. These values were selected to achieve a good sensitivity of the test and to avoid floor or ceiling performance.

The stimuli were presented via a loudspeaker placed at a distance of one meter in front of the listener. The participants were provided with a touchpad and were asked to begin the procedure by pressing the *Start* button, whenever they felt ready. After the first stimulus presentation, the target word appeared on the screen with empty spaces instead of the target phonemes. All response alternatives were displayed underneath. In the consonant list, the response alternatives for the onset consonant were displayed on the left side and the response alternatives for the medial consonant were displayed on the right side on the screen and visibly separated from each other. The participants were informed that they would not have a possibility to repeat the sound and that they must provide a response even if they need to guess. The participants decided when to proceed to the next trial by pressing the *Next* button.

Before the actual test started, a practice list of 30 items was administered at +10 dB SNR to familiarize the participants with the procedure, and to make sure that all of them were able to use the touchpad.

Sentence intelligibility

Sentence intelligibility was measured using the Danish version of the Hearing-In-Noise Test (Nielsen and Dau, 2011). Two lists of 20 items spoken by a male talker were presented in stationary background noise with a frequency magnitude spectrum matched to that of the speaker's voice. The noise started one second before and ended one second after the presentation of the sentence. The level of the noise was fixed at 65 dB SPL. The level of the speech was initially set to be equal to the level of the noise and then changed adaptively according to a staircase procedure to define the speech recognition threshold (SRT), i.e. the SNR, at which the participant correctly repeated 50% of the test items (Brand and Kollmeier, 2002). A correct word scoring procedure (see e.g., Bianchi et al., 2019; Wendt et al., 2017) was used, i.e. the magnitude of the change of the speech level depended on the number of correctly repeated words and varied between +2 dB (zero correctly repeated words out of five) to -2 dB (five correctly repeated words out of five) in 0.8 dB steps. These numbers were doubled for the first four sentences.

The stimuli were presented via a loudspeaker placed at a one-meter distance in front of the listener. The participants were wearing their hearing aids during the test. They were instructed to repeat each sentence as precisely as possible and encouraged to guess if uncertain. Responses were scored by a native Danish-speaking audiologist.

2.2.7 Analysis

Separate linear mixed-effects models implemented in R (R Core Team, 2020) in *lmer* package (Kuznetsova et al., 2017) were used for each outcome measure: C1, C2 and V identification scores as well as SRTs. Each model included fixed factors of *Visit* (*pre, post*) and *Group* (*training, control*) as well as a *Visit x Group* interaction, and a random *Participant* factor. A backward elimination of non-significant terms was performed on each model. In case of a significant *Visit x Group* interaction, planned comparisons using the least-squares means approach (Lenth, 2016) were performed. This allowed for identifying significant differences between the relevant levels of the interaction (i.e. differences between the *pre* and *post* levels of *Visit* within each level of *Group*).

Additionally, a correlational analysis was performed to investigate if participants' characteristics could be used to predict training outcomes. Differences between post- and pre-training scores were calculated for those of outcome measures, for which a mixed-effect model revealed a significant *Visit x Group* interaction and for which subsequent post-hoc analysis showed a significant difference between pre-training and post-training levels. Correlations between these differences and three predictor variables (age, PTA and auditory digit span score) were computed.

An analysis of retention effect included only those outcome measures that revealed significant training effect. The normality of the samples was investigated using quantile-quantile plots and a Shapiro-Wilk test. Due to inconclusive outcomes, resulting from a very small samples size (n = 7), a non-parametric paired one-tailed Wilcoxon signed-rank test was chosen to compare the post-training and follow-up scores. A rejection of the alternative hypothesis that the median difference between the samples is not equal to zero would indicate that the training effect did not diminish over time.

To quantify an improvement in performance through the training, the mean SNR value was computed for each training session and participant. Then, a mixed-effects model with a fixed factor of *Session* (1-6) and a random factor of *Participant* was fitted. A post-hoc analysis was performed to investigate if there was a significant difference between the first and the last session.

2.3 Results

All participants completed the assigned activity program and the planned (two to three) measurement visits unless prevented by the epidemiological situation. None of the participants expressed the desire to withdraw from the study.

2.3.1 The effect of the intervention on phoneme identification scores

Figure 2.3 shows the pre-intervention (blue bars) and post-intervention (red bars) identification scores for the onset consonant C1 for the participants 1 to 20 as well as the group means (on the right). The top panel shows the results for the training group and the bottom panel shows the results obtained with the control group. The analysis revealed no significant *Visit x Group* interaction ($F(1,18) = 0.9$, $p = .35$), indicating that there was no difference between the pre- and post-intervention scores across the groups. The main effects of *Visit* ($F(1,18) = 0.1$, $p = .75$) and Group ($F(1,18) = 0.09$, $p = .76$) were not significant either.

In Figure 2.4, the pre-intervention (blue bars) and post-intervention (red bars) identification scores are presented for the vowels (V) for the participants 1 to 20 together with the group means (on the right). The top panel shows the results for the training group and the bottom panel shows the results for the control group. The analysis showed a marginally significant *Visit x Group* interaction ($F(1,18) = 3.92$, $p = .06$) as well as significant main effects of *Visit* ($F(1,18) = 8.48$, $p < .01$). The main effect of *Group* ($F(1,18) = 0.48$, $p = .49$) was not significant. A post-hoc analysis revealed a significant ($t = -3.46$, $p < .01$)

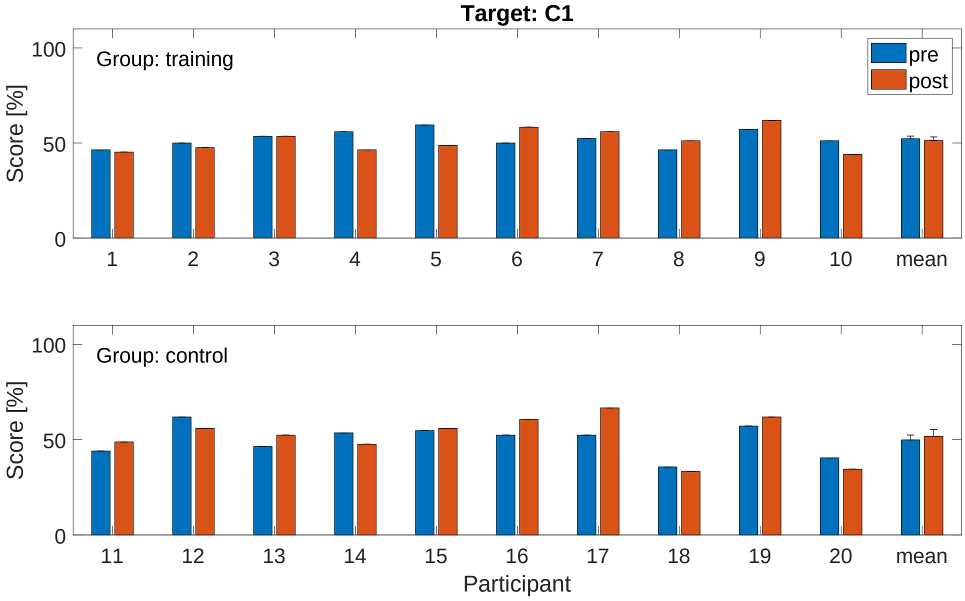

Figure 2.3: Pre-intervention (blue bars) and post-intervention (red bars) onset consonant (C1) identification scores in % correct. Results for the training group are shown in the upper panel and results for the control group are represented in the bottom panel. Numbers 1-20 correspond to the individual participants. The mean scores for each group are shown on the right. Errorbars represent one standard error of the mean.

difference between the pre- and post-intervention scores (*Mean (M) = 7.78 pts.*) only for the group that received the training.

Figure 2.5 shows the pre-intervention (blue bars) and post-intervention (red bars) identification scores for the post-vowel consonant C2 for the participants 1 to 20 as well as the group means (on the right). The top panel shows the results for the training group and the bottom panel represents the results for the control group. There was a significant *Visit x Group* interaction ($F(1,18) = 6.72, p = .018$) as well as significant main effects of Visit ($F(1,18) = 6.27, p = .02$). The Group effect ($F(1,18) = 0.01, p = .91$) was not significant. A post-hoc analysis revealed that the difference between the pre- and the post-intervention C2-scores was significant for the training group ($M = 6.9$ *pts.*, $t = -3.6, p < .01$).

Differences between the post-training and pre-training scores were calculated for the targets C2 and V. Pearson correlations between these differences and the three predictor variables (age, PTA and auditory working memory) were computed. There were no significant correlations between the predictors and the individual improvements in C2 and V identification scores.

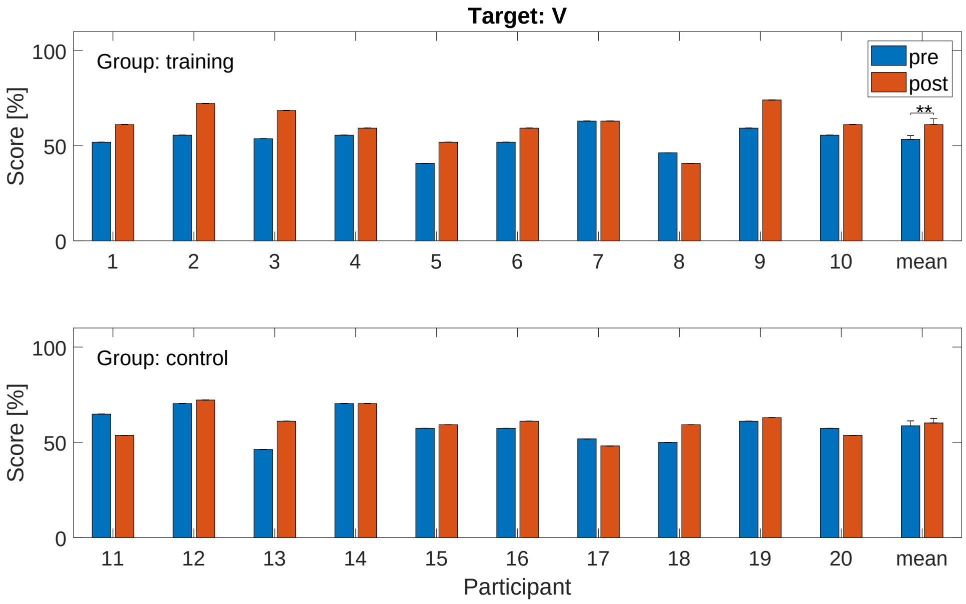

Figure 2.4: Pre-intervention (blue bars) and post-intervention (red bars) vowel (V) identification scores in % correct. Results for the training group are shown in the upper panel and results for the control group are represented in the bottom panel. Numbers 1-20 correspond to the individual participants. The mean score for each group is shown on the right. Errorbars represent one standard error of the mean. Asterisks indicate a significance level ** p<.01.

2.3.2 The effect of the intervention on sentence intelligibility

Figure 2.6 shows pre- and post-intervention SRTs for the training group (left panel) and the control group (right panel). The average change in SRT (the mean of differences between the post- and pre-intervention scores) was -0.7 dB for the training group and -0.3 dB for the control group. For eight out of ten participants in the training group, a negative difference in SRT between the post- and the pre-intervention results was observed, i.e. speech intelligibility improved following training. In the control group, half of the participants showed lower SRTs after the intervention than before.

The analysis showed a significant main effect of *Visit* (*F(1,18) = 4.49, p < .05*). Neither the main effect of *Group* (*F(1,18) = 2.81, p = .11*) nor *Visit x Group* interaction (*F(1,18) = 0.76, p = .39*) were significant.

2.3.3 The retention of the training effects

Figure 2.7 shows individual and mean phoneme identification scores obtained at the pre-training (blue bars), post-training (red bars) and follow-up (orange

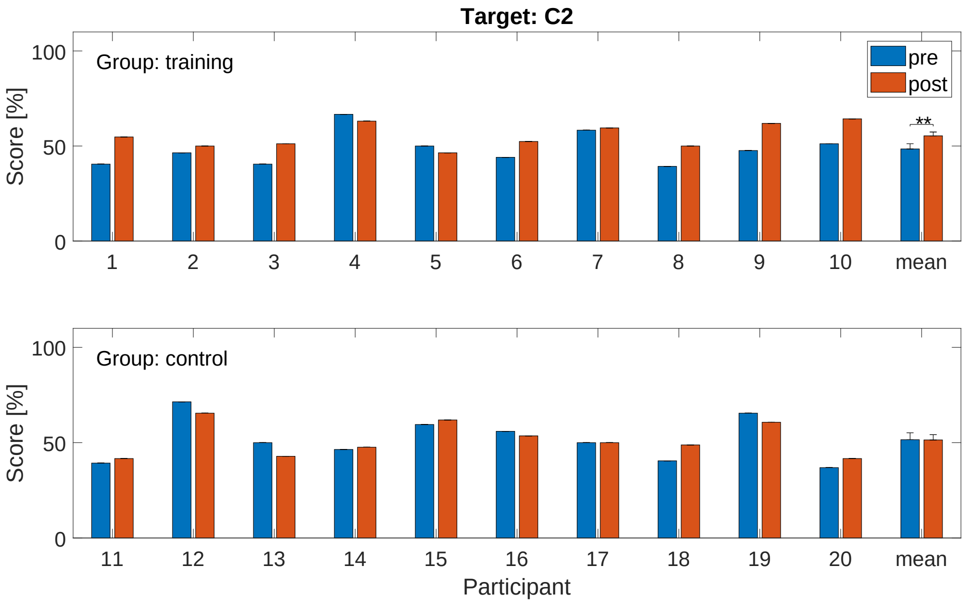

Figure 2.5: Pre-intervention (blue bars) and post-intervention (red bars) post-vowel consonant (C2) identification scores in % correct. Results for the training group are shown in the upper panel and results for the control group are represented in the bottom panel. Numbers 1-20 correspond to the individual participants. The mean score for each group is shown on the right. Errorbars represent one standard error of the mean. Asterisks indicate a significance level ** p<.01.

bars) visit for the C1 (top), V (middle) and C2 (bottom) targets. The individual results for those seven participants who completed all three visits and the mean results (on the right) are shown. All scores are expressed as % correct.

For target C1 (top panel), there was no indication of an improvement (pre-training: *mean (M) = 53.23%, standard deviation (SD) = 4.65*; post-training: *M = 51.70%, SD = 6.41*; follow-up: *M = 51.36%, SD = 8.14*). Regarding the V target (middle panel), the participants scored higher at the follow-up visit (*M = 60.05%, SD = 10.74*) than at the pre-training visit (*M = 52.65%, SD = 5.84*). Even though the scores at the follow-up visit were slightly below those at the post-training visit (*M = 63.76%, SD = 8.06*), they still deviated from the values obtained before training. For target C2 (bottom panel), on average, higher scores were obtained at the follow-up visit (*M = 55.10%, SD = 5.90*) than at the pre-training visit (*M = 47.96%, SD = 8.98*). Notably, the score from the follow-up visit did not deviate much from the post-training (*M = 54.25%, SD = 6.18*) result. The Wilcoxon signed-rank test showed that the difference between follow-up and post-training performance was non-significant, both for C2 ($p = .63$) and V target ($p = .09$).

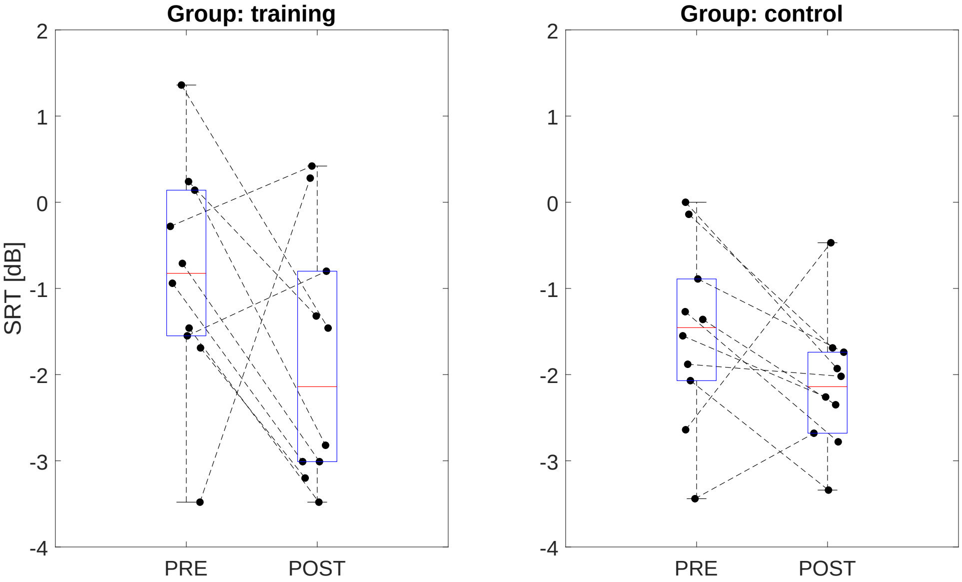

Figure 2.6: Pre-intervention and post-intervention SRTs for the training group (left panel) and the control group (right panel). The medians are indicated by the horizontal red lines. The 25^{th} and 75^{th} percentiles are represented by the bottom and the top edge of the box respectively. The whiskers indicate the most extreme observations not considered outliers. Individual data points are represented by the black dots. Results for the pre-intervention and the post-intervention condition obtained from the same listener are indicated by the connecting dashed lines.

2.3.4 The effect of training across the sessions

The session scores averaged across the participants ranged from 59.6% to 65.7% for C1, from 57.8% to 66.3% for C2 and from 63.2% to 73.9% for V. These scores were computed based on the first response attempts (i.e. correct responses provided at the second or the third attempt for a given logatome are not considered here). Figure 2.8 shows the mean SNR value computed for each training session and participant (individual data) for C1 (top panel), V (middle panel) and C2 (bottom panel). The analysis revealed that the effect of session was significant in the case of C1 ($F(1,5) = 4.42, p < .01$) and V ($F(1,5) = 12.97, p < .001$) but not C2 ($F(1,5) = 0.8, p = .0.55$). The post-hoc analysis showed that there were significant differences between the first and the last session for target C1 ($t = 3.75, p < .01$) and for target V ($t = 6.57, p < .0001$).

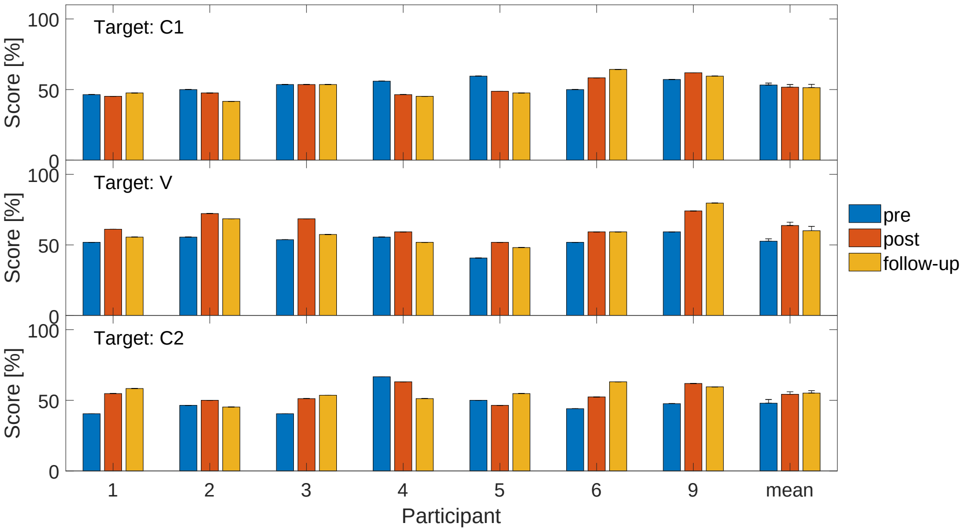

Figure 2.7: Phoneme identification scores obtained at pre-training (blue), post-training (orange) and follow-up (yellow) visit. Results are shown for target C1 (upper panel), V (middle panel) and C2 (bottom panel). Numbers 1-6 and 9 represent the individual participants. The mean scores are shown on the right. Errorbars represent one standard error of the mean.

2.4 Discussion

The present study investigated whether auditory training can further improve speech perception in individuals who have been using hearing aids. The proposed training focused on phoneme identification in background noise. The parameters of the training were aimed at providing an adequate level of difficulty such that the participants remained both challenged and motivated. The hypotheses were that the training (i) would improve phoneme identification in noise, (ii) would provide a benefit also with untrained, higher-context speech, and (iii) that any improvement would still be present three months after the training. Additionally, the participants' performance across the six training sessions was analyzed and a correlational analysis between significant training outcomes and the participants' age, degree of hearing loss and auditory working memory score was provided.

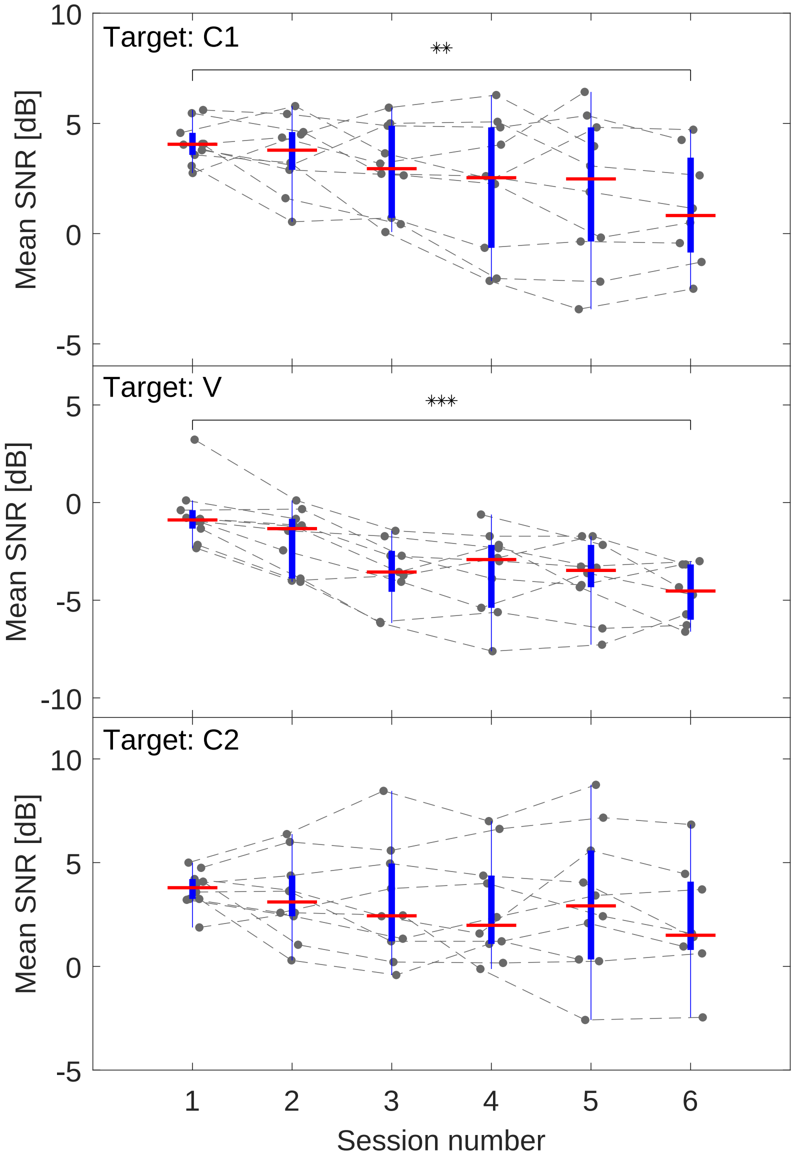

Figure 2.8: Mean training SNR plotted against the number of training session dedicated to the target C1 (upper panel), V (middle panel) and C2 (bottom panel). The boxplots represent group data. The medians are indicated by the horizontal red lines. The 25th and 75th percentiles are represented by the bottom and top edge of the box respectively. The whiskers indicate the most extreme observations not considered outliers. Individual data points are represented by the grey dots. Results for different training sessions in the individual listeners are indicated by the connecting dashed lines. Asterisks indicate significant (**p < .01, *** p < .001) differences between the sessions found in the post-hoc analysis.

2.4.1 The effect of training on the outcome measures of speech perception

Phoneme identification

The results of the phoneme identification test showed that the training was accompanied by significantly higher recognition scores for two out of the three tested phoneme targets: vowels and post-vowel consonants (C2). The applied testing procedure differed sufficiently from the training paradigm such that it could be ruled out that the improvement was caused solely by procedural learning. The major differences between the test and the training were: the talkers (female talker F1 in the test vs. female talker F2 and male talker M1 in the training), the number of response alternatives, the SNR as well as the task (i.e. identifying one consonant vs. identifying both consonants within one trial). The fact that the improvement was observed in a setting with untrained parameters for two out of three tested targets supports the first hypothesis that training can improve phoneme identification in noise.

The finding that the effects of training were observed only for one position of the consonant (C2 but not C1) was unexpected, since Woods et al. (2015b) in a study considering CVC units found no effect of consonant position on training benefits. The effect of position might therefore be specific to the DANOK material and linked to the presence of the second vowel (the final /i/). In DANOK, the initial consonant is involved in coarticulatory effects with the following vowel only. C2, on the other hand, is part of a VCV sequence. Thus, in the case of C2, more coarticulation cues are available for the listener. Assuming that the training improved the ability to utilize those cues, the benefit might be higher for the second consonant. The listeners might have adapted their strategy to optimize their score and therefore might have focused their attention on the target, which is easier to identify, neglecting the first consonant. This explanation is also consistent with the minor drop in the C1 score in the training group, which was not observed in the control group.

The results obtained at the follow-up session for the V and C2 targets did not differ significantly from the post-training performance, indicating that the training-related improvement was still present after three months. This result supports the third hypothesis about the retention of the training effects.

Sentence intelligibility

The analysis of HINT results did not reveal a significant interaction of Group and Visit. This finding suggests that the improvement in sentence intelligibility after the phoneme-in-noise training did not differ from the improvement attributed to the effects of test-retest and the 'placebo' intervention. Hence, the second hypothesis concerning the generalization of the training effects towards higher-level context speech could not be supported in the present study. Nonetheless, the magnitude of the change of the SRT was larger for the training (0.7 dB) than for the control group (0.3 dB) and comparable to the result previously obtained using HINT (0.8 dB in sentence recognition thresholds; Woods et al., 2015b).

The reports on the generalization of phoneme training to untrained intelligibility measures remain mixed. Ferguson et al. (2014) did not find any significant improvement of speech-in-noise measures in the group of unaided individuals with mild-to-moderate hearing loss (even though an improvement in terms of cognitive measures and self-reported outcomes was reported). Similarly, in Woods et al. (2015b), the training did not result in any improvement of sentence recognition thresholds. In contrast, the improvement of monosyllabic word recognition scores in noise after phoneme-training has been reported in normal-hearing listeners (Schumann et al., 2016). For CI patients, an improvement of sentence intelligibility scores was revealed only in a moderately difficult (+5 dB SNR) noise condition (Schumann et al., 2015). The discrepancies between these findings may stem from the differences between the target populations. For example, the mechanisms and the degree of training-related improvements can differ greatly between hearing-aid and CI users as the underlying hearing deficits are largely different.

The duration of the training is another factor the generalization to untrained measures may rely on. In the present study, the participants trained only for two weeks, which is shorter compared to other studies (e.g., three weeks in Schumann et al., 2015). Such a two-week regimen may be sufficient to observe an on-task improvement, but more training would be required to utilize the trained skill in other tasks and conditions. This explanation is in line with the Reverse Hierarchy Theory, which states that prolonged learning of lower-level representations may eventually lead to increased generalization to global contexts (Ahissar et al., 2008). The observed improvement in phoneme identification may reflect the modification of low-level representations, but not enough time was given to modify the high-level representations and hence, the lack of the effect on sentence intelligibility.

Finally, the study design is of importance when assessing the generalization to the untrained measures. For example, the studies by Schumann et al. (2016) and Schumann et al. (2015) involved only a non-contact control group. Thus, the reported improvement of speech intelligibility scores could, at least partially, be attributed to the psychological effect of the training, i.e. increased motivation, increased self-confidence, etc.

2.4.2 The challenge of achieving appropriate task demands when training phoneme identification in speech-weighted noise

The training paradigm investigated in the present study was motivated by the need to re-train the usage of speech cues in hearing-impaired listeners. Using background noise in the training had two main advantages. First, by creating a training environment in which the speech is presented simultaneously with interfering sound, one can simulate to some degree the condition that is most problematic for the listeners in their everyday lives. This should facilitate the generalization of the training effects to situations in which improving speech intelligibility is most crucial. Secondly, the presence of noise can mask some information in the phoneme spectrum and force listeners to use other accessible cues. It has been demonstrated by Varnet et al. (2019) that when discriminating between two stop consonants (/d/ and /g/) in noise, individuals with high-frequency hearing loss (provided with an audiogram-based gain) relied less heavily than normal-hearing listeners on the high-frequency cue than on low-frequency cues. Favoring low-frequency information by hearing-impaired individuals is not surprising as this information remains relatively reliable compared to high-frequency ones, even in progressing hearing loss. As proposed by Woods et al., 2015b, training can create the opposite effect where the low-frequency cues are masked by noise and thus become less reliable than the high-frequency cues (provided that those can be restored with amplification).

Introducing masking noise creates a need for defining an SNR that will provide optimal task demands. It is known that the training should take place in the condition that corresponds to the individual's threshold and that the task difficulty should adaptively follow the gradual improvement of this threshold. The main challenge in terms of defining the threshold for phoneme identification lies in the high degree of variability in the phoneme material (especially consonants) and the fact that not all the trained items were equally susceptible to masking in the presence of speech-weighted noise (Phatak and Allen, 2007; Woods et al., 2010). The variability of phonemes makes it impossible to select an SNR that would even roughly represent an identification threshold for all trained items. This was demonstrated by Stecker et al. (2006), where the SNR during the training was adjusted on a trial-to-trial basis using a 1-up, 1-down tracking procedure. In such a case, the adaptation was guided mostly by differences between the consecutive trained tokens rather than changes in the listener's identification abilities. Essentially, the only way of precisely defining the threshold in this type of training is to do it on an individual consonant-to-consonant basis. Such an approach was implemented in the study with English consonants (Woods et al., 2015b), where the SNR adaptation was based on the sensitivity index (*d'*) derived for each trained consonant. However, it was not feasible to incorporate this method in the present study as it would require more repetitions of each target phoneme than could be provided within a two-week training.

In the present study, the SNR adaptation based on the evaluation of the score after every fifth trial was proposed. To compensate, at least partially, for the variability between consonant thresholds in speech-weighted noise, the consonants were grouped into two categories and an easy-to-identify category was always presented at a fixed offset of -4 dB added to the SNR defined by the performance-based adaptation rule. A limitation of this method is that the adaptation did not precisely track any pre-defined threshold. Thus, the decrease of the mean SNR across the sessions can only approximately characterize the training-induced improvement. The average scores for individual training sessions were in the range between 57.8% and 73.9%. These values roughly correspond to the desired range of performance since an adequate training difficulty is commonly associated with a performance level between 60-80% (Serman et al., 2018). Therefore, it seems that task demands were reasonable, i.e. the training was challenging and manageable at the same time.

2.4.3 Perspectives

Not ‘if’ but ‘how?’ – investigating mechanisms behind the improvement

The current study investigated whether auditory training can improve phoneme perception in hearing-aid users but did not explore mechanisms that might underlie such improvement. It remains unknown whether the training indeed stimulated a change in terms of utilizing certain acoustic cues in speech or whether the improvement is attributed to the development of some other skills, such as the enhanced ability to separate the target from the background noise. This question could be addressed in an experiment that would directly investigate if the weighting of acoustic cues can be altered through perceptual training. It would be particularly valuable to know if the listening strategies of hearing-impaired individuals can be re-trained to become more like those of normal-hearing listeners.

Varnet et al. (2013), developed a method that allows for identifying fine acoustic cues used by a listener to categorize speech sounds, called auditory classification images (ACIs). This approach was used to reveal differences in the perceptual strategies used by musicians and non-musicians (Varnet et al., 2015) as well as listeners with and without hearing loss (Varnet et al., 2019) to discriminate between consonants in noise. ACIs should also have the potential to provide a more detailed description of how training alters the mechanisms of speech perception in hard-of-hearing individuals.

Measuring the generalization of training – the impact of a test’s sensitivity

The importance of selecting an appropriate and sensitive training outcome measure has previously been emphasized in the literature on auditory training (Henshaw and Ferguson, 2013a). Several training studies that used more than one speech outcome reported that the effects of training were revealed only for a subset of those measures (Stropahl et al., 2019). It has been shown, for example, that the improvement was present only in an adequately challenging noise condition but not when the task was either too easy or too difficult (Henshaw and Ferguson, 2013b). Apart from the difficulty that can be steered by manipulating the characteristics of the masker, the degree of predictability of the speech material can affect the sensitivity of the test. Woods et al., 2015b argued that the dominant role of top-down processing in the comprehension of predictable HINT sentences limits the benefits of training on sentence recognition thresholds.

A low-context test of speech intelligibility, such as words and digits, could be a more sensitive outcome. However, a drawback of such measures is their low relevance for real-life scenarios, where listening to speech with rich semantic context is a more common situation than identifying words in isolation. An alternative way of measuring the effects of training would be to characterize its impact on processing effort. This approach might be especially relevant in the case of an analytic phoneme-based training approach as the one considered in the present study. Improved discriminability of individual phonemes, even if not reflected in the change of intelligibility thresholds, might reduce the proportion of misperceived speech segments. Thereby, a potential benefit of phoneme training is that listeners need to resolve fewer ambiguities present in the perceived signal. This, in turn, could alleviate the cognitive load that would otherwise be required to achieve comprehension. Combining a speech intelligibility test with a physiological correlate of listening effort might be a suitable method to assess the benefits of a phoneme-in-noise training intervention and could be an interesting topic for future investigations.

2.5 Summary and conclusion

This study investigated the efficacy of auditory phoneme-in-noise training in experienced hearing-aid users. After only two weeks of training, a significant increase in phoneme identification scores was found in the training group as opposed to the active control group for vowels and post-vowel consonants but not for onset consonants. The improved performance was retained until three months after completing the training. In spite of a promising trend, no significant benefit on SRTs in sentences in noise was found, suggesting that there was no generalization of the trained skill to an untrained condition after two weeks. The lack of the effect on sentence intelligibility can potentially be linked to the relatively short duration of the training or to the insufficient sensitivity of the outcome measure. Nonetheless, based on the findings of this study, the claim that this type of training will improve speech comprehension in real life cannot be supported.

3

The influence of auditory training on listening effort in hearing-aid users - insights from a pupillometry study[a]

Abstract

Auditory training is one of the possible strategies to improve speech perception in individuals with hearing loss. So far, most studies have assessed the effectiveness of auditory training using intelligibility metrics. Little is known about how auditory training affects peoples' effort needed to recognize speech. In the present study, pupillometry was used to characterize listening effort during a hearing in noise test (HINT) before and after phoneme-in-noise identification training. Twenty experienced hearing-aid users participated in the study, half of whom completed the training, while the other half formed an active control group. Higher peak pupil dilations (PPDs) were obtained at the end of the study compared to the beginning in both groups of the participants, suggesting increased effort. The analysis of pupil dilation in an extended time window suggested, however, that the listening effort increased more in the training than in the control group. The effect of training on effort was observed in pupil responses even when no improvement in HINT was found. On the other hand, the pupil responses proved to be sensitive to placebo effect, as observed in the results of the control group.

---

[a] This chapter is based on Koprowska A., Wendt D., Serman M., Dau T. & Marozeau J. (*in preparation*).

3.1 Introduction

Auditory training has been considered to support aural rehabilitation in individuals with hearing loss (Stropahl et al., 2019). Training programs for hearing-impaired populations are designed to improve specific perceptual or cognitive skills relevant for speech perception. The desired scenario is that the skills refined via training will contribute to better speech understanding in everyday life situations, including a variety of untrained stimuli and conditions. The most common way of assessing the generalization of training is to administer an untrained speech intelligibility task before and after the intervention (see for review: Henshaw and Ferguson, 2013a; Stropahl et al., 2019). Significant improvement in the training recipients' performance (compared to a control group) is interpreted as evidence for the training's efficacy. Nevertheless, the literature suggests that evaluating the success of an intervention based only on speech recognition metrics may have certain limitations.

The result of a speech intelligibility test does not quantify how difficult it is for an individual to achieve a given performance. It has been shown that hearing-impaired listeners need to invest more resources to recognize words in noise than an age-matched group with normal hearing (McCoy et al., 2005). This finding suggests that hearing loss can turn daily communication into a highly demanding and exhausting task in the longer run. Individuals with hearing disabilities indeed report higher perceived listening effort and fatigue levels than people without auditory deficits (Alhanbali et al., 2017). Increased conversation effort appears to be a highly distressing consequence of a hearing loss, as it has been associated with a higher perceived handicap (Gatehouse and Noble, 2004). Overall, it appears that quantifying listening effort is essential to capture the peoples' listening and communication difficulties.

The literature also shows that speech recognition is not the only aspect important for characterizing the benefit of hearing loss interventions. Humes (1999) showed that metrics of subjective listening effort in noise represent a different dimension of hearing-aid outcome than metrics related to speech recognition. A study by Wendt et al. (2017), demonstrated that hearing-aid signal processing can reduce listening effort even in a condition where intelligibility was unchanged. Thus, measuring listening effort adds additional value when evaluating the success of hearing loss remediation strategies. Nevertheless, so far only a few studies have attempted to characterize the relationship between auditory training and listening effort.

Henshaw and Ferguson (2013b) administered a dual auditory and memory task designed by Howard et al. (2010) in a group of hearing-aid users who underwent phoneme-in-noise discrimination training. The primary (auditory) task was to recognize and repeat words presented in quiet or in noise (at a signal to noise ratios (SNRs) of -4 or 0 dB), whereas the secondary (memory) task was to remember digits visually presented on a screen prior to the auditory task, and to recall them after the auditory task was completed. In the dual-task paradigm, the performance in the secondary task is assumed to reflect spare cognitive resources not utilized to solve the primary task and thus, has been used to characterize listening effort (McGarrigle et al., 2014). Henshaw and Ferguson (2013b) reported an improvement in the dual-task score (primary and secondary task taken together) in one out of three noise conditions (at 0 dB SNR). Their study focused mainly on the sensitivity of the outcome measures so no final conclusions about the impact of training on listening effort was drawn from that result.

Kuchinsky et al. (2014) used pupillometry to investigate the impact of a frequent-word training program developed by Humes et al. (2009) on listening effort in older adults with hearing loss. Pupil dilation is known to reflect cognitive load during a behavioral task (Piquado et al., 2010). The training resulted in increased overall dilation, interpreted by the authors as an indicator of an increased arousal, and a faster peak of the responses, seen as a biomarker of a more rapid target discrimination of words in noise. None of these effects was observed in the control group. The study demonstrated that combining a speech intelligibility test with pupillometry can reveal additional aspects of training. Yet, the task used to obtain pupil responses was very similar to the one used in training (closed-set monosyllabic word recognition in noise). It remains unknown, how training effects on effort are reflected in pupil responses to an untrained and more complex speech material.

Koprowska et al. (in preparation) demonstrated that phoneme-in-noise training improved consonant and vowel identification skills in experienced hearing-aid users. However, no significant effect was found on sentence intelligibility measured in hearing in noise test (HINT). A similar lack of a significant effect in HINT in a consonant-training study by Woods et al. (2015b) was explained by the predominant role of top-down processes in comprehension of high-context speech material. Reconstructing the meaning of a sub-optimally perceived sentence is possible based, e.g., on contextual cues, but this process imposes additional demands on cognitive resources, as described, for example, by the Ease of Language Understanding (ELU) model (Rönnberg et al., 2013). Improving phoneme discrimination should lower the need to involve extra resources in compensating for ambiguities in the speech signal. Thus, the effort needed to understand the sentences might be lower after the training, even if the intelligibility remains similar.

The present study verified this hypothesis by measuring listening effort in experienced hearing-aid users before and after phoneme-in-noise training. Pupillometry was selected as a method to characterize listening effort. As a physiological measure, pupillometry allows for avoiding participants' bias as opposed to self-report measures. Additionally, pupil metrics have been reported to show a good-to-excellent test-retest reliability (Alhanbali et al., 2019), suggesting that they can be used as an outcome measure of a training intervention. Task-evoked pupil responses were recorded while participants performed the Danish hearing in noise test (HINT; Nielsen and Dau, 2011) in background noise. It was investigated if phoneme training had an impact on the effort required to process untrained speech stimuli that were more complex and realistic than the trained material.

Three research questions were addressed. First, it was investigated if listening effort is lower after the training than before. Two pupil metrics were considered: the Peak Pupil Dilation (PPD), defined as the maximum of the pupil response function, and the overall response magnitude. After the training, reduced PPDs, as well as reduced overall responses, were expected as indicators of lower effort. Second, the sensitivity of pupillometry to the potential effect of training on listening effort in a situation when there is no change in intelligibility was assessed. For this purpose, the pupillary responses were measured at two signal-to-noise ratios (SNRs) corresponding to performance levels of 50% (SNR50) and 80% (SNR80). It was expected that the effect of training on intelligibility would only be seen in the SNR50 condition since in the SNR80 condition, the performance would be close to the ceiling. It was also anticipated that the release from effort would be reflected in the pupil metrics regardless of the performance level. The third question was whether active participation without the training elicits significant changes in pupil responses. As opposed to the study by Kuchinsky et al. (2014), the present experiment involved not a passive but an active control group. Kuchinsky et al. (2014) used growth curve analysis (GCA) to model pupil responses and reported no significant changes in the obtained parameters for the control group. The present study applied the same method to investigate if any changes for the active control group occurred.

3.2 Methods

3.2.1 Participants and study design

Twenty elderly hearing-aid users were enrolled in the study. Apart from fulfilling the inclusion criteria listed in *Chapter 2*, none had eye diseases or took medications that could affect the pupillary response (Winn et al., 2018). The participants' demographic and clinical data, i.e., age, sex, pure-tone average (PTA) of hearing thresholds, auditory digit span (DS) scores and hearing-aid usage, are provided in Table 2.1 in *Chapter 2*. All individuals participating in the study completed two measurement visits with a two-week interval. Between these visits, one group of the participants (n = 10) completed phoneme-in-noise training while the other group (n = 10) received a control intervention based on listening to audiobooks and answering multiple-choice questions.

The summary of the groups' characteris (age, hearing loss and working memory) is presented in Table 3.1. The groups were homogeneous in terms of age and hearing loss but a significant difference between the auditory digit span scores was observed (*t = -2.98, p < .01*). The training group had, on average, lower working memory capacity than the control group.

Table 3.1: Group means, standard deviations (SD) and the results of between-group t-test (t-statistics and p-values) for age, PTA and auditory digit span (DS) score. Significant difference between the groups is indicated in bold.

Variable	**Group**	**Mean**	**SD**	***t***	***p***
Age	Training	71.70	5.16	0.91	.37
	Control	69.40	5.52		
PTA [dB HL]	Training	44.50	6.33	1.43	.17
	Control	39.38	8.65		
DS score	Training	24.88	5.99	**-2.98**	**<.01**
	Control	69.40	5.52		

3.2.2 Experimental procedure

During the pre- and post-intervention assessment, sentence intelligibility was measured using the Danish hearing in noise test (HINT; Nielsen and Dau, 2011). The five-word sentences spoken by a male talker were presented in stationary speech-shaped noise. The stimuli were played from a loudspeaker placed at a one-meter distance from the listener. The participants were instructed to repeat the sentences as precisely as possible and were encouraged to guess if they were uncertain. The responses were scored by an audiologist who was a native Danish speaker. The participants performed the task while wearing their hearing aids.

The measurement started with an estimation of the signal-to-noise ratio which for a given individual corresponded to an intelligibility score of 50% (SNR50) and 80% (SNR80), respectively. This part of the measurement did not involve eye-tracking. After a training list used to familiarize the participants with the task, two sentence lists of 20 sentences each were administered. The noise fixed at 65 dB SPL started one second before and ended one second after the sentence. The speech level was adjusted adaptively from trial to trial to achieve the desired performance. A word-scoring procedure was applied. To estimate SNR50, the magnitude of the speech level adjustment varied between +2 dB and -2 dB, depending on the number of correctly repeated words in the sentence. The speech level was increased by 2 dB if none of the five words were correctly repeated and decreased by 2 dB when all the five words was correctly repeated. For one, two, three and four correctly recognized words, the step size was +1.2, +0.4, -0.4 and -1.2 dB, respectively. To estimate SNR80, the speech level could change between 3.2 dB and -0.8 dB based on analogical principle. For the first four sentences, the magnitude of adjustment was multiplied by two. The order of SNR50 and SNR80 estimation was balanced across the participants.

In the following part of the assessment, the participants' pupil size was registered while they performed the sentence intelligibility task at a constant SNRs. The pupil size was recorded at a sampling rate of 500 Hz with an EyeLink camera (SR Research, Canada). The data was collected in a room without windows and the luminance conditions were the same for all measurement sessions. The camera was fixed to the participant chair with an adjustable arm and located in front of the person. The participants were asked to fixate a visual target located on the wall behind the loudspeaker. The adjustable headrest was used to ensure the participants' comfort and stabilize the head position. After one training list to familiarize participants with the eye-tracker, two test conditions were administered: one corresponding to 50% intelligibility and one corresponding to 80% intelligibility. The noise level at a one-meter distance was fixed at 65 dB SPL and the speech level was set to achieve the desired SNR (SNR50 or SNR80). The order of the two conditions was balanced across the participants.

Each trial (corresponding to one sentence) followed a specific sequence of events to obtain a reliable task-evoked pupil response (Winn et al., 2018). The trial began with two seconds of silence and three seconds of noise. Keeping the initial noise interval constant across the trials was intended to prepare the participant for the stimulus and avoid a situation where the sentence occurred unexpectedly, possibly causing a task-unrelated dilation. Even though the onset of the noise could elicit a brief pupil response, the pupil size should have stabilized before the beginning of the sentence. After the pre-stimulus phase, the target sentence was presented while the noise continued playing. The average duration of the sentence was 1.5 seconds. A three-second retention interval followed the sentence to separate the dilation evoked by listening to the stimulus from the dilation related to verbal response. The noise continued during this phase. The listeners were instructed not to repeat the sentence until the noise ended. After the offset of the noise, the participant provided a verbal response. The experimenter manually triggered the beginning of the next trial a few seconds after the response to allow the pupil to return to its baseline size.

At the post-intervention visit, the test lists were administered at SNR50 and SNR80 defined before the intervention such that for each individual the acoustic conditions (and not the performance level) were held constant across the visits. Two adaptive lists were nevertheless run also during the post-intervention visits in order to keep the duration of both measurement sessions approximately the same. A substantial disproportion of the measurement time across the visits would have most likely resulted in a different within-session fatigue, which would have made the comparison of the obtained pupillometry results less reliable.

It has been shown that the pupil diameter is smaller in the afternoon compared to the morning (Eggert et al., 2012). To avoid within-subject variation in baseline pupil size, the measurements were planned such that for each individual the pre-intervention and post-intervention visit took place at the same time of the day. Only in three instances, due to logistic challenges, the second visit had to be rescheduled by 4-5 hours.

3.2.3 Pupillometry data processing

Raw pupil data were pre-processed in MATLAB using the PUPILS pipeline (Relaño-Iborra and Bækgaard, 2020). Blinks were detected by identifying values differing more than two standard deviations from the mean of the acquisition corresponding to one sentence list. Datapoints within 50 milliseconds before and 150 milliseconds after the blinks (Winn et al., 2018) were removed and linearly interpolated. Next, the data was denoised by applying low-pass filtering with a cut-off frequency of 10 Hz. Trials with more than 25% of the data interpolated within the relevant time interval (between the onset and the offset of the sound) were removed from further analysis.

The data within each trial were baseline-corrected by subtracting an average pupil size within one second before the sentence onset. The baseline value usually takes a couple of first trials to stabilize, so the responses to first three sentences in each list were not considered in the analysis. Furthermore, all traces were visually inspected for baseline artifacts, which would affect the subsequent pupil estimation or gross distortions that were not detected by the automatic procedure. Trials in which these artifacts were found were also discarded from further analysis. The baseline-corrected data were normalized relative to the pupil range to obtain a better test-retest reliability (Neagu et al., in preparation). The lower and upper limits of the individual's pupil range were defined as the 0.025 and 0.975 quantiles of the datapoints within five seconds after the sentence onset (considering only non-discarded trials). The normalization involved a subtraction of the lower limit and a subsequent division by the range. Four mean response curves were calculated for each participant by averaging the non-rejected trials for each of the two experimental conditions within each visit. As a starting point, a minimum of ten preserved trials was set as a requirement to include the resulting mean response in the analysis. In four instances, fewer than ten trials remained after interpolation, rejecting the three first trials, and manual cleaning. For two of the test subjects, one of the four response curves did not fulfill the requirement. After visual inspection, these curves were kept for further analysis to preserve the complete pairs of pre- and post-intervention responses. Only one test subject in the control group, for whom more than one response would be an average of nine or fewer than nine trials, was excluded from the analysis.

3.2.4 Analysis

The sentence intelligibility results, expressed as the percentage of correctly repeated words within the test list, were analyzed in R using the *lmerTest* package (Kuznetsova et al., 2017). A mixed-effect linear model with fixed effects of *Visit* (levels: *pre*, *post*) and *Group* (levels: *training*, *control*) as well as a *Visit x Group* interaction and a random effect of *Participant* was fitted for each condition. An interaction of fixed effects represented the impact of the intervention on the outcome measure. The significance of the interaction would indicate that the impact of the intervention on speech intelligibility differed across the groups. In such case, a post-hoc analysis including planned comparisons between the levels of *Visit* within the factor of *Group* was performed. The Tukey adjustment method was applied to account for multiple comparisons (Lenth, 2016).

Table 3.2: Growth Curve Analysis model tests for pupil responses extracted in the SNR50 condition. The base model formula is provided with the random structure shown in parenthesis. After adding each new model parameter, the improvement of fit was evaluated using an Akaike Information Criterion (AIC) and a chi-square test (χ^2). Significant changes in the goodness of fit are indicated in bold.

Condition: SNR50			
Base model formula:	*pupil dilation = Linear + Quadratic + Cubic + (Linear + Quadratic + Cubic|Participant)*		
	AIC	χ^2	***p***
Base model	-252626		
Group			
Intercept	-252624	0.41	.51
Linear	-252623	1.08	.30
Quadratic	-252622	0.47	.49
Cubic	-252620	0.15	.70
Visit			
Intercept	**-262305**	**9687.45**	**<.0001**
Linear	**-266372**	**4069.32**	**<.0001**
Quadratic	**-267097**	**726.64**	**<.0001**
Cubic	**-267262**	**166.66**	**<.0001**
Group x Visit			
Intercept	**-269408**	**2148.53**	**<.0001**
Linear	**-269520**	**113.78**	**<.0001**
Quadratic	-269518	0.002	.97
Cubic	**-269544**	**28.39**	**<.0001**

The peak-pupil dilation (PPD) was used as a metric reflecting the cognitive load during the behavioral task. PPDs were extracted as the maximum value of each mean response curve between 0.5 to 3 seconds after the stimulus onset (i.e., up to approximately 1.5 seconds after the end of the sentence). Based on the previous literature, the maximum of the response should be contained within this time window (Winn et al., 2018). The statistical analysis of the obtained PPDs employed tools and approaches analogical to those applied to the intelligibility scores (described in the previous paragraph).

Table 3.3: Growth Curve Analysis model tests for pupil responses extracted in the SNR80 condition. The base model formula is provided with the random structure shown in parenthesis. After adding each new model parameter, the improvement of fit was evaluated using an Akaike Information Criterion (AIC) and a chi-square test (χ^2). Significant changes in the goodness of fit are indicated in bold.

Condition: SNR80			
Base model formula:	*pupil dilation = Linear + Quadratic + Cubic + (Linear + Quadratic + Cubic|Participant)*		
	AIC	χ^2	***p***
Base model	-244416		
Group			
Intercept	-244414	0.50	.48
Linear	-244413	0.46	.50
Quadratic	**-244415**	**3.92**	**<.05***
Cubic	-244413	0.05	.83
Visit			
Intercept	**-246581**	**2170.20**	**<.0001**
Linear	**-246587**	**8.33**	**<.01**
Quadratic	**-247115**	**530.03**	**<.0001**
Cubic	**-247130**	**16.47**	**<.0001**
Group x Visit			
Intercept	**-247784**	**656.44**	**<.0001**
Linear	**-248152**	**370.11**	**<.0001**
Quadratic	**-248330**	**179.08**	**<.0001**
Cubic	**-248349**	**21.41**	**<.001**

To capture the effect of the training on the overall shape of the response, a growth curve analysis (GCA) was applied. GCA is a multilevel regression technique suitable for analyzing time-course data (Mirman, 2014) commonly used in pupillometry studies (e.g., Kuchinsky et al., 2013; Wendt et al., 2018; Winn, 2016). GCA allows for modeling the shape of the response using polynomial functions. In the output of a GCA model, the intercept represents the overall area under the curve, while the linear term reflects the overall slope of the response. The quadratic term is interpreted as the symmetric rise and fall rate around the peak. The higher-order terms become more difficult to interpret but are generally associated with the presence of second and subsequent inflection points (Mirman, 2014). In the pupillometry-related literature, the cubic term has also been interpreted as the latency of the primary peak (Kuchinsky et al., 2013). In the present study, the GCA was used to model the shape of the response curve within 0 to 5 seconds after the onset of the sentence. This time window contains the stimulus presentation (0 to approximately 1.5 seconds) and the retention interval before the verbal response (3 seconds after the sentence offset). As such, it should reflect cognitive load involved in listening and in rehearsing and reconstructing the sentence in memory before repeating it back to the experimenter.

The GCA was performed in R using the *lme4* package (Bates et al., 2015). Two separate models were fitted for each experimental condition. Within the time window selected for the analysis, the response curve typically does not show more than two inflection points. Therefore, the selected model structure included up to the third-order time term (i.e., the intercept, linear, quadratic and cubic term). The random effect structure of the model allowed all these terms to vary with the listener. Orthogonal polynomials were used such that each term could be interpreted independently. The impact of experimental factors (*Group, Visit* and *Group x Visit* interaction) on the curve characteristics was investigated by gradually adding interactions between these factors and the polynomial terms. After increasing the model complexity by one parameter (i.e., adding each subsequent interaction), the fit was evaluated using an Akaike Information Criterion (AIC) and a χ^2 test. The model tests are summarized in Table 3.2 for SNR50 and in Table 3.3 for SNR80. The reduction of the AIC value indicated improved goodness of fit. The p-value obtained from the χ^2 test showed whether adding a new parameter significantly improved the model fit compared to the previous most-complex model. The models were gradually built up until the most complex structure of interest was reached (i.e., model that included the *Visit x Group* interaction on all the time terms). The models that provided the best description of the data (i.e., involved only those effects, which significantly contributed to the model fit) were used for the subsequent analysis.

After defining the best-fitting models, the estimates that characterized the change from pre- to post-visit in both groups of the participants were extracted together with their significance levels. The parameters for the training group were obtained by fitting the respective models with the "training-pre" as a reference. To obtain the parameters for the control group, the reference in the models was changed to "control-pre".

3.3 Results

3.3.1 Speech intelligibility

Figure 3.1 represents the pre-intervention (empty bars) and post-intervention (filled bars) speech intelligibility scores at SNR50 (left panel) and SNR80 (right panel). The results of the training group are shown on the left and the results of the control group are displayed on the right side of each panel. The pre-intervention scores roughly corresponded to the targeted intelligibility levels (50% and 80% correct). At SNR50, there was a marginally significant *Visit x Group* interaction (*F(1,18) = 4.24, p = .05*). A post-hoc analysis revealed that the percentage of correct responses was significantly higher after the intervention only for the training group (*t = -2.87, p = .01*). Neither the interaction nor main effects were significant in the second condition.

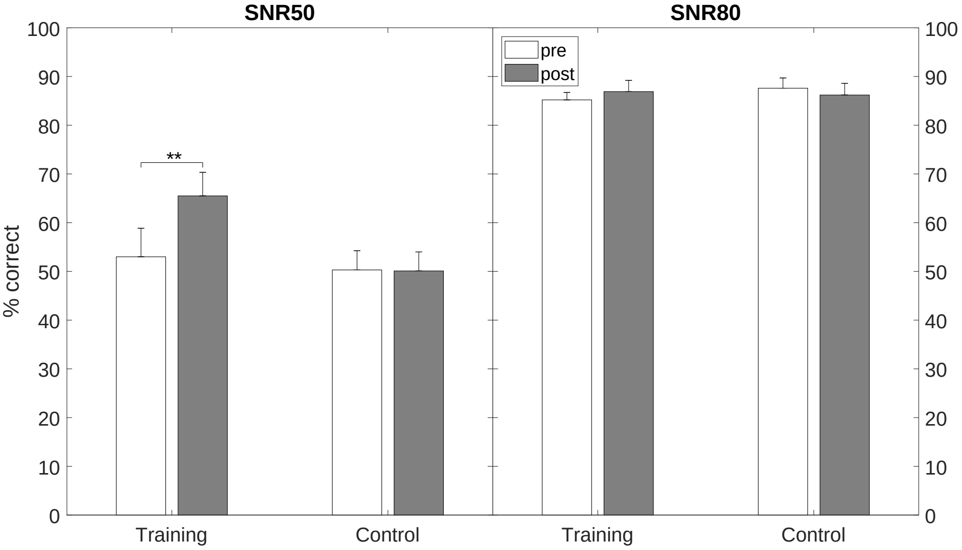

Figure 3.1: Pre-intervention (empty bars) and post-intervention (filled bars) word scores in percent correct for SNR50 (left panel) and SNR80 (right panel). The mean results of the training group are shown on the left of the panels and the mean results of the control group are on the right of the panels. Errorbars represent one standard error of the mean. Asterisks indicate significance level ** p<.01.

3.3.2 Pupillometry

Figure 3.2 depicts the mean pupil responses for the training (green curves) and the control group (blue curves) obtained at SNR50 (left panel) and at SNR80 (right panel). The responses measured before the intervention are indicated by brighter colors, whereas the responses obtained after the intervention are shown as darker colors. The time interval from -3 to 0 seconds relative to the sentence onset corresponds to the pre-stimulus noise playback. The interval from -1 to 0 seconds was used for baseline estimation (indicated with the vertical dashed lines). The pupil started dilating just after the beginning of the sentence and reached its maximum between 2 and 3 seconds after the sentence onset, which corresponds to approximately 0.5-1.5 seconds after the end of the sentence. The shape of the response after the peak differed between the conditions. For SNR80, the pupil showed a peak-release pattern with faster decrease towards the baseline. For SNR50, the responses showed a more sustained pattern without clearly pronounced decay.

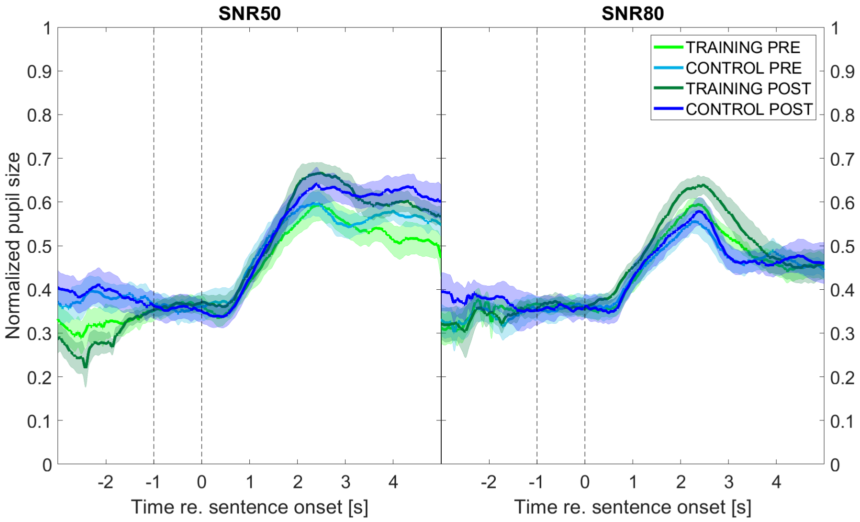

Figure 3.2: Normalized pupil responses averaged across participants plotted as a function of time. The results for SNR50 are plotted on the left panel and the results for SNR80 are shown on the right panel. Green lines represent the training group and blue lines correspond to the control group. Responses obtained before the intervention are plotted with brighter colors and responses collected after the intervention are indicated with darker colors. The shaded areas represent +/-1 standard error of the mean. The vertical dashed lines mark the baseline estimation window.

Figure 3.3 shows PPDs obtained before (empty bars) and after (filled bars) the intervention in the two experimental conditions: SNR50 (left panel) and

SNR80 (right panel). The bars representing the training and the control group respectively, are shown on the panels' left and right sides. Both groups had similar PPDs before the intervention and showed consistently higher values after the training in both experimental conditions. The analysis revealed no significant *Visit x Group* interaction nor a significant *Group* effect in any of the conditions. The effect of *Visit* was significant both for SNR50 ($F(1,17) = 5.85$, $p < .05$) and SNR80 ($F(1,17) = 5.91$, $p < .05$), indicating that the increase in PPD was similar for both groups.

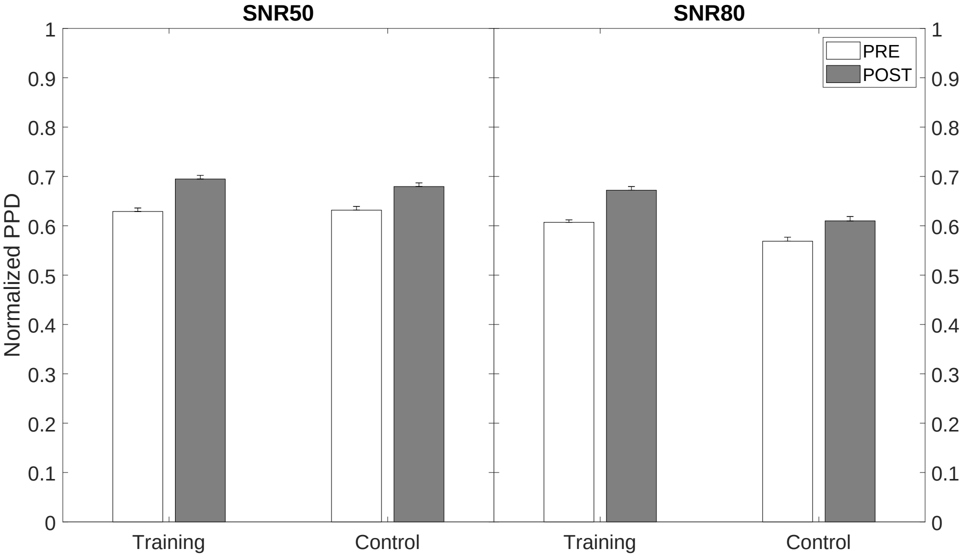

Figure 3.3: Normalized peak pupil dilations (PPDs) before (empty bars) and after (filled bars) the intervention shown for SNR50 (left panel) and SNR80 (right panel). The mean results of the training group are shown on the left and the mean results of the control group are on the right of the panels. Errorbars represent one standard error of the mean.

Figure 3.4 shows the output of the GCA models for SNR50 (left) and SNR80 (right) for the time window 0-5 seconds from the sentence onset. The upper panels represent the outcomes for the training group and the lower panels show the results for the control group. The modeled responses are indicated by solid lines: pre-intervention condition is marked by brighter colors and the post-intervention condition is displayed in darker colors. The results measured before (dotted lines) and after (dashed lines) the intervention are shown for comparison. The models capture the overall shape of the responses and the differences between the "pre" and "post" traces within the groups. However, the peaks of the curves tend to be underestimated, especially for the SNR80.

Significant *Group x Visit* interactions in GCA model for the SNR50 (Table

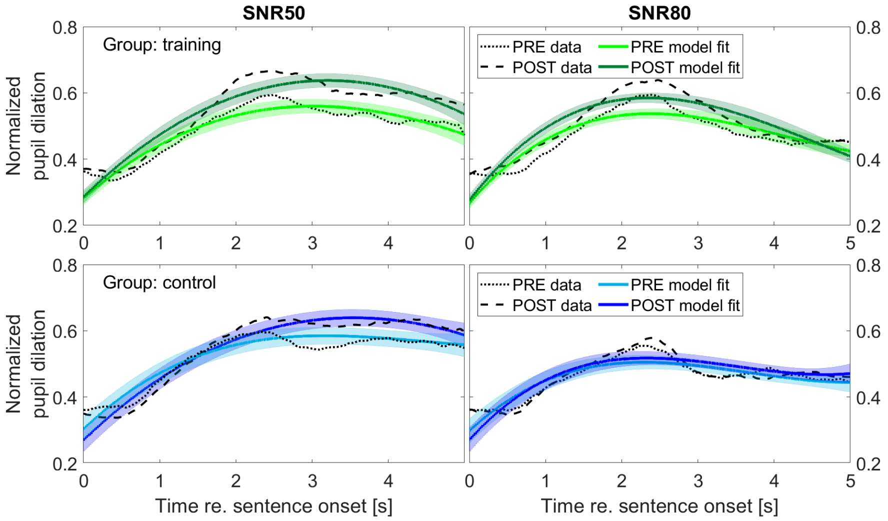

Figure 3.4: The output of the GCA models for SNR50 (left column) and SNR80 (right column). The upper panels show the results for the training group and the bottom panels represent the control group. The modeled responses for the "pre" visit are plotted in brighter colors and the modeled responses for the "post" visit are shown in darker colors. The shaded areas represent the models' variability. Data obtained at pre- and post-intervention visits are shown for comparison in dotted and dashed lines, respectively.

3.2) indicated that the interventions administered in the training and control group differently affected the overall area under the curve (represented by the intercept), the overall steepness of the curve (represented by the linear term) as well as the secondary inflection (represented by the cubic term). The *Group x Visit* interaction was not significant for the quadratic term, which however was affected by the *Visit* factor. This means that the symmetric rise and fall rate around the peak was similarly affected in both groups. The GCA model output for the SNR80 condition (Table 3.3), revealed a significant *Group x Visit* interaction for all terms indicating that the two interventions affected the curve parameters considered in the model differently. There was a significant *Group* effect on the quadratic term, suggesting a training-unrelated difference in the steepness around the peak between the two groups. The control group had a flatter response, as indicated by a more positive parameter estimate.

The parameters characterizing the change in pupil response across visits for each of the groups are summarized in Table 3.4 for the SNR50 and in Table 3.5 for the SNR80. In the SNR50 condition (Table 3.4), a significant increase of the overall area under the curve (intercept) and the overall steepness of the curve (linear term) was observed in both groups ($p < .0001$). The steepness around the peak increased as well (quadratic, *p < .0001*; the parameter estimate is identical for both groups because the final model structure did not involve a non-significant *Group x Visit x Quadratic* interaction). The estimate of the cubic term also changed significantly for both groups (*p < .0001*). Overall, the trends were consistent across the groups, but the magnitude of the change was larger for the intercept and smaller for the linear and the cubic term in the training group.

Table 3.4: The output of the GCA model for SNR50. The absolute parameter estimates are reported for the reference level ("pre"). Changes in the parameter estimates relative to the reference are shown for the "post" level. For the reference level ("pre"), p<.05 indicates that the parameter is significantly different from 0. For the "post" level, p<.05 indicates that the parameter is significantly different from the reference. Significant changes from "pre" to "post" are shown in bold.

Condition: SNR50						
	Group: Training			**Group: Control**		
	Estimate	*t*	*p*	Estimate	*t*	*p*
Intercept (pre)	0.49	28.79	<.0001	0.53	29.57	<.0001
Linear (pre)	0.78	4.37	<.0001	0.93	4.95	<.0001
Quadratic (pre)	-0.76	-7.85	<.0001	-0.76	-7.85	<.0001
Cubic (pre)	0.07	0.77	.44	0.15	1.57	.12
Intercept:post	**0.06**	**108.23**	**<.0001**	**0.02**	**38.37**	**<.0001**
Linear:post	**0.33**	**40.22**	**<.0001**	**0.46**	**52.87**	**<.0001**
Quadratic:post	**-0.16**	**-27.36**	**<.0001**	**-0.16**	**-27.36**	**<.0001**
Cubic:post	**-0.05**	**-5.82**	**<.0001**	**-0.11**	**-12.86**	**<.0001**

In the SNR80 condition (Table 3.5), the intervention was followed by an increased area under the curve (intercept) and increased steepness around the peak (quadratic) in both groups (*p < .0001*). The change was larger for the training group and rather subtle for the control group. The overall steepness of the curve (linear) also changed significantly for both groups (*p < .0001*) but in the opposite directions. The steepness decreased in the training group and increased in the control group. The significant change of the cubic term was observed only in the control group (*p < .0001*) but not in the training group (*p < .82*).

Summarizing the results, significant changes in pupil response features were measured in both groups, not only in the group that completed the training. The trends that were common across the groups and conditions were the increase

Table 3.5: The output of the GCA model for SNR80. The absolute parameter estimates are reported for the reference level (“pre”). Changes in the parameter estimates relative to the reference are shown for the “post” level. For the reference level (“pre”), p<.05 indicates that the parameter is significantly different from 0. For the “post” level, p<.05 indicates that the parameter is significantly different from the reference. Significant changes from “pre” to “post” are shown in bold.

Condition: SNR80						
	Group: Training			**Group: Control**		
	Estimate	*t*	*p*	Estimate	*t*	*p*
Intercept (pre)	0.47	43.58	<.0001	0.46	40.17	<.0001
Linear (pre)	0.36	1.77	.08	0.32	1.47	<.14
Quadratic (pre)	-0.90	-7.57	<.0001	-0.63	-5.01	<.0001
Cubic (pre)	0.23	2.49	<.05	0.23	2.34	<.05
Intercept:post	**0.03**	**52.03**	**<.0001**	**0.009**	**13.88**	**<.0001**
Linear:post	**-0.10**	**-11.16**	**<.0001**	**0.16**	**15.99**	**<.0001**
Quadratic:post	**-0.24**	**-26.05**	**<.0001**	**-0.06**	**-6.26**	**<.0001**
Cubic:post	-0.002	-0.22	.83	**-0.06**	**6.17**	**<.0001**

in magnitude and steepness of the peaks, which was generally larger in the training group. This was not true for the change in an overall steepness of the response, which was in the training group was less positive than for the control. This direction of change might be linked to the fact that in the training group the decay rate after the peak seemed to increase from "pre" to "post" visit.

Surprisingly, a change in cubic term in the SNR80 condition was present in the control but not in the training group. The third-order polynomial’s coefficient represents the sharpness of two peaks if there is a second inflection point (Mirman, 2014) but in the pupillometry-related literature has also been interpreted as the latency of the primary peak (Kuchinsky et al., 2013). As it can be seen in Figure 3.4, this aspect might not be captured by the models with great fidelity. The amplitudes of the modeled curves do not always align with the maxima in data and the magnitudes of the peaks tend to be underestimated, especially in the SNR80 condition.

3.4 Discussion

The present study used pupillometry to characterize the impact of phoneme-in-noise auditory training on listening effort exerted in HINT. Task-evoked pupil responses to HINT sentences were obtained in noise conditions corresponding to 50% intelligibility (SNR50) and 80% intelligibility (SNR80), respectively. Three research questions were addressed: (i) whether the training would results in reduced effort, (ii) whether pupillometry would reveal the effect of training on effort in a condition when intelligibility remains unchanged and (iii) whether active participation in the study can elicit significant changes in pupillary responses.

The training outcomes were assessed with reference to an active control group, which received an intervention meant to maintain a similar level of engagement in the study. The training and the control groups were well-matched in terms of age and hearing loss. However, a significant difference between the two groups in terms of auditory working memory scores was found. Despite this difference, no significant main effect of *Group* was found in any of the analyses, except for the SNR80 condition, where an effect of *Group* on the quadratic term was found.

3.4.1 The effect of training on listening effort

The hypothesized outcome of the training was that less reliance on the cognitive resources will be needed for speech intelligibility, leading to decreased listening effort. Lower effort is typically associated with smaller PPDs. The results of the study showed, however, that the PPDs were higher at the post-intervention visit, suggesting the opposite. Even more surprisingly, the magnitude of the increase of PPDs was the same for both groups.

The analysis that considered an extended time window revealed, however, that the changes in the response patterns from pre- to post-intervention in fact differed across the groups in terms of the response's overall magnitude and steepness as well as the growth/decay rate and the peak timing. The increase in the overall magnitude of the response was more pronounced in the training group, suggesting that the effort increased more for the group that received the training than for the controls.

According to the Framework for Understanding Effortful Listening (FUEL; Pichora-Fuller et al., 2016), the effort allocated to a task is influenced, i.a., by input-related demands, the overall capacity of available mental resources and the allocation policy modulated by the current state of an individual's motivation or fatigue (Kahneman, 1973). An increased effort could therefore result from a change in the allocation policy, due to the fact that the participants' motivation or fatigue status changed during the intervention. It is plausible that motivation increased in both groups, since both received an intervention that they expected to help them understand speech better. This could increase their self-confidence and willingness to allocate more resources in the task, as they trusted that the investment of effort will results in a higher success rate. Such an explanation would account for the trends seen in the PPDs, where the effect did not differ across the groups. However, the change in the overall response's magnitude indicated that for the training group the effort increased more than for the control group, which suggests that another training-induced mechanism might have been involved.

Another factor affecting the effort allocation policy is the current fatigue status. According to FUEL, fatigue moderates the subjective evaluation of both the benefits of successful performance and the cost of achieving this performance. Thus, it can hinder the willingness to invest resources in a cognitively demanding task when the estimated cost of successful performance outweighs the potential gain. This scenario is quite likely to occur in the case of hearing-impaired individuals, who generally report high fatigue levels (Alhanbali et al., 2017). The relationship between subjective fatigue and listening effort reflected by PPDs was demonstrated by Wang et al. (2018): listeners with higher self-reported fatigue levels showed smaller PPDs in speech-in-noise task.

A desired outcome of any hearing-loss intervention, be it auditory training or hearing aids, is the alleviation of the challenges experienced during everyday communication. If the intervention succeeds, the person should be less tired listening to speech in everyday life, and the overall fatigue should decrease in the long run. When the efficacy of the intervention is assessed at two points separated in time, the participants might be willing to exert more effort upon the follow-up since their overall fatigue will be lower. The higher elevation of effort observed in the training group could be attributed to relief from fatigue among the training recipients. This explanation has, of course, a speculative character since the fatigue status of the listeners was not monitored in the present study.

Finally, the training might have increased the overall capacity of cognitive resources, resulting in participants being able to allocate more effort to complete the speech-in-noise task. Even though the phoneme training targeted relatively low-level skill (phoneme identification) it could have benefited participants' overall cognitive ability. In fact, such a result has already been reported by Ferguson et al. (2014) who found a performance improvement in complex cognitive tasks (divided attention and working memory updating) following a four-week phoneme discrimination training in individuals with mild-to-moderate hearing loss.

3.4.2 Sensitivity of the outcome measures: HINT and pupillometry

As expected, the number of correctly repeated words increased significantly after the training in the SNR50 condition but not the SNR80 condition. The fact that the training-related improvement can be seen only in appropriately sensitive test conditions (which are challenging enough but neither too easy nor too difficult) has been previously demonstrated by, e.g., Henshaw and Ferguson (2013b). Unfortunately, the present study is an example of a situation when a sensitive outcome measure is not necessarily realistic. The situation when a person understands 80% of the information corresponds roughly to the bottom range of the Lowest Acceptable Performance Level (Boothroyd and Schauer, 2015), i.e. the intelligibility level at which people are willing to maintain a conversation for a short time. It is unlikely that people would engage in a conversation if they understand only 50% of what is being said. Hence, out of the two considered experimental conditions, the SNR80 resembles a real-life challenging communication scenario more closely than the SNR50. However, traditional sentence tests are not sensitive enough to detect training benefits at higher-performance levels.

Considering the outcome of the current analysis, the lack of a significant effect in HINT reported by Koprowska et al. (in preparation) seems to be more likely due to the sensitivity of the outcome measure rather than due to lack of generalization. It appears that the choice of the test procedure (e.g., the adaptive threshold estimation vs. measuring at fixed individualized SNR) determines the absence or presence of a significant effect even when using the same sentence material. This finding implies that one should be cautious not to claim an intervention inefficient solely based on the lack of significant improvement in the speech perception metric.

The results of the present study demonstrated that pupillometry can reveal the effect of the training on listening effort in a condition, when no benefit on intelligibility is observed. This finding demonstrates the importance of using outcome measures other than speech intelligibility tests to evaluate the effectiveness of auditory training.

3.4.3 The effect of a non-training activity on pupillary responses

An important element of the present investigation was the inclusion of an active control group. Previous study that used pupillometry as an outcome measure of auditory training (Kuchinsky et al., 2014) involved a passive control group, which completed two measurement sessions interleaved by a period of time equal to the duration of investigated training. No changes in pupil traces for that group were reported. Here, in contrast, the control group participated in a non-training activity, which should have resulted in the expectation that the training might have improved their auditory skills. As such, an opportunity arose to investigate if these expectations would be reflected in pupillometry results.

The control group showed a significant change in PPDs and the shape of the pupil response, which might have been associated with the participants' expectation regarding the effectiveness of the received intervention. It has been previously demonstrated that the expectations have an impact on the outcome measures, e.g., labelling a hearing aid as "new" or as "conventional" affected users' preference, sound quality ratings and even speech recognition benefit (Dawes et al., 2013). Similar mechanism may also underlie the effect observed in the present study.

In a research setting, when the results are analyzed on a group level, it is possible to control for this effect using a double-blind design. There is, however, an ongoing debate about a potential use of pupillometry as an individualized measure of effort in a clinical setting. The demonstrated sensitivity of pupillary responses to "placebo" effects should be carefully taken into consideration in the context of future applications of pupillometry to assess the efficacy of hearing-loss intervention at the level of an individual listener.

3.5 Summary and conclusion

The present study investigated the impact of phoneme-in-noise identification training on listening effort in hearing-aid users. Task-evoked pupil responses were registered while the participants performed a sentence intelligibility test before and after the training. The sentence test was administered at two performance levels, corresponding to 50% (SNR50) and 80% (SNR80) intelligibility (defined before the intervention). The outcomes were assessed with reference

to an active control group. PPDs were significantly higher at the post- compared to the pre-intervention visit in both groups. The analysis of time-dependent effects in pupil dilation revealed that the responses' magnitude and shape were affected differently across the groups, with a larger increase in the responses in the group that received the training. A significant improvement in intelligibility was found only in the training group and in the more challenging condition (SNR50). The pupil responses showed the impact of the training on effort in both HINT conditions, also when intelligibility remained unchanged (SNR80). The findings indicate that training led to increased listening effort, which may be linked to changes in the participants' motivation, fatigue status as well as overall cognitive capacity. The study demonstrated the sensitivity of pupillometry to the effect of training at high intelligibility levels. The significant changes in PPDs and the shape of the pupil responses observed in the active control group indicate that this method is also sensitive to placebo effects, which can be a challenge for the potential use in clinical settings.

4

The development of a home-based rhythmic training method to improve speech intelligibility in hearing-impaired listeners[a]

Abstract

Recent research showed some evidence that musical training can improve auditory perception in hearing-impaired listeners, potentially leading to better speech intelligibility in background noise. This study explored the feasibility of a home-delivered music-based training program for elderly listeners and its impact on the users' speech intelligibility. Based on previously documented similarities between rhythm perception in music and speech, it is here hypothesized that training rhythm-related skills could benefit speech perception by enhancing the accuracy of temporal predictions. An eighteen-day-long training program that comprised approximately ten minutes of daily exercises in synchronized tapping to music was implemented as a mobile application. Thirteen listeners with sensorineural hearing loss participated in the study. Seven listeners performed the training, while the remaining six completed another activity as a control group. Before and after the respective interventions, the participants' synchronized tapping and beat perception skills were assessed using the Beat Alignment Test (BAT, Iversen and Patel, 2008). Additionally, their speech recognition thresholds in noise were measured in Danish hearing in noise test (HINT; Nielsen and Dau, 2011) and Danish matrix sentence test (DANTALE II; Wagener et al., 2003). No influence of the training on the BAT scores nor speech recognition thresholds was found. Possible reasons for

---

[a] This chapter is based on Koprowska A., Serman M., Dau T. & Marozeau J. (*in preparation*)

the lack of the training effect as well as potential improvements in the training method are discussed.

4.1 Introduction

Recent evidence points toward the beneficial impact of musical training on understanding speech in noise in elderly adults with normal hearing and mild hearing loss (Zendel et al., 2019; Dubinsky et al., 2019). The term 'musical training' is rather broad and not limited to the protocols that have already been investigated in the context of auditory rehabilitation for older adults (i.e., piano practice and choir singing). In fact, different kinds of musical training put emphasis on different skills, which are needed to achieve satisfactory performance. As an example, playing a string instrument requires sensitivity to small pitch deviations, which is not so crucial when it comes to other instruments (e.g., piano), while playing the percussion places higher demands on timing abilities than any other type of instrumental practice. These various skills that musical training can enhance may contribute to speech-in-noise perception differently. Hence, the duration and intensity of training needed to obtain a significant benefit might also vary across alternative approaches. Therefore, identifying a specific aspect of musical abilities that is most closely linked to understanding speech in noise could facilitate the design and improve the efficiency of a music-based training program.

The question of which aspect of musical experience is most relevant for speech-in-noise perception was addressed by Yates et al. (2019). In their study, the relationship between different musical skills and speech recognition thresholds (SRTs) in noise was explored. The participants' sensitivity to melodic and rhythmic patterns was measured using the Musical Ear Test (MET; Wallentin et al., 2010) and the perception of musical beat was assessed with the Beat Alignment Test (BAT; Iversen and Patel, 2008). As opposed to the scores obtained in the "melody subtest" of MET, the results of the "rhythm subtest" of MET and the outcomes of BAT were correlated with the participants' SRTs. Out of these two metrics, BAT scores turned out to be the best predictor of individual performance. The findings of Yates et al. (2019) suggest that the expertise in rhythm-related tasks may play a significant role in speech-in-noise understanding. A similar picture emerges from a study by Slater and Kraus (2015), who investigated speech-in-noise perception in two subgroups of musicians. In their study, percussionists, who are extensively trained in rhythm perception and production, outperformed professional singers in terms of sentence recognition performance in four-talker babble noise.

Rhythm is a fundamental component of a musical piece, which refers to the placement of sounds in time. One of the most prominent aspects of the rhythmical organization of music is the presence of an isochronous beat, which is a periodically occurring pulsation that the majority of the listeners can perceive and to which they can entrain their movements. Another important feature of rhythm in music is its hierarchical nature – the individual notes are grouped into strong (accented) and weak (unaccented) beats, which form a recurring higher-order structure referred to as meter. Even though speech lacks a strictly isochronous pulsation, listeners can still extract and entrain their movements to quasi-periodic temporal regularities in the spoken language (Lidji et al., 2011). Moreover, the patterns of stressed and unstressed syllables are governed by a higher-order structure, referred to as a metrical grid (Liberman and Prince, 1977). Hence, temporal regularities and hierarchical organization emerge as shared characteristics in both domains, which appear to have similar consequences for human perception.

It has been demonstrated that temporal patterns in speech have implications for comprehension. Ghitza and Greenberg (2009) measured word recognition in time-compressed sentences while manipulating the syllable rate by insertion of periodic or aperiodic silent intervals. The amount of errors turned out to be the lowest when the insertion of periodic silence resulted in a syllable rate close to the one characteristic for conversational speech. McAuley et al. (2020) investigated the role of speech rhythm in selective listening to target in background noise. Disrupting the natural speech rhythm degraded the target word recognition both in stationary speech-shaped noise and in multi-talker babble. On the other hand, applying the same rhythmic distortion to the background talker when the natural rhythm of the target was preserved led to enhanced speech recognition. The findings of both studies suggest that listeners can utilize rhythmic properties in speech to improve their performance in difficult listening conditions.

The role of speech rhythm in speech comprehension has been linked to the Dynamic Attending Theory (DAT; Jones and Boltz, 1989; Jones et al., 2002; Large and Jones, 1999). According to DAT, the allocation of attention in time is driven by the rhythmic pattern of an external stimulus. Periodic or quasi-periodic characteristics of the signal allow for forming temporal predictions and orienting the anticipatory attention towards those points in time when important information is likely to occur. Thus, the performance in an auditory task is enhanced when stimulus occurrence matches the temporal expectations. The effect of temporal expectations in speech processing has been demonstrated in studies, where the precedence of rhythmic primes improved, e.g., reaction times in a phoneme detection task (Cason and Schön, 2012) and word recognition in babble background noise (Sidiras et al., 2017). Dynamic attending was also proposed as a mechanism underlying the effects of speech rhythm disruption on selective attending to the target in McAuley et al. (2020). The authors argued that the listener's attention is more likely to entrain to the speech signal with natural than distorted rhythm. The ability to extract temporal regularities in order to form accurate predictions about the upcoming signal has also been discussed by Yates et al. (2019) as a mechanism that underlay the significant correlation between beat perception skills and speech recognition thresholds in background noise.

In the present study, a novel method of training musical beat processing to benefit speech understanding in noise in individuals with age-related hearing loss is presented. Instead of a purely perceptual task, a synchronized tapping to music is proposed here. The presence of a motor component should be an advantage since multimodality of musical training has been shown to enhance neural plasticity (Lappe et al., 2011). The primary aim of the current study was to evaluate the feasibility of such a program in a clinical population of older hearing-impaired listeners. The training program was intended to be intuitive and easy to understand for people without formal musical knowledge, as well as suitable for unsupervised training at home. The participants' compliance with the training schedule and their ability to perform the task without supervision were taken into consideration. The second goal of the study was to investigate the impact of the training on outcome measures of synchronized tapping, musical beat perception as well as speech intelligibility in noise.

4.2 Methods

4.2.1 Participants

Fifteen 60-81 years old hearing-aid users were recruited from the volunteer database of Hearing Systems Section at the Technical University of Denmark as well as from patient databases at the hospital of the Capital Region of Copenhagen and the Odense University Hospital. All of them were native Danish speakers with a symmetrical bilateral hearing loss. At the moment of starting the experiment, all the participants had had at least half a year of experience using hearing aids. The participants signed an informed consent and were reimbursed on an hourly basis for their participation in the study. Two participants did not complete the study due to illness and a lost hearing-aid, respectively. The information about the thirteen participants that completed the study is summarized in Table 4.1. A more extensive data collection was impossible due to limited possibilities of recruiting and testing participants during the COVID-19 pandemic.

4.2.2 Study design

All the thirteen participants that fulfilled the eligibility criteria completed three visits. The first visit comprised otoscopy, pure-tone audiometry (unless a recent audiogram was available in the database), Montreal Cognitive Assessment (Nasreddine et al., 2005), Digit Span test (Wechsler, 2008), Goldsmith Musical Sophistication questionnaire (Müllensiefen et al., 2014) and Real-Ear Measurements. During the second and third visit, the participants completed speech-in-noise tests (DANTALE II; Wagener et al., 2003 and HINT; Nielsen and Dau, 2011) as well as a musical beat perception test (Beat Alignment Test; Iversen and Patel, 2008), described in more detail further below. The second and the third visit were interleaved with an intervention (training or control) period of three weeks. The flowchart illustrating the study design is shown in Figure 4.1.

The participants were assigned to one of the two groups. Each group received a different version of a mobile application and was asked to perform auditory exercises six days a week over the three-week period between the second and the third measurement visit. One version of the application was the training program aimed at improving the ability to tap along with the musical beat. The other version was the control intervention, where the participants

Table 4.1: Characteristics of the participants

No.	**Age**	**Sex**	**DS score**	**PTA better ear [dB HL]**	**HA brand**	**HA experience [years]**	**Gold-MSI**
Training group							
1	75	M	26	50.00	Oticon	8	50.00
2	72	M	17	43.75	Oticon	6	53.00
3	75	F	31	33.75	Signia	8	54.00
4	75	M	24	32.50	Siemens	2.5	56.00
5	60	F	33	43.75	Widex	30	23.00
6	61	F	42	38.75	Rexton	12	61.50
7	81	M	23	30.00	Oticon	2	44.00
Mean	**71.3**		**28**	**38.93**		**9.8**	**48.79**
Control group							
8	66	M	35	37.5	Rexton	0.5	58.00
9	79	M	23	48.75	Oticon	8	56.00
10	77	F	31	45.00	Rexton	5	42.00
11	70	M	26	48.75	ReSound	16	39.00
12	70	F	40	48.75	Bernafon	13	51.00
13	61	F	30	37.50	Widex	8	61.00
Mean	**70.5**		**30.8**	**44.38**		**8.4**	**51.17**

DS – Digit Span; PTA – pure-tone audiometry; HA – hearing aid; Gold-MSI – Goldsmiths Musical Sophistication Index

listened to the excerpts from audiobooks and answered multiple-choice questions.

All participants received a mobile phone (Motorola Moto G8) with the assigned version of the mobile application and a pair of headphones (Sennheiser 380). Each person was provided with verbal and written instructions during visit number two and completed one set of daily exercises under the experimenter's supervision. Each phone had a SIM card with mobile data and the results files were automatically uploaded to the on-line storage site. After the first and the second week of the training, the participants received an e-mail with a graph representing their progress accompanied with a congratulation message.

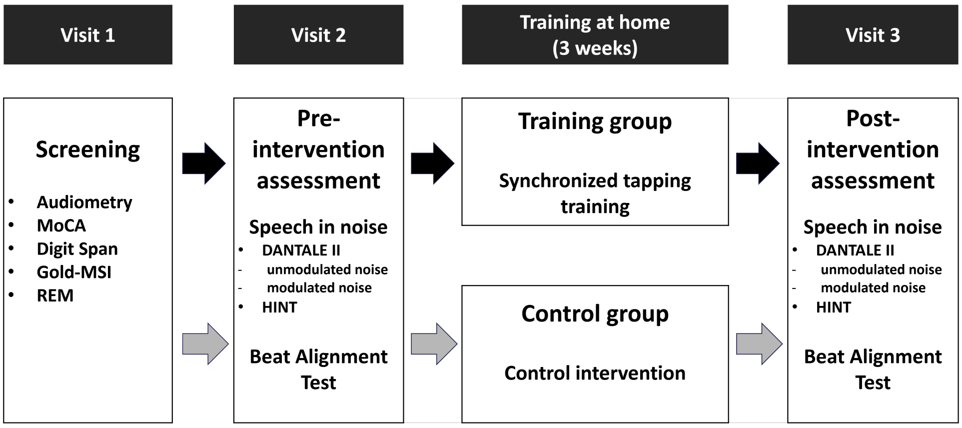

Figure 4.1: Study design. Participants completed three visits. The first visit comprised pure-tone audiometry, Montreal Cognitive Assessment (MoCA), auditory digit span test, Goldsmith Music Sophistication Index (Gold-MSI) questionnaire and Real-Ear Measurements (REM). Test of speech-in-noise (DANTALE II and HINT) as well as the Beat Alignment Test were performed upon the second and the third visit. During the three weeks between the second and the third visit, the participants completed the allocated program at home.

4.2.3 Training program

Stimuli and schedule

Thirty-six musical excerpts were selected from the database of the music collaboration platform Soundtrap (*https://soundtrap.com*). The nominal tempo of the excerpts varied between 60 and 150 beats per minute (bpm) and the duration was eight bars (14 excerpts) or four bars (22 excerpts). Most of the music examples had quadruple meter (with four beats per measure) except for two triple-meter examples (with three beats per measure). The stimuli were organized into a six-day long training schedule.The participants were asked to train six days a week and the schedule was repeated three times, resulting in eighteen days of training in total distributed over three weeks. Each week, the training started with tempos around 100 bpm, which is close to natural preferred tapping rate in humans (London, 2012). Gradually, slower and faster tempos were introduced in a step of 10 bpm per day (with the exception of tempo 80 bpm, which was trained on two consecutive days). For example, on the second day of the weekly schedule, the trained tempos were 90 bpm and 110 bpm, on the third day 80 bpm and 120 bpm etc. There were six training sessions per day (three in a slower and three in a faster tempo). During the first week, the participants were provided with visual cues to help them understand the task, as will be explained in more detail below.

The mobile application: user interface and procedure

The source code for the training application was written in PureData (Puckette, 1996) and run on the mobile devices using MobMuPlat software suite (Iglesia, 2016).

The user accessed the training session from the main menu of the application. The sessions' labels informed the user which day and week of the training they were assigned. After selecting the exercise, the user was directed to the training interface, which is shown in Figure 4.2. The interface was equipped with a *Start/Stop* button, a response button, a counter of late, early and on-time responses as well as - during the first week of the training - a panel providing the visual feedback.

The user triggered the trial by pressing the *Start* button. The music started playing and the user's task was to tap along to the perceived beat whenever they felt ready. Each tap would generate a 660 Hz beep and the beginning of each music bar was marked with a 440 Hz beep. The excerpts were looped such that the music played continuously until 80 tap times were collected. The offset of the music marked the end of the trial and the user could return to the menu to select the next session.

Two ways of ongoing feedback about the user's performance were provided. The first one was the visual representation using the panel with a circle. The circle's circumference represented the bar's duration and the empty dots corresponded to the occurrence of the beats. The filled dot represented the timing of the user's response. As the user started tapping on the response button, the filled dot changed its position on the circumference. The distance of the filled dot from the targets (empty dots) represented the accuracy, with a smaller distance indicating better accuracy. The goal of this representation was to help the user understand the task and to encourage tapping in a desired metrical level (i.e., avoiding a situation when the user would tap at double or half tempo) and phase (i.e., preventing the user from tapping in anti-phase, which would result in a filled dot being placed approximately in between the target positions). This visual aid was available to the user only during the first week of training, so that the user's synchronization behavior was based solely on the auditory input throughout the majority of the training.

The second feedback information was the overall number of early, delayed and on-time responses within the session, which was displayed on the screen and updated every time the user tapped the response button. The responses were counted as correct if they fell within an 80-millisecond interval before or after the beat.

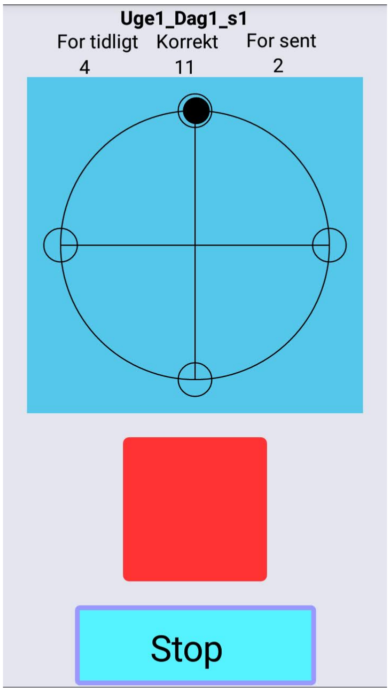

Figure 4.2: Training interface. The session label, the number of early (*For tidligt*), on-time (*Korrekt*) and late (*For sent*) responses, the panel with visual feedback (circle on the blue background), the response button (the red square) and the *Start/Stop* button are displayed on the screen.

Amplification and volume settings

One of the requirements set for the training was that it should be suitable for individuals with hearing loss. Thus, care had to be taken to compensate for the elevation of the hearing threshold by applying a frequency-dependent gain to the stimuli. All the stimuli were prefiltered using the 'Cambridge formula' (Moore and Glasberg, 1998) based on the individual hearing thresholds averaged across the ears. For each participant, a set of stimuli with a personalized gain was used.

A stationary calibration noise with a long-term spectrum matching the overall spectrum of the musical stimuli (before the filtering) was used to adjust the volume settings on the devices. The volume control was set in a position corresponding to output sound pressure level (SPL) equal to 65 dB. Before the filtering, all the musical stimuli were equalized to the root-mean-square value of the stationary noise used for calibration. During the first training session guided by the experimenter, the participants had a chance to increase or decrease the music volume to achieve the comfort level but were asked not to change the volume setting on the device when training at home.

4.2.4 Outcome measures

Speech intelligibility assessment

Two Danish sentence tests varying in terms of temporal predictability were used in the experiment: a matrix sentence test with a regular rate of word utterance (DANTALE II) and a test with more natural speech rhythm (HINT). Both tests were conducted using the recordings of a male talker and were administered in the presence of stationary background noise with the same long-term spectrum as the speech.

To investigate if the training benefit varied for different masker types, DANTALE II was additionally administered in the presence of a modulated masker. A sinusoidal amplitude modulation at a frequency of 8 Hz and a modulation depth of 80% was applied to the speech-shaped stationary noise. These noise modulation parameters were shown to provide good test-retest reliability for an English matrix test (Yates et al., 2019).

The tests were administered in a listening room fulfilling the IEC 268-13 standard at the Technical University of Denmark, Kongens Lyngby, or in an acoustically treated booth at the University of Southern Denmark, Odense. The stimuli were presented from a loudspeaker located at a one-meter distance from the participant’s seat. The speech and the masking noise were co-located and both initially presented at 65 dB SPL (before the speech level was adjusted according to the staircase procedures described further below).

The participants were instructed to repeat each sentence as precisely as possible and encouraged to guess if in doubt. A native Danish-speaking audiologist scored the responses. The participants were wearing their hearing aids during the test.

DANTALE II DANTALE II (Wagener et al., 2003) consists of five-word Hagerman-type sentences in Danish. Each sentence has the same semantic structure:

name-verb-numeral-adjective-object. The sentences are semantically correct but not necessarily meaningful (e.g. *Linda købte fjorten nye huse.*, English: *Linda bought fourteen new houses.*). The test material consists of 160 unique sentences organized into 16 ten-sentence lists. Ten alternatives for each word exist in the corpus (i.e., ten different names, verbs, etc.).

Three training lists (30 sentences) were administered for each visit and masker type to ensure that the learning effects had minimal impact on the test results. The training lists were followed by three test lists (30 sentences) per noise condition. The order of unmodulated and modulated condition was balanced across the participants.

An adaptive procedure was used to define the speech recognition threshold (SRT), which is defined as the signal-to-noise ratio corresponding to the performance of 50% correctly repeated test items (here: words). After each sentence presentation, the speech level was adjusted based on the number of correctly repeated words in the sentence. For the first five sentences within each training or test session, the step size was -3 dB when five words were correctly repeated, -2 dB for four correct words, -1 dB for three correct words and +1, +2 +3 dB when two, one or none of the words were correctly repeated. For the remaining sentences, the respective step sizes were: -2, -1, 0, 0, +1 and +2 dB.

HINT Danish HINT (hearing in noise test; Nielsen and Dau, 2011) uses 5-word sentences, which - in contrast to the matrix test - do not follow the same semantic structure.

One training and one test list (20 sentences each) were administered upon the second and the third visit. The SRTs were defined using a sentence-scoring procedure: after a correctly repeated sentence, the level of the speech was decreased by 2 dB and after an incorrectly repeated sentence increased by 2 dB. For the first five sentences, these step sizes were doubled.

Beat Alignment Test

To examine the participants' ability to tap along with acoustic stimuli as well as their perceptual sensitivity to musical beat, the following subsets of the Beat Alignment Test (BAT; Iversen and Patel, 2008) were used:

1. Synchronized tapping
   Participants were presented with twelve music excerpts representing three

music genres: rock, jazz and orchestral pop (average duration: 15.9 seconds) and were asked to tap along to the perceived beat. Each excerpt was played twice (the second presentation directly after the first one). The order of the excerpts was randomized across the participants.

2. Perceptual judgement of the beat
   Participants were presented with the same musical excerpts as in the previous subtest. Each piece was presented with a superimposed sequence of beeps, starting five seconds after the onset of the music. The beeps were regularly spaced in time and were either aligned or misaligned with the beat of the music. Two types of misalignment were used: a phase shift and a tempo change. In the phase shift condition, the beeps were presented with the inter-onset-interval (IOI) matching the tempo of the piece but anticipating or lagging after the actual beat by 30% of the inter-beat-interval (IBI). In the tempo change condition, the beeps were presented at IOIs equal to 0.9 or 1.1 of the actual IBI (tempos 10% slower or faster compared to the actual tempo).

   Each of the twelve music excerpts was presented three times: once in the on-beat and once in each of the two off-beat conditions, resulting in 36 trials in total. The early and delayed beep sequences were distributed evenly across the twelve trials with the phase shift. The slowed down and accelerated beep sequences were distributed evenly across the trials with the tempo change.

The stimuli presentation and response acquisition were controlled by MATLAB-implemented software. Acoustic stimuli were presented via Sennheiser HD 650 or Sennheiser HDA 200 headphones (Sennheiser, Germany). Tapping times were registered using a SubZero MiniPad MIDI Controller (Gear4music, UK). The participants performed the testing without their hearing aids. To provide audibility, the whole set of the stimuli was pre-processed using the ’Cambridge formula’ (Moore and Glasberg, 1998) based on each individual’s hearing thresholds. At the beginning of each measurement session, the participants were presented with examples of music and a beep sequence and were asked to adjust the volume to the most comfortable level.

4.2.5 Analysis

Synchronized tapping task: training data

The participants' performance in the synchronized tapping task throughout the training was analyzed using circular statistics (Fisher, 1993; application of the method to tapping data: Dalla Bella and Sowinski, 2015). The analysis was run in MATLAB using the CircStat toolbox (Berens, 2009). The participants' responses were mapped on the polar coordinate system, where 0°angle marks the occurrence of the beat. Each tap time relative to the inter-beat-interval (IBI) was converted to degrees and treated as an angle of a unitary vector. A mean resultant vector $\overrightarrow{R}$ was computed for each training session. The length of the $\overrightarrow{R}$ vector (between 0 and 1) represents the consistency of tapping; the closer to one, the more consistent was the performance. The angle θ of the $\overrightarrow{R}$ vector represents the accuracy of tapping; the smaller the angle, the smaller, on average, was the interval between the finger tap and the beat. A negative θ value means that the subject tends to tap before the beat and a positive value indicates that the taps occur after the beat. Figure 4.3 shows an example of using circular statistics to analyze synchronized tapping data.

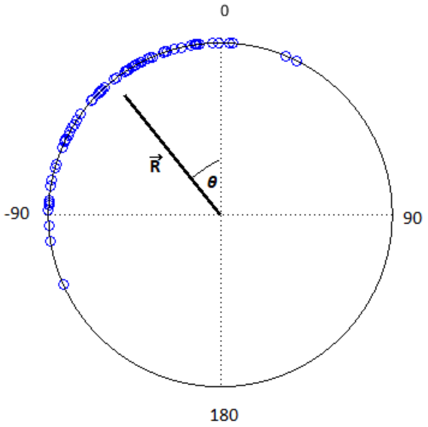

Figure 4.3: An example of using circular statistics to analyse synchronized tapping data. Tap times relative to the inter-beat-interval (IBI) define the positions of the blue dots on the unit circle. The length of the mean resultant vector $\overrightarrow{R}$ represents the tapping consistency and the angle θ - tapping accuracy.

For each subject, a mean consistency per week was computed. A meaningful accuracy value can be obtained only if the distribution of tap times is not random, i.e., when there is a phasic relationship between the timing of beat and the timing of the responses. Therefore, the distribution of the responses for each training session was assessed using a Rayleight test. A significant result of the test indicated that the response pattern was not random. Only those trails, for which Rayleigh test yielded a significant result and the vector length was greater than 0.4 (Dalla Bella et al., 2017) were considered when calculating average accuracy per week. The number of trials with a random tapping pattern was registered for each participant.

All participants' response times used for the calculations described in this section were corrected by subtracting an estimated system delay. The delay was computed as the difference of two time intervals: 1) the interval between the timestamp marking the bar onset and the timestamp of the participant's response saved in the result file and 2) the interval between the beep generated at the bar onset and the finger tap extracted from an audio file recorded with an external microphone. The average delay computed for 20 trial tap times equaled 164 ms (standard deviation (SD) = 6 ms).

Beat Alignment Test

The test setup used in the tapping subset of BAT allowed for obtaining the time intervals between the participants' taps but not the differences between the taps and the beats. Therefore, applying circular statistics to this dataset was not possible and alternative analysis method based on the inter-tap-intervals (ITIs) was proposed.

Two series of ITIs were registered for each of the twelve music excerpts per participant and visit. The data was cleaned by removing artifacts and outliers according to the criteria used in Dalla Bella et al. (2017). An ITI was considered an artifact when its duration was shorter than 100 ms. Outliers were defined as ITIs shorter than $Q_1 - 3 * IQR$ or longer than $Q_3 + 3 * IQR$ (where Q_1 indicates the first quartile, Q_3 denotes the third quartile and IQR represents the interquartile range of the sample). Trials in which fewer than eight consecutive taps not considered outliers were registered were discarded from the analysis. The mean absolute error (MAE) was computed for each of the remaining trials by averaging the differences between the ITIs and the IBI of the piece. The IBIs were established based on the values reported by the authors of the test (Iversen and Patel, 2008). If the participants' average tapping rate was closer to half (or one third for one piece with a triple meter) or double the value of nominal IBI, then the IBI was adjusted accordingly before the calculation. The MAE value for each of twelve excerpts was obtained by averaging the MEAs calculated for two consecutive presentations of a given piece. Each of the twelve MAE values were then divided by IBI. The resulting values, which represented the tapping error as a proportion of the IBI, were then averaged across the excerpts to obtain one number representing the synchronized tapping performance per subject and visit.

For the perceptual subset, the sensitivity index *d'* was computed based on the number of hits (trials, in which the misalignment was correctly detected) and the number of false alarms (trials, in which the misalignment was incorrectly reported).

Speech intelligibility

A mixed-linear model approach was applied to the results of each speech test with a main effect of *Visit* and *Group* and the interaction of these effects. The random effect of *Participant* was also included in each model. The analysis was run in R (R Core Team, 2020) using *lmerTest* package (Kuznetsova et al., 2017).

4.3 Results

4.3.1 Training compliance and performance

Training compliance was calculated per individual as a percentage of the completed sessions (out of 108 sessions in total). Out of seven participants that were assigned to the training group, five showed excellent compliance between 96.3% and 100%. The two participants with lower compliance completed 39.81% (participant 3) and 71.3% (participant 4) of the training sessions. Participant 3, who completed less than half of the training, was not considered in the analysis.

Figure 4.4 represents the average weekly consistency (upper left panel) and accuracy (upper right panel) for each of the seven participants, as well as the percentage of rejected trials per participant and week (bottom panel).

In terms of the synchronization consistency, two subgroups of the participants can be easily distinguished: three individuals (no. 1, 2 and 5) showed better (higher) consistency than the other three (no. 4, 6, 7). Such a grouping of the participants is in agreement with a cutoff value of 0.51, which was proposed by Dalla Bella and Sowinski (2015) as a criterion for detecting poor synchronizers, i.e., people who experience difficulties synchronizing their movement

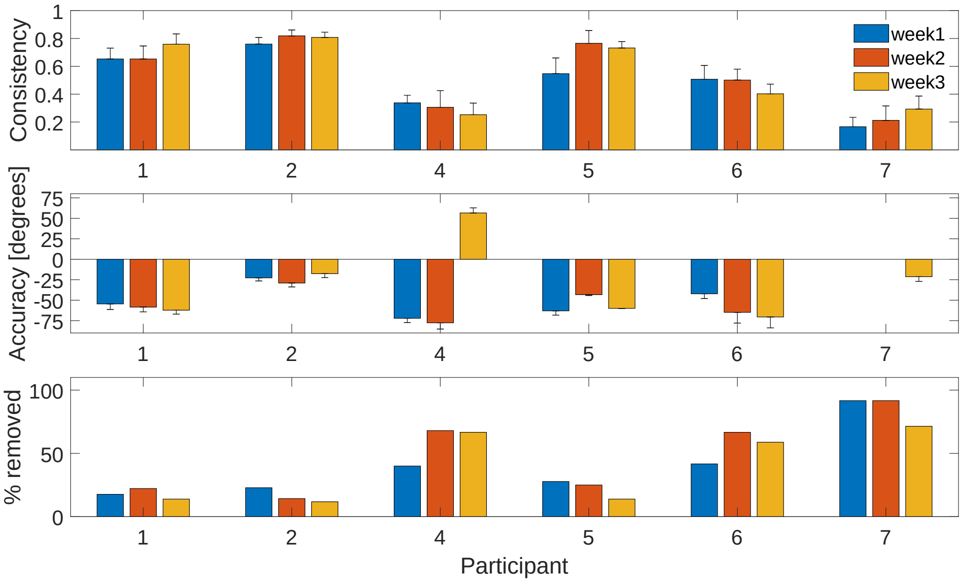

Figure 4.4: Average synchronization consistency (upper panel) and accuracy (middle panel) as well as the percentage of rejected trials (bottom panel) for each participant and week of training. The first week is represented with blue bars, the second week with orange bars and the third week with yellow bars. Errorbars show one standard error of the mean.

with auditory rhythms. A subdivision into groups of good and poor performers was not observed in synchronization accuracy data. Nevertheless, the amount of rejected trials was much higher for poor synchronizers than for the good synchronizers.

In two of the participants from the poor synchronizers group (no. 4 and 6), a substantial increase in the number of rejected trials (i.e., trials with a random tapping pattern) was seen after the first week. This effect might have been linked to the removal of the visual cue. The good synchronizers seemed unaffected or affected to a much lesser extent (e.g., participant no. 1) by lack of the visual cue, both regarding consistency and the proportion of rejected trials.

Since the change in performance between the first week of training and the following two weeks was most likely affected by the removal of visual cues, the difference between the average scores for the third and the second week was used to represent the on-task improvement. The changes in consistency are summarized in Table 4.2. Consistency is a more sensitive metric to characterize the improvement than accuracy, which is based on a different proportion of trials for each individual. The consistency scores improved by 2.6, 7.9 and 15% in the group of good synchronizers and by 0, 4.3 and 5.4% in the group of poor synchronizers.

Table 4.2: Changes in tapping consistency from week 2 to week 3

No.	Consistency week 2	Consistency week 3	Delta consistency (week 3 - week 2)	Change in %
1	0.65	0.75	0.10	15
2	0.77	0.79	0.02	2.6
4	0.34	0.34	0	0.00
5	0.63	0.68	0.05	7.9
6	0.37	0.39	0.02	5.4
7	0.23	0.24	0.01	4.3

4.3.2 The effect of training on synchronized tapping and beat perception

Figure 4.5 shows the results of the Beat Alignment Test obtained before (blue boxes) and after (orange boxes) the training and control intervention, respectively. The data for the synchronized tapping task are displayed in the left panel and the data for the perceptual task are represented in the right panel. Smaller values of MAE (expressed as a proportion of IBI) are associated with a better synchronized tapping performance. Higher *d'* values indicate greater perceptual sensitivity to beat misalignment.

The analysis of the BAT scores did not show a significant *Group x Visit* interaction, nor main effects of *Group* and *Visit* for any of the test subsets. Visual inspection of the data reveled that the numbers of participants, whose performance improved from the pre- the post-intervention visit was roughly the same in both groups.

4.3.3 The effect of training on speech intelligibility

The results of the speech test are shown in Figure 4.6. The statistical analysis revealed no significant *Group x Visit*, *Group* or *Visit* effects for DANTALE II in unmodulated noise and HINT. For DANTALE II in modulated noise, a significant effect of *Visit* was found, ($F(1,10) = 40.58, p < .001$), indicating an overall improvement of the performance, regardless of the received intervention.

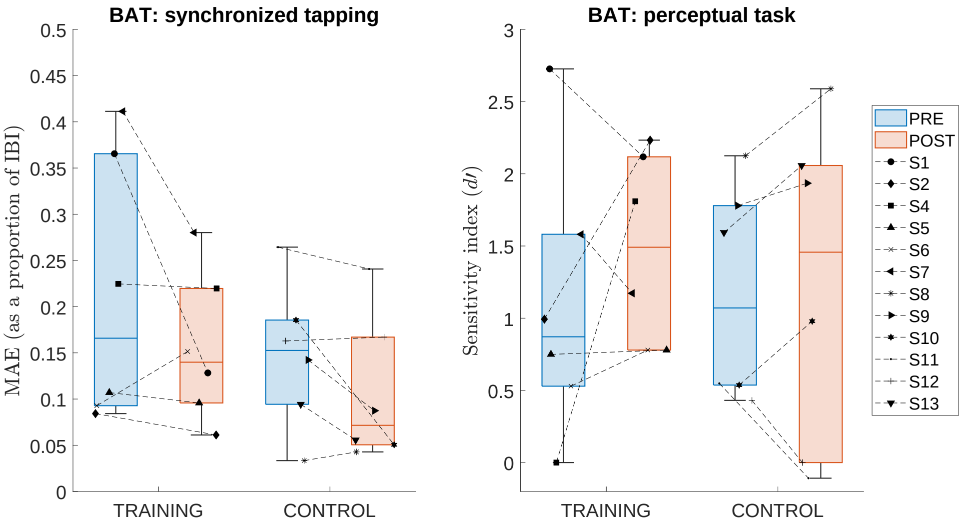

Figure 4.5: The outcomes of the Beat Alignment Test: the synchronized tapping task (left panel) and the perceptual task (right panel), administered before (blue) and after (orange) the intervention. The data for the training group are shown on the left side of the panels and the results for the control group are displayed on the right side of the panels. The horizonatal lines in the middle of the boxes represent the medians of the samples. The boxes' bottom and upper edges indicate the 25^{th} and the 75^{th} and percentile, respectively. The whiskers correspond to the most extreme observations not categorized as outliers. Each participant is represented with a different symbol and the observations from the same subject are connected with a dashed line.

4.4 Discussion

The study presented a novel rhythm-based musical training method intended to improve speech-in-noise perception in elderly hearing-impaired listeners. The participants' compliance and abilities to perform the task without supervision were considered in evaluating the suitability of this program for training at home. Furthermore, the impact of training on speech intelligibility was explored.

4.4.1 The outcomes of the training

The overall compliance was high, with five participants who completed more than 96% of all training sessions, one participant who completed around 70% of all training sessions and one person with compliance lower than 40%. These results are quite optimistic, however, the lower compliance rates in two participants might indicate that the training could have had more features oriented towards maintaining high motivation and engagement in the training process.

The ability to perform the task was assessed using tapping data analyzed with circular statistics. This analysis method allowed for identifying sessions

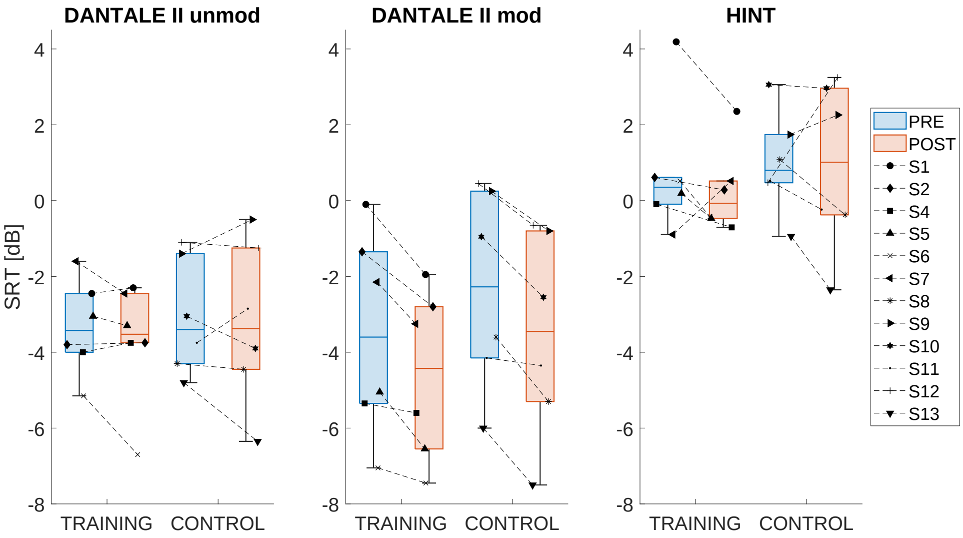

Figure 4.6: The results of the speech intelligibility tests: DANTALE II in unmodulated noise (left panel), DANTALE II in modulated noise (middle panel) and HINT (right panel). The boxplots showing data for the training group are displayed on the left and the boxplots representing the control group are displayed on the right side of each panel. The horizonatal lines in the middle of the boxes represent the medians of the samples. The boxes' bottom and upper edges indicate the 25^{th} and the 75^{th} percentile of the sample, respectively. Blue and orange colors represent, respectively, the pre-intervention and the post-intervention visit. The whiskers show the most extreme observations not categorized as outliers. Each participant is represented with a different symbol and the observations from the same subject are connected with a dashed line.

in which the participant's tapping pattern was random (i.e., did not show any phasic relationship to the musical beat). Such a response pattern indicated that the participant was unable to perform the task. Three out of the six participants included in the analysis tapped at a random phase in more than 50% of all sessions. The fact that the training caused substantial difficulties for half of the participants indicates that the current version of the application is not suitable for unsupervised training at home. The reasons for these substantial difficulties in performing the task should be further investigated.

The second goal of this study was to explore the potential transfer of the training effect from music to speech domain. The results did not show any training-related performance improvements in the speech-in-noise tests. The only condition in which the SRTs improved from pre- to post-intervention visit was DANTALE II in the presence of modulated noise. However, the improvement was observed in both groups and hence, was most likely caused by factors unrelated to training, for example, increased familiarity with the test procedure and material.

A robust improvement in the musical skill targeted by the training should be the prerequisite for any potential transfer to speech perception. Lack of significant training effect on the two BAT subsets as well as very modest increments of the synchronized tapping consistency during the training indicate that the ability expected to support speech-in-noise perception was not developed to a sufficient degree. Therefore, the presence or absence of transfer between the two domains could not be validly assessed in this experiment.

4.4.2 Possible improvements in the training method

The fact that training failed to improve the targeted skill might stem from the limitations of the applied method. The major constraint of the training was its short duration. The participants trained for 18 days, with daily exercises estimated to take 10-15 minutes. For 100% compliance, such a schedule would result in 180-240 minutes in total. It is not known if a longer duration or higher intensity of training would have given better results. Bégel et al. (2018) reported a significant improvement in BAT results after a two-week training with a perceptual rhythmic game in young normal-hearing adults. The total time spent playing the game during the training period was 300 minutes. Despite the differences in task and population, the study of Bégel et al. (2018) points towards the need for a longer and/or more intensive training schedule.

Another limiting factor in the present version of the training program might have been the artificial character of the stimuli. The musical excerpts used during the training had a duration of four to eight bars and were replayed in a loop until the user provided the required number of responses. Such a degree of repetitiveness and predictability is rarely experienced in real music. The task in which the user got familiar with the tune after a few first presentations might not have placed enough demands on the predictive mechanisms of beat perception. The lack of improvement in BAT might suggest that training in these artificial conditions did not generalize to more 'natural' and therefore 'unpredictable' music. In such a situation, any transfer of learning to speech, which has even more unpredictable rhythm, is very unlikely. Perhaps, less repetitive and more diverse music material should be used in this type of training in the future.

Successful training might also require a more individually-oriented approach. The emerging subdivision between the good and the poor synchronizers suggests that these two groups might need slightly different training protocols. Interestingly, two of the participants classified as poor synchronizers (no. 4 and 6) showed substantial worsening of their performance after the first week of training, which is most likely linked to the unavailability of visual feedback from the second week. These two participants seemed to rely more heavily on the visual cues than the good synchronizers. In the future, more individually-tailored training could be provided by identifying the subgroups of poor and good performers and administering two different program versions: one for the good synchronizers, where the amount of visual aid will not be increased and one for the poor synchronizers, with the visual feedback available until further stages of the training or throughout the whole duration of the training.

Finally, the synchronized tapping to the beat of music might not have been the optimal choice of the training task. The selection of the task was based on a strong relationship between the beat perception (measured in BAT) and speech recognition thresholds (Yates et al., 2019). An alternative training method informed by this correlation could target the perceptual sensitivity to misalignment between a pacing stimulus (e.g., metronome) and the beat of music. However, a multimodal character of the tapping task appeared as an advantage, as it was expected to enhance training-induced neural plasticity (Lappe et al., 2011). Ironically, it is possible that the nature of the underlying mechanism that is most relevant for the potential transfer to the speech domain was lost by introducing the motor aspect. In fact, Dalla Bella et al. (2017) reported a lack of correlation between the perceptual sensitivity measured in the BAT and the paced tapping performance. A similar tendency was observed in the present study; poor synchronizers did not show a poorer sensitivity in the perceptual task than good synchronizers. These findings suggest that sensitivity to misalignment between the beat of music and a metronome and motor synchronization ability might involve different beat-processing mechanisms. Hence, their relevance for speech processing may also differ. Therefore, an alternative rhythm-based training program that involves a purely perceptual task is worth considering.

4.5 Summary and conclusion

The present study investigated the feasibility of using a rhythm-based remote training program in a group of older hearing-impaired listeners as a potential tool to improve their speech perception. Six out of seven participants showed good to excellent training compliance. However, the training performance data indicated that the task demands were not well suited for all the participants and that the improvement in the trained synchronized tapping task was rather modest. No effect on speech intelligibility in noise was found, probably due to insufficient development of the trained skill in the music domain. Any potential benefit of such a rhythmic training approach can only be revealed if more individually-tailored task demands are provided to the participants (e.g., poor synchronizers should receive more visual feedback). Longer duration as well as purely auditory character of the training should also be considered for the future improvement of the method.

5

Overall discussion

This thesis investigated the effect of auditory training on the ability to understand speech in noise in hearing-impaired listeners who use hearing aids. Two auditory training approaches were considered: one based on the recognition of Danish phonemes and one based on the motor synchronization with musical stimuli. Three major research questions aimed to expand the current knowledge about these two respective approaches were addressed in the three main chapters of the thesis:

1. Does phoneme-in-noise identification training improve speech intelligibility in an untrained sentence material?
2. Does phoneme-in-noise identification training alleviate the cognitive effort required to understand sentences in noise?
3. Can home-based training of synchronization with musical beat be used as intervention to improve speech intelligibility in elderly hearing-aid users?

5.1 Summary of main results

In *Chapter 2*, the effect of phoneme-in-noise identification training on speech intelligibility was investigated. The training method was based on an existing training program (Schoolo; Serman, 2012), which had previously shown promising outcomes in cochlear-implant recipients (Schumann et al., 2015) and normal-hearing listeners (Schumann et al., 2016). In the present study, the program was used with a Danish nonsense speech material (DANOK; Nielsen and Dau, 2019). The structure of DANOK logatomes allowed for training three categories of phoneme targets: an onset consonant (C1), a medial vowel (V) and a post-vowel consonant (C2). The outcome measures included a phoneme identification test (which also used the DANOK material but the talker and procedure were different from those used throughout the training) and a hearing in noise test (HINT; Nielsen and Dau, 2011). A two-week long training comprising six visits was administered in a group of 63-79 years old hearing-aid users with a mild-to-moderate hearing loss. Throughout the training, an adaptive rule was applied to adjust the signal-to-noise ratio based on the user's performance. A decrease in the mean SNR within a training session (interpreted here as the performance improvement) was found in the sessions where the users had to identify the onset consonant (C1) and the vowel (V) but not the post-vowel consonant (C2). The training led to an improvement of the phoneme-identification for the vowels (V) and the post-vowel consonants (C2) but not for the onset consonants (C1). The average improvement in the HINT was more prominent in the training group than in the control group, but the effect did not reach statistical significance. After three months, the significant increase of the phoneme identification scores (V and C2) in the training group were still present. Overall, the results presented in *Chapter 2* showed that a relatively short duration of training improved phoneme identification in an untrained context and with a novel talker for two out of three considered target positions. This effect persisted over time. However, no evidence for generalization to sentence intelligibility in noise was found.

The study results presented in *Chapter 2* reflect a 'traditional' approach to assessing the efficacy of auditory training by comparing the difference between the pre- and post-intervention intelligibility scores across two groups: a group that received the training and a control group. In *Chapter 3*, measuring listening effort was proposed as a method to complement the intelligibility metrics. Listening effort was characterized by pupil dilations obtained while the participants performed a HINT at two fixed SNR conditions: SNR50 - corresponding to 50% correctly repeated words - and SNR80 - corresponding to 80% correctly repeated words. These two SNR values were defined before the intervention and were held constant across each individual's pre- and post-intervention assessment. The analysis of the HINT results revealed a significant improvement in terms of the percentage of correctly repeated words in the training group in the SNR50 condition but not the SNR80 condition. The pupillometry data were analyzed in two ways: 1) by evaluating the change in the peak pupil dilation (PPD) and 2) by evaluating the change in the magnitude and the shape of the response in the temporal window spanning five seconds from the onset of the sentence. The PPDs increased significantly in both groups, with no distinction between the trained and the untrained listeners. However, the analysis of the longer temporal window revealed that the training did affect the magnitude and morphology of the pupil response. These pupillometry results suggested that listening effort increased both in the training and in the active control group, with a more prominent difference in the training group. While motivation was presumably driving the effect in the control group, the change in the fatigue status and overall cognitive capacity might have contributed to the effect observed in the training group. Even though the effect of training on speech intelligibility was not the main focus of this chapter, a significant training-induced improvement of the recognition score for an untrained speech material was found for the more demanding (SNR50) of the two HINT conditions.

While *Chapter 2* and *Chapter 3* employed an existing phoneme-in-noise training paradigm, *Chapter 4* proposed a novel music-based training method with the aim of providing a benefit in speech understanding in noise. Seven 61-81 years old experienced hearing-aid users completed an approximately three-week long remote rhythmic training. The training task was to tap in synchrony with the beat of a musical piece. Another six individuals formed an active control group that matched the training recipients in terms of age, degree of hearing loss and years of hearing-aid usage. Five out of the seven participants assigned to the training group showed an excellent compliance with the training program (96-100%), whereas one of the listeners showed a poor compliance and was therefore excluded from the analysis. Tapping consistency was used to evaluate the participants' performance and indicate the progress throughout the training. Half of the participants who completed the training (n = 3) were classified as "good synchronizers". The other half (n = 3) were categorized as "poor synchronizers" as their performance lay below the consistency score cut-off value of 0.51 proposed by Dalla Bella and Sowinski (2015). The increments of the consistency scores – which represent the performance improvement – were modest in both groups and equaled 2.6, 7.9 and 15% for "good synchronizers" and 0, 4.3 and 5.4% for "poor synchronizers". The results of the Beat Alignment Test (BAT; Iversen and Patel, 2008) administered before and after the training did not reveal any significant improvement both in terms of the synchronized tapping to music and in terms of the perceptual sensitivity to beat misalignment. The lack of a significant effect in BAT suggested that the training was insufficient to improve the targeted skill. Thus, any transfer to speech intelligibility outcomes was very unlikely to have happened. Consistently, no significant effect of training was observed in the speech recognition results obtained with three combinations of speech tests and maskers: 1) DANTALE II in the presence of stationary noise, 2) DANTALE II presented in modulated noise and 3) HINT presented in stationary noise. The significant main effect of *Visit* for the SRTs in DANTALE II in modulated maskers indicated that the threshold decreased by about the same value regardless of the group assignment, thus reflecting procedural rather than training-related learning. Overall, the results presented in *Chapter 4* did not support clear conclusions related to the potential transfer of synchronized tapping training to the speech domain due to the small sample size and/or lack of improvement in the trained skill itself. The results suggested, however, that a longer and more individually-tailored training method might help to address this question properly.

5.2 Comparison of different approaches to auditory training

5.2.1 Speech-based versus music-based training

Both training programs investigated in this dissertation had a relatively short duration (two-three weeks). As reported in *Chapter 2*, a two-week-long phoneme-in-noise training was sufficient to lead to an apparent and a significant improvement of the phoneme identification scores for two out of three phoneme categories. Moreover, a significant effect on sentence intelligibility in the SNR50 HINT condition was observed in the same group of listeners, as described in *Chapter 3*. On the other hand, no indication of an improvement in the trained skill and in terms of speech intelligibility was found after a three-week-long music-based training investigated in *Chapter 4*. These findings suggest that for these two domains, different amounts of time might be needed to develop new skills and enable generalization to new conditions (in the case of speech-based training) or to transfer to the other domains (in the case of music-based training).

These discrepancies can be discussed in light of the Reversed Hierarchy Theory (Ahissar et al., 2008). The two different approaches targeted different levels of representation in the auditory hierarchy. Rhythm perception is based on the encoding of the temporal envelope of the signal (Fujii and Wan, 2014). Encoding of such a basic stimulus parameter corresponds to the lowest levels of the sensory stimulus representations. On the contrary, phonemes are more complex categorical entities encoded by higher hierarchy levels. Out of these two types of representations, the modification of the lower-level sound envelope encoding might require a more extended training regime (Ahissar et al., 2008). Even more time might be needed to allow for a generalization to the global context. Studies that reported benefits of the musical training on speech-in-noise perception in elderly listeners with and without hearing loss employed protocols that were substantially longer than three weeks; i.e., ten weeks of choir singing (Dubinsky et al., 2019) or six months of piano lessons (Zendel et al., 2019).

Considering the time course of learning effects, speech-based auditory training might be a more time-efficient method than music-based training. Music-based interventions might require much more commitment and would need to be practised continuously with great diligence to enable any meaningful benefits for speech-in-noise perception.

5.2.2 In-clinic versus at-home training

The two training approaches presented in this dissertation differed in terms of the administration method. The phoneme-in-noise training was delivered under supervision in the study location. The rhythmic training was self-administered by the participants at home. The method of administration (clinic-based versus home-based) has implications for the ease of access to the training but also affects the adherence to the training schedule.

The key advantage of home-based training is its low cost and high resource effectiveness. Moreover, it increases the accessibility of the training for people with reduced mobility and those living far from the clinic. With the increasing use of technology, home-based training appears as an attractive alternative to the more traditional model which involves meeting the clinician in person. However, the home-delivered training seems to create challenges associated with the participants' adherence to the training schedule.

Several studies that examined the effectiveness of computerized home-based auditory training programs demonstrated a relatively low compliance among the participants. Sweetow and Sabes (2010) reported a compliance of 30%, whereas Abrams et al. (2015) mentioned that the participants completed slightly above 50% of the recommended training time. In the present investigation, the only challenges with compliance occurred in the case of home-based rhythmic training but not in the case of the phoneme-in-noise training that took place in the clinical research facilities. This finding fits the general trend that compliance seems to drop when the training is unsupervised. Noncompliance with the training schedule may lead to sub-optimal outcomes of the intervention (Chisolm et al., 2013).

While home-delivered solutions may facilitate incorporating the auditory training in daily routine, more needs to be done to maintain good training compliance in the long term. One way to achieve this goal would be using gamification to make the training more engaging (Lumsden et al., 2016). On the other hand, involvement of the clinician in the training process - even if not necessary on a daily basis - might be crucial to keep the patient's motivation high. Regular contact and an opportunity to discuss the progress and intervention outcomes with a hearing care professional might be more stimulating than an 'anonymous' feedback message provided by the training software.

5.3 Assessing the training efficacy

5.3.1 Identifying appropriate outcome measures

In their systematic review of evidence for the efficacy of computer-based auditory training, Henshaw and Ferguson (2013a) emphasized the importance of an appropriate outcome measure selection for an accurate characterization of the training effects. An "appropriate" outcome measure should be characterized by a good test-retest reliability and be sensitive enough to detect small effect sizes. It should allow for avoiding floor and ceiling performance and - ideally - reflect the real-world benefit. Henshaw and Ferguson (2013a) also postulated a standardization of outcome measures to facilitate comparisons between different studies and enable future meta-analyses. Despite the variety of available methods of audiological assessment, the choice of the outcomes that would fulfill these criteria is not trivial.

Speech-in-noise perception

Using well-established speech-in-noise paradigms like HINT and DANTALE II has substantial advantages. Normative data for these tests are available, making the results easier to interpret. Thanks to the fact that both HINT and Hagerman matrix sentence test have counterparts in many languages, it is possible to compare results obtained with different language populations. A limitation of these tests is that they have been validated for SRT measurements in the presence of a stationary speech-shaped noise, which is a rather artificial condition - in terms of the masker type, the spatial setup and the targeted performance level (50%). Capturing the benefits of training for higher performance levels might be more relevant for real-life situations. However, these "traditional" speech tests are not sensitive in such conditions, as demonstrated in *Chapter 3*.

For the sake of obtaining a more "ecologically valid" masker condition, the stationary speech-shaped noise could be replaced with more natural background noise, e.g., speech-fluctuating noise, environmental noise, multi-talker babble, etc. Nevertheless, it is not known how such manipulations would affect the test-retest reliability of the SRT estimation. For example, *icra5* noise (that represents the modulations of a single male talker) was shown to increase the test-retest SRT difference compared to stationary maskers for DANTALE II (Wagener and Brand, 2005). Another example of how the presence of a novel masker can influence the test reliability was shown in *Chapter 4*. The thresholds obtained with DANTALE II in the presence of an amplitude-modulated noise were, on average, lower at the post-intervention visit. The trend was statistically significant, consistent across the participants and independent of group assignment. Hence, it could not be attributed to the effect of rhythmic training. Since the participants were blinded to the intervention they received, one could argue that this effect could have aroused from active participation in the study (e.g., increased self-confidence due to expected improvement). In such a case, however, a similar consistent improvement in SRTs should also have been observed for DANTALE II in the presence of stationary noise, which was not the case. The observed improvement in thresholds could not be attributed to the learning of the test vocabulary either, since the recommended number of training sentences (30; Wagener et al., 2003) was completed during each visit and for each masker. The improvement was specific only to this particular test condition and, hypothetically, linked to the increased familiarity with the fluctuating masker. This finding demonstrates that any speech-in-noise paradigm that has not been validated should be treated with caution as novel maskers can lower the reliability of the test, limiting its suitability as an outcome measure.

Listening effort

Studying the impact of auditory training on listening effort is of interest from the perspective of functional, real-life benefits. Limiting the effort required to follow a conversation is a highly desired outcome since it might encourage social participation and have a meaningful, positive impact on the quality of life. Recent findings demonstrate that the majority of measures considered to reflect listening effort show good test-retest reliability (Alhanbali et al., 2019), which indicates their suitability for measuring the outcome of the training intervention at two points separated in time. On the other hand, different metrics of listening effort are not necessarily correlated with each other. When choosing the outcome measure, it is important to consider which dimension of effort is assessed and how the result should be interpreted.

In *Chapter 3*, listening effort before and after the training was characterized using pupillometry. Reduced pupil dilation has been associated with lower effort and has traditionally been interpreted in the literature as a positive outcome, e.g., when assessing the effectiveness of a noise reduction algorithm in hearing aids (Wendt et al., 2017). Nevertheless, larger pupil dilation should not be treated as an undesired finding in some situations. Winn et al. (2018) argued that a larger dilation might reflect higher engagement in the task, which should be treated as a success of an audiological intervention, especially in the longer term.

The findings of *Chapter 3* suggest, however, that higher engagement may not necessarily stem from the fact that the intervention was successful but just from the fact of receiving an intervention. People who committed to several weeks of auditory training might want to prove themselves that this investment was worth time and trouble. In such a situation, the effort investment during the test will be high, leading to larger pupil dilations. The sensitivity of pupillometry to aspects related to participant's motivation makes the interpretation of the results rather challenging. Another limitation of this method is that it reflects a momentary, task-evoked processing load. Hence, the result might not be relevant for more realistic situations, like continuous listening to speech in the presence of competing talkers for several minutes.

Investigating the impact of auditory training on listening effort remains a topic for future research on auditory training. It would be interesting to measure the effect of training on self-perceived everyday effort and fatigue, for example using ecological momentary assessment (EMA; Holube et al., 2020). A significant reduction in one or both of these two metrics would indicate that the training elicited positive changes relevant from the patient's perspective.

5.3.2 The importance of an appropriate control group

A double-blinded design with an active control group is considered necessary to ensure good quality of evidence in studies investigating the efficacy of auditory training (McKay, 2021). Ideally, both the training and the control group should be equally engaged in the study and have similar expectations regarding the potential benefits. Moreover, the performance improvement can only be attributed to the training if the statistically significant effect is revealed using an across group analysis as opposed to comparing the pre- and the post-intervention scores within each group.

In the investigations of this thesis, the control group listened to audiobooks and answered multiple-choice questions about their content. Similar activity was offered to control groups in two previous high-quality studies (Humes et al., 2019; Saunders et al., 2016). For the three outcome measures in the present investigation, the statistical analysis revealed a significant *Visit* effect (post vs. pre) when no significant *Group x Visit* interaction was present. These results included: a decrease of SRT in the HINT (*Chapter 2*), an increase of peak pupil dilations during the HINT (*Chapter 3*) and a decrease of SRT using DANTALE II sentences in the presence of modulated background noise (*Chapter 4*). If no control group had been involved, the obtained significant changes from pre- to post-intervention visit might have been erroneously interpreted as the training-related improvements. These findings demonstrated that a passive control group is not always sufficient to rule out placebo effects in auditory training studies.

5.4 Limitations of the study

The investigations presented in this thesis are not free from certain limitations, which should be avoided in future studies aspiring to what is considered a "golden standard" in auditory training research.

Several aspects of the study design failed to comply with the double-blinded randomized controlled trial criteria. The number of participants was small and not pre-defined based on estimating the size effects and the power calculation. Furthermore, no stratified randomization procedure was chosen that would ensure equal characteristics of participants. Due to those decisions, an unexpected statistically significant difference between the groups in auditory working memory scores was found in the investigation of the phoneme-in-noise training (reported in *Chapter 3*). Therefore, it cannot be confirmed or denied whether this difference introduced any bias in the results. The double-blinding was not achieved since the group assignment was known to the experimenter.

Another potential limitation may stem from the character of the control intervention and the lack of measurement of the participants' expectations. Even though listening to audiobooks has previously been employed in auditory training studies (Humes et al., 2019; Saunders et al., 2016), it might not have created a convincing impression of training. In fact, two participants from the control groups reported to the experimenter that they did not experience the administered intervention as something that could improve their speech perception.

Lastly, some degree of participant selection bias could not be avoided in the present investigations. Individuals willing to commit to being part of audiological research are most likely to exhibit higher than average motivation to comply with the training schedule. It is unknown whether the outcomes obtained with the selected groups of participants are representative of the general population.

5.5 Perspectives

Currently, auditory training plays a marginal role in hearing rehabilitation. It is estimated that less than 10% of audiologists offer auditory training to their adult patients (Li-Korotky, 2012). It seems that the current state of evidence is not sufficient to guarantee that the benefits of auditory training will outweigh the costs. Clinicians might therefore not feel encouraged to incorporate auditory training in their practice.

To ensure the high efficiency of auditory training, a more individualized approach is needed. The next step in the auditory training research could be to identify profiles of listeners who could benefit most from the training. Hearing-aid benefit could be one of the criteria for training candidacy. Low-benefit hearing-aid users might be more likely to experience meaningful changes in real life after training. On the other hand, for individuals with high hearing-aid benefit, training might not provide meaningful additional improvements in terms of speech-in-noise understanding.

Another step to achieve high training efficiency would be to investigate more carefully how long the training benefits persist over time. So far, the retention of training-related changes has been assessed at only one follow-up measurement. As a result, only the presence or absence of the benefit at that specific point in time was evaluated. A longitudinal study design where the follow-up measurements are repeated in regular time intervals after the training period would reveal more precisely when the training effects begin to disappear. This knowledge would inform about how often the training needs to be repeated in order to maintain the benefits over time.

Auditory training should always be seen in the context of a holistic approach to aural rehabilitation. Boothroyd (2007) proposed four aspects of such an approach, namely 1) sensory management, including the provision of hearing devices, 2) perceptual training, 3) instruction in the use of hearing assistive devices and 4) counseling. While providing hearing aids that restore access to auditory information is the first and essential step in hearing loss remediation, the importance of other aspects should not be underestimated. Complementing hearing devices with auditory training might lead to better rehabilitation outcomes. Still, even this combination might not be sufficient to address the current challenges with improving the communication skills in hearing-impaired listeners. Therefore, more attention should be given to counseling, as this element of auditory rehabilitation can address aspects that cannot be tackled by technology or auditory training. In the holistic approach to auditory rehabilitation, hearing care professionals would need to be more involved - not only in providing auditory training to those who might benefit from it - but also in coordinating different areas of action to ensure the best possible interplay between them.

Bibliography

Abavisani, A. and J. B. Allen (2017). “Evaluating hearing aid amplification using idiosyncratic consonant errors”. In: *The journal of the Acoustical Society of America* 142.6, pp. 3736–3745.

Abrams, H. B., K. Bock, and R. L. Irey (2015). “Can a Remotely Delivered Auditory Training Program Improve Speech-in-Noise Understanding?” In: *American Journal of Audiology* 24, pp. 333–337.

Ahissar, M., M. Nahum, I. Nelken, and S. Hochstein (2008). “Reverse hierarchies and sensory learning”. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 364, pp. 285 –299.

Akeroyd, M. A. (2008). “Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults”. In: *International Journal of Audiology* 47.sup2, S53–S71.

Alhanbali, S., P. D. Dawes, S. K. W. Lloyd, and K. J. Munro (2017). “Self-Reported Listening-Related Effort and Fatigue in Hearing-Impaired Adults”. In: *Ear and Hearing* 38, e39–e48.

Alhanbali, S., P. D. Dawes, R. E. Millman, and K. J. Munro (2019). “Measures of Listening Effort Are Multidimensional”. In: *Ear and Hearing* 40, pp. 1084 –1097.

Anderson, S., T. White-Schwoch, A. Parbery-Clark, and N. Kraus (2013). “Reversal of age-related neural timing delays with training”. In: *Proceedings of the National Academy of Sciences* 110, pp. 4357 –4362.

Arlinger, S. and H. Dryselius (1990). “Speech Recognition in Noise, Temporal and Spectral Resolution in Normal and Impaired Hearing”. In: *Acta Oto-Laryngologica* 109.sup469, pp. 30–37.

Bachmann, F. L. and J. Hjortkjær (in preparation). “Test-retest reliability of subcortical responses to running speech in older hard of hearing adults”.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1, pp. 1–48.

Başkent, D. and E. Gaudrain (2016). “Musician advantage for speech-on-speech perception.” In: *The Journal of the Acoustical Society of America* 139 3, EL51–6.

Berens, P. (2009). “CircStat: A MATLAB Toolbox for Circular Statistics”. In: *Journal of Statistical Software* 31.10, 1–21.

Bianchi, F. et al. (2019). “Benefit of Higher Maximum Force Output on Listening Effort in Bone-Anchored Hearing System Users: A Pupillometry Study”. In: *Ear and Hearing* 40.5, pp. 1220–1232.

Blackburn, H. L. and A. L. Benton (1957). “Revised administration and scoring of the digit span test”. In: *journal of consulting psychology* 21 2, pp. 139–43.

Boebinger, D., S. Evans, S. Rosen, C. F. Lima, T. Manly, and S. K. Scott (2015). “Musicians and non-musicians are equally adept at perceiving masked speech.” In: *The Journal of the Acoustical Society of America* 137 1, pp. 378–87.

Boot, W. R., D. J. Simons, C. Stothart, and C. Stutts (2013). “The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects”. In: *Perspectives on Psychological Science* 8.4, pp. 445–454.

Boothroyd, A. (2007). “Adult Aural Rehabilitation: What Is It and Does It Work?” In: *Trends in Amplification* 11.2, pp. 63–71.

Boothroyd, A. (2010). “Adapting to changed hearing: the potential role of formal training”. In: *Journal of the American Academy of Audiology* 21.9, pp. 601–611.

Boothroyd, A. and A. Schauer (Oct. 2015). *Lowest acceptable performance level.* Presented at the Meeting of the International Collegium of Rehabilitative Audiology. Berkeley, CA.

Brand, T. and B. Kollmeier (2002). “Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests”. In: *The journal of the Acoustical Society of America* 111, pp. 2801–10.

Bégel, V., A. Seilles, and S. D. Bella (2018). “Rhythm Workers: A music-based serious game for training rhythm skills”. In: *Music & Science* 1, p. 2059204318794369.

Calandruccio, L. and K. A. Doherty (2008). “Spectral weighting strategies for hearing-impaired listeners measured using a correlational method”. In: *The journal of the Acoustical Society of America* 123.4, pp. 2367–2378.

Carson, N., L. Leach, and K. J. Murphy (2018). “A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores”. In: *International journal of Geriatric Psychiatry* 33, pp. 379 –388.

Cason, N. and D. Schön (2012). “Rhythmic priming enhances the phonological processing of speech”. In: *Neuropsychologia* 50, pp. 2652–2658.

Cherry, E. C. (1953). “Some Experiments on the Recognition of Speech, with One and with Two Ears”. In: *The Journal of the Acoustical Society of America* 25.5, pp. 975–979.

Chisolm, T. H., G. H. Saunders, M. T. Frederick, R. A. McArdle, S. L. Smith, and R. H. Wilson (2013). “Learning to Listen Again: The Role of Compliance in Auditory Training for Adults With Hearing Loss”. In: *American Journal of Audiology* 22.2, pp. 339–342.

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

Clayton, K. K., J. Swaminathan, A. Yazdanbakhsh, J. Zuk, A. D. Patel, and G. Kidd (2016). “Executive Function, Visual Attention and the Cocktail Party Problem in Musicians and Non-Musicians”. In: *PLoS ONE* 11.

Coffey, E. B. J., A. M. P. Chepesiuk, S. C. Herholz, S. Baillet, and R. J. Zatorre (2017a). “Neural Correlates of Early Sound Encoding and their Relationship to Speech-in-Noise Perception”. In: *Frontiers in Neuroscience* 11, p. 479.

Coffey, E. B. J., N. B. Mogilever, and R. J. Zatorre (2017b). “Speech-in-noise perception in musicians: A review”. In: *Hearing Research* 352, pp. 49–69.

Cosh, S. M., C. Helmer, C. Delcourt, T. G. Robins, and P. J. Tully (2019). “Depression in elderly patients with hearing loss: current perspectives”. In: *Clinical Interventions in Aging* 14, pp. 1471 –1480.

Dalla Bella, S. and J. S. Sowinski (2015). “Uncovering beat deafness: detecting rhythm disorders with synchronized finger tapping and perceptual timing tasks.” In: *Journal of Visualized Experiments : JoVE* 97.

Dalla Bella, S. et al. (2017). “BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities”. In: *Behavior Research Methods* 49, pp. 1128–1145.

Dawes, P., R. Hopkins, and K. J. Munro (2013). “Placebo effects in hearing-aid trials are reliable”. In: *International Journal of Audiology* 52.7, pp. 472–477.

Dubinsky, E., E. A. Wood, G. Nespoli, and F. A. Russo (2019). “Short-Term Choir Singing Supports Speech-in-Noise Perception and Neural Pitch Strength in

Older Adults With Age-Related Hearing Loss”. In: *Frontiers in Neuroscience* 13, p. 1153.

Eggert, T., C. Sauter, R. Popp, J. Zeitlhofer, and H. Danker-Hopfe (2012). “The pupillographic sleepiness test in adults: Effect of age, gender, and time of day on pupillometric variables”. In: *American Journal of Human Biology* 24.

Elberling, C., C. Ludvigsen, and P. E. Lyregaard (1989). “DANTALE: a new Danish speech material.” In: *Scandinavian Audiology* 18 3, pp. 169–75.

Ferguson, M., H. Henshaw, D. Clark, and D. Moore (2014). “Benefits of Phoneme Discrimination Training in a Randomized Controlled Trial of 50- to 74-Year-Olds With Mild Hearing Loss”. In: *Ear and hearing* 35, e110–121.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data.* Cambridge University Press.

French, N. R., C. W. Carter, and W. Koenig (1930). “The words and sounds of telephone conversations”. In: *The Bell System Technical Journal* 9.2, pp. 290–324.

Fujii, S. and C. Y. Wan (2014). “The Role of Rhythm in Speech and Language Rehabilitation: The SEP Hypothesis”. In: *Frontiers in Human Neuroscience*, p. 777.

Gatehouse, S. and W. Noble (2004). “The Speech, Spatial and Qualities of Hearing Scale (SSQ)”. In: *International Journal of Audiology* 43, pp. 85–99.

Ghitza, O. and S. Greenberg (2009). “On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence”. In: *Phonetica* 66, pp. 113 –126.

Henshaw, H. and M. A. Ferguson (2013a). “Efficacy of Individual Computer-Based Auditory Training for People with Hearing Loss: A Systematic Review of the Evidence”. In: *PLoS ONE* 8.

Henshaw, H. and M. Ferguson (2013b). “Assessing the benefits of auditory training to real-world listening: identifying appropriate and sensitive outcomes”. In: *Proceedings of the International Symposium on Auditory and Audiological Research* 4, 45–52.

Holube, I., P. v. Gablenz, and J. Bitzer (2020). “Ecological Momentary Assessment (EMA) in hearing research – current state, challenges, and future directions”. In: *Ear and Hearing* 41.S1, 79S–90S.

Howard, C. S., K. J. Munro, and C. J. Plack (2010). “Listening effort at signal-to-noise ratios that are typical of the school classroom”. In: *International Journal of Audiology* 49, pp. 928 –932.

Humes, L. E. (1999). “Dimensions of hearing aid outcome.” In: *Journal of the American Academy of Audiology* 10 1, pp. 26–39.

Humes, L. E., M. H. Burk, L. E. Strauser, and D. L. Kinney (2009). “Development and Efficacy of a Frequent-Word Auditory Training Protocol for Older Adults with Impaired Hearing”. In: *Ear and Hearing* 30, pp. 613–627.

Humes, L. E., K. G. Skinner, D. L. Kinney, S. E. Rogers, A. K. Main, and T. M. Quigley (2019). “Clinical Effectiveness of an At-Home Auditory Training Program: A Randomized Controlled Trial.” In: *Ear & Hearing*.

Iglesia, D. I. (Nov. 2016). “The Mobility is the Message : the Development and Uses of MobMuPlat”. In: *Proceedings of the 5th International Pure Data Convention, PdCon16*. New York, NY.

Iversen, J. R. and A. D. Patel (2008). “The Beat Alignment Test (BAT): Surveying beat processing abilities in the general population”. In: *10th International Conference on Music Perception Cognition*. Ed. by M. A. e. a. Ken’ichi Miyazaki Yuzuru Hiraga. Sapporo, Japan, pp. 465–468.

Jones, M. R. and M. Boltz (1989). “Dynamic attending and responses to time.” In: *Psychological review* 96 3, pp. 459–91.

Jones, M. R., H. Moynihan, N. MacKenzie, and J. Puente (2002). “Temporal Aspects of Stimulus-Driven Attending in Dynamic Arrays”. In: *Psychological Science* 13, pp. 313 –319.

Kahneman, D. (1973). *Attention and Effort*. NJ: Prentice-Hall, Inc.

Koprowska, A., J. Marozeau, T. Dau, and M. Serman (in preparation). “Test-retest reliability of subcortical responses to running speech in older hearing impaired adults”.

Kuchinsky, S. E., J. B. Ahlstrom, S. L. Cute, L. E. Humes, J. R. Dubno, and M. A. Eckert (2014). “Speech-perception training for older adults with hearing loss impacts word recognition and effort”. In: *Psychophysiology* 51 10, 1046–1057.

Kuchinsky, S. E. et al. (2013). “Pupil size varies with word listening and response selection difficulty in older adults with hearing loss.” In: *Psychophysiology* 50 1, pp. 23–34.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen (2017). “lmerTest Package: Tests in Linear Mixed Effects Models”. In: *Journal of Statistical Software* 82.13, 1–26.

Lappe, C., L. J. Trainor, S. C. Herholz, and C. Pantev (2011). “Cortical Plasticity Induced by Short-Term Multimodal Musical Rhythm Training”. In: *PLOS ONE* 6.6, pp. 1–8.

Large, E. W. and M. R. Jones (1999). “The dynamics of attending: How people track time-varying events.” In: *Psychological Review* 106, pp. 119–159.

Lenth, R. V. (2016). “Least-Squares Means: The R Package lsmeans”. In: *journal of Statistical Software* 69.1, pp. 1–33.

Li-Korotky, H.-S. (2012). “Age-related hearing loss: quality of care for quality of life.” In: *The Gerontologist* 52 2, pp. 265–71.

Liberman, M. and A. Prince (1977). “On Stress and Linguistic Rhythm”. In: *Linguistic Inquiry* 8, pp. 249–336.

Lidji, P., C. Palmer, I. Peretz, and M. Morningstar (2011). “Listeners feel the beat: Entrainment to English and French speech rhythms”. In: *Psychonomic Bulletin & Review* 18, pp. 1035 –1041.

London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter, 2nd Edn.* Oxford University Press.

Lumsden, J., E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò (2016). “Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy”. In: *JMIR Serious Games* 4.2, e11.

Madsen, S. M. K., M. Marschall, T. Dau, and A. J. Oxenham (2019). “Speech perception is similar for musicians and non-musicians across a wide range of conditions”. In: *Scientific Reports* 9 1, p. 10404.

Madsen, S. M. K., K. L. Whiteford, and A. J. Oxenham (2017). “Musicians do not benefit from differences in fundamental frequency when listening to speech in competing speech backgrounds”. In: *Scientific Reports* 7.

McAuley, J. D., Y. Shen, S. Dec, and G. R. Kidd (2020). “Altering the rhythm of target and background talkers differentially affects speech understanding.” In: *Attention, perception & psychophysics.*

McCoy, S. L., P. A. Tun, L. C. Cox, M. Colangelo, R. A. Stewart, and A. Wingfield (2005). “Hearing Loss and Perceptual Effort: Downstream Effects on Older Adults’ Memory for Speech”. In: *Quarterly Journal of Experimental Psychology* 58, pp. 22 –33.

McGarrigle, R. et al. (2014). “Listening effort and fatigue: what exactly are we measuring? A British society of audiology cognition in hearing special interest group ’white paper’”. English. In: *International Journal of Audiology* 53.7, pp. 433–440.

McKay, C. M. (2021). “No Evidence That Music Training Benefits Speech Perception in Hearing-Impaired Listeners: A Systematic Review”. In: *Trends in Hearing* 25, p. 2331216520985678.

Micheyl, C., K. Delhommeau, X. Perrot, and A. J. Oxenham (2006). “Influence of musical and psychoacoustical training on pitch discrimination”. In: *Hearing Research* 219, pp. 36–47.

Mick, P., I. Kawachi, and F. R. Lin (2014). “The Association between Hearing Loss and Social Isolation in Older Adults”. In: *Otolaryngology–Head and Neck Surgery* 150.3, pp. 378–384.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. CRC Press.

Mishra, S. K., M. R. Panda, and C. Herbert (2014). “Enhanced auditory temporal gap detection in listeners with musical training”. In: *The Journal of the Acoustical Society of America* 136.2, EL173–EL178.

Moore, B. C. J. and B. R. Glasberg (1998). “Use of a Loudness Model for Hearing-Aid Fitting. I. Linear Hearing Aids”. In: *British Journal of Audiology* 32.5, pp. 317–335.

Moore, D. R., J. F. Rosenberg, and J. S. Coleman (2005). “Discrimination training of phonemic contrasts enhances phonological processing in mainstream school children”. In: *Brain and Language* 94.1, pp. 72–85.

Müllensiefen, D., B. Gingras, J. Musil, and L. Stewart (2014). “The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population”. In: *PLOS ONE* 9, pp. 1–23.

Nasreddine, Z. S. et al. (2005). “The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment”. In: *journal of the American Geriatrics Society* 53.4, pp. 695–699.

Neagu, M.-B., T. Dau, P. Bækgaard, H. Relaño-Iborra, A. A. Kressner, and D. Wendt (in preparation). “Investigating the reliability of pupillometry as a measure of individualized listening effort”.

Nielsen, J. B. and T. Dau (2011). “The Danish hearing in noise test”. In: *International Journal of Audiology* 50.3, pp. 202–208.

Nielsen, J. and T. Dau (2019). “A Danish nonsense word corpus for phoneme recognition measurements”. In: *Acustica United with Acta Acustica* 105.1, pp. 183–194.

Noble, W. and S. Gatehouse (2004). “Interaural asymmetry of hearing loss, Speech, Spatial and Qualities of Hearing Scale (SSQ) disabilities, and handicap”. In: *International journal of Audiology* 43.2, pp. 100–114.

Parbery-Clark, A., E. Skoe, and N. Kraus (2009a). “Musical Experience Limits the Degradative Effects of Background Noise on the Neural Processing of Sound”. In: *The journal of Neuroscience* 29, pp. 14100 –14107.

Parbery-Clark, A., E. Skoe, C. K.-Y. Lam, and N. Kraus (2009b). “Musician Enhancement for Speech-In-Noise”. In: *Ear and Hearing* 30, pp. 653–661.

Phatak, S. A. and J. B. Allen (2007). “Consonant and vowel confusions in speech-weighted noise”. In: *The journal of the Acoustical Society of America* 121.4, pp. 2312–2326.

Pichora-Fuller, M. K. et al. (2016). “Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL)”. In: *Ear and Hearing* 37, 5S–27S.

Picou, E. M. (2020). “MarkeTrak 10 (MT10) Survey Results Demonstrate High Satisfaction with and Benefits from Hearing Aids.” In: *Seminars in hearing* 41 1, pp. 21–36.

Piquado, T., D. M. Isaacowitz, and A. Wingfield (2010). “Pupillometry as a measure of cognitive effort in younger and older adults.” In: *Psychophysiology* 47 3, pp. 560–9.

Puckette, M. (Feb. 1996). “Pure Data: another integrated computer music environment”. In: *Proceedings of the Second Intercollege Computer Music Concerts*, pp. 37–41.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Relaño-Iborra, H. and P. Bækgaard (2020). *PUPILS pipeline: A flexible Matlab toolbox for eyetracking and pupillometry data processing*. arXiv:2011.05118.

Rönnberg, J. et al. (2013). “The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances”. In: *Frontiers in Systems Neuroscience* 7.

Ruggles, D. R., R. L. Freyman, and A. J. Oxenham (2014). “Influence of Musical Training on Understanding Voiced and Whispered Speech in Noise”. In: *PLoS ONE* 9.

Saunders, G. H., S. L. Smith, T. H. Chisolm, M. T. Frederick, R. A. McArdle, and R. H. Wilson (2016). “A Randomized Control Trial: Supplementing Hearing Aid Use with Listening and Communication Enhancement (LACE) Auditory Training”. In: *Ear and Hearing* 37, pp. 381–396.

Scheidiger, C., J. B. Allen, and T. Dau (2017). “Assessing the efficacy of hearing-aid amplification using a phoneme test”. In: *The Journal of the Acoustical Society of America* 141.3, pp. 1739–1748.

Schellenberg, E. G. (2015). “Music training and speech perception: a gene-environment interaction”. In: *Annals of the New York Academy of Sciences* 1337, pp. 170–177.

Schön, D. and B. Tillmann (2015). “Short- and long-term rhythmic interventions: perspectives for language rehabilitation”. In: *Annals of the New York Academy of Sciences* 1337.

Schumann, A., L. G. Garcia, and U. Hoppe (2016). “Trainierbarkeit der Diskrimination von Phonemen im Störgeräusch bei normalhörenden Erwachsenen”. In: *Laryngo-rhino-otologie* 96, pp. 98–103.

Schumann, A., M. Serman, O. Gefeller, and U. Hoppe (2015). “Computer-based auditory phoneme discrimination training improves speech recognition in noise in experienced adult cochlear implant listeners”. In: *International journal of audiology* 54, pp. 1–9.

Schumann, A., M. Serman, and U. Hoppe (2018). *Phoneme discrimination training in background noise can improve speech recognition in noise in elderly Hearing Aid users*. Poster presented at the 4th International Conference on Cognitive Hearing Science for Communication. Linköping, Sweden.

Serman, M. (2012). *SchooLo: new speech training method based on changes in speech cues*. Paper presented at the 15th Annual meeting of the German Society of Audiology. Erlangen, Germany.

Serman, M., K. Kaja, M. Alina, and G. Ulrich (2018). *Finding the sweet spot for phoneme in noise training*. Poster presented at the 10th Workshop on Speech in Noise. Glasgow, UK.

Sidiras, C., V. Iliadou, I. Nimatoudis, T. Reichenbach, and D.-E. Bamiou (2017). “Spoken Word Recognition Enhancement Due to Preceding Synchronized Beats Compared to Unsynchronized or Unrhythmic Beats”. In: *Frontiers in Neuroscience* 11, p. 415.

Slater, J. and N. Kraus (2015). “The role of rhythm in perceiving speech in noise: a comparison of percussionists, vocalists and non-musicians”. In: *Cognitive Processing* 17, pp. 79–87.

Stecker, G. C., G. A. Bowman, E. W. Yund, T. J. Herron, C. M. Roup, and D. L. Woods (2006). “Perceptual training improves syllable identification in new and experienced hearing aid users.” In: *Journal of Rehabilitation Research and Development* 43 4, pp. 537–52.

Stropahl, M., J. Besser, and S. Launer (2019). “Auditory Training Supports Auditory Rehabilitation: A State-of-the-Art Review”. In: *Ear and Hearing* 41, p. 1.

Swaminathan, J., C. R. Mason, T. Streeter, V. Best, J. Gerald Kidd, and A. D. Patel (2015). “Musical training, individual differences and the cocktail party problem”. In: *Scientific Reports* 5.

Sweetow, R. W. and J. H. Sabes (2006). “The need for and development of an adaptive Listening and Communication Enhancement (LACE) Program.” In: *Journal of the American Academy of Audiology* 17 8, pp. 538–58.

Sweetow, R. W. and J. H. Sabes (2010). “Auditory training and challenges associated with participation and compliance.” In: *Journal of the American Academy of Audiology* 21.9, pp. 586–593.

Varnet, L., T. T. Wang, C. Peter, F. Meunier, and M. Hoen (2015). “How musical expertise shapes speech perception: evidence from auditory classification images”. In: *Scientific Reports* 5.

Varnet, L., K. Knoblauch, F. Meunier, and M. Hoen (2013). “Using auditory classification images for the identification of fine acoustic cues used in speech perception”. In: *Frontiers in Human Neuroscience* 7, p. 865.

Varnet, L., C. Langlet, C. Lorenzi, D. S. Lazard, and C. Micheyl (2019). “High-Frequency Sensorineural Hearing Loss Alters Cue-Weighting Strategies for Discriminating Stop Consonants in Noise”. In: *Trends in Hearing* 23, p. 2331216519886707.

WHO (2021). *World report on hearing: executive summary*. Tech. rep. Geneva: World Health Organization.

Wagener, K. C. and T. Brand (2005). “Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters”. In: *International Journal of Audiology* 44.3, pp. 144–156.

Wagener, K., J. L. Josvassen, and R. Ardenkjær (2003). “Design, optimization and evaluation of a Danish sentence test in noise”. In: *International Journal of Audiology* 42.1, pp. 10–17.

Wallentin, M., A. Højlund, M. Friis-Olivarius, C. Vuust, and P. Vuust (2010). “The Musical Ear Test, a new reliable test for measuring musical competence”. In: *Learning and Individual Differences* 20, pp. 188–196.

Wang, Y. et al. (2018). “Relations Between Self-Reported Daily-Life Fatigue, Hearing Status, and Pupil Dilation During a Speech Perception in Noise Task”. In: *Ear and Hearing* 39, pp. 573 –582.

Wayne, R. V. and I. S. Johnsrude (2015). “A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline”. In: *Ageing Research Reviews* 23.Part B, pp. 154–166.

Wechsler, D. (2008). *Wechsler adult intelligence scale.* Tech. rep. San Antonio, TX: The Psychological Corporation.

Wendt, D., R. Hietkamp, and T. Lunner (2017). “Impact of Noise and Noise Reduction on Processing Effort: A Pupillometry Study: A Pupillometry Study”. English. In: *Ear and Hearing* 38.6, pp. 690–700.

Wendt, D., T. Koelewijn, P. Książek, S. E. Kramer, and T. Lunner (2018). “Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test”. In: *Hearing Research* 369, pp. 67–78.

Whitton, J., K. Hancock, J. Shannon, and D. Polley (2017). “Audiomotor Perceptual Training Enhances Speech Intelligibility in Background Noise.” In: *Current Biology* 27, pp. 1–11.

Winn, M. B. (2016). “Rapid Release From Listening Effort Resulting From Semantic Context, and Effects of Spectral Degradation and Cochlear Implants”. In: *Trends in Hearing* 20, 1–17.

Winn, M., D. Wendt, T. Koelewijn, and S. Kuchinsky (2018). “Best Practices and Advice for Using Pupillometry to Measure Listening Effort: An Introduction for Those Who Want to Get Started”. In: *Trends in Hearing* 22, pp. 1–22.

Woods, D. L., T. Arbogast, Z. Doss, M. Younus, T. J. Herron, and E. W. Yund (2015a). “Aided and Unaided Speech Perception by Older Hearing Impaired Listeners”. In: *PLOS ONE* 10.3, pp. 1–29.

Woods, D. L., E. W. Yund, T. J. Herron, and M. A. I. U. Cruadhlaoich (2010). “Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise”. In: *The journal of the Acoustical Society of America* 127.3, pp. 1609–1623.

Woods, D. L. et al. (2015b). “Speech Perception in Older Hearing Impaired Listeners: Benefits of Perceptual Training”. In: *PLOS ONE* 10.3, pp. 1–25.

Yates, K. (2016). “Musical training as a potential tool for improving speech perception in background noise”. PhD thesis. University of Nottingham.

Yates, K., D. R. Moore, S. Amitay, and J. G. Barry (2019). “Sensitivity to Melody, Rhythm, and Beat in Supporting Speech-in-Noise Perception in Young Adults”. In: *Ear and Hearing* 40, pp. 358 –367.

Zendel, B. R., G. L. West, S. Belleville, and I. Peretz (2019). “Musical training improves the ability to understand speech-in-noise in older adults”. In: *Neurobiology of Aging* 81, pp. 102–115.

Zuk, J., C. Benjamin, A. Kenyon, and N. Gaab (2014). “Behavioral and Neural Correlates of Executive Functioning in Musicians and Non-Musicians”. In: *PLOS ONE* 9, pp. 1–14.

Contributions to Hearing Research

Vol. 1: *Gilles Pigasse*, Deriving cochlear delays in humans using otoacoustic emissions and auditory evoked potentials, 2008.
External examiners: Mark Lutman, Stefan Stenfeld

Vol. 2: *Olaf Strelcyk*, Peripheral auditory processing and speech reception in impaired hearing, 2009.
External examiners: Brian Moore, Kathrin Krumbholz

Vol. 3: *Eric R. Thompson*, Characterizing binaural processing of amplitude-modulated sounds, 2009.
External examiners: Michael Akeroyd, Armin Kohlrausch

Vol. 4: *Tobias Piechowiak*, Spectro-temporal analysis of complex sounds in the human auditory system, 2009.
External examiners: Jesko Verhey, Steven van de Par

Vol. 5: *Jens Bo Nielsen*, Assessment of speech intelligibility in background noise and reverberation, 2009.
External examiners: Björn Hagerman, Ejnar Laukli

Vol. 6: *Helen Connor*, Hearing aid amplification at soft input levels, 2010.
External examiners: Inga Holube, Birgitta Larsby

Vol. 7: *Morten Løve Jepsen*, Modeling auditory processing and speech perception in hearing-impaired listeners, 2010.
External examiners: Birger Kollmeier, Ray Meddis

Vol. 8: *Sarah Verhulst*, Characterizing and modeling dynamic processes in the cochlea using otoacoustic emissions, 2010.
External examiners: David Kemp, Stephen Neely

Vol. 9: *Sylvain Favrot*, A loudspeaker-based room auralization system for auditory research, 2010.
External examiners: Bernhard Seeber, Michael Vorländer

Vol. 10: *Sébastien Santurette*, Neural coding and perception of pitch in the normal and impaired human auditory system, 2011.
External examiners: Christopher Plack, Christian Lorenzi

Vol. 11: *Iris Arweiler*, Processing of spatial sounds in the impaired auditory system, 2011.
External examiners: Joost Festen, Jürgen Tchorz

Vol. 12: *Filip Munch Rønne*, Modeling auditory evoked potentials to complex stimuli, 2012.
External examiners: Bob Burkard, Stephen Neely

Vol. 13: *Claus Forup Corlin Jespersgaard*, Listening in adverse conditions: Masking release and effects of hearing loss, 2012.
External examiners: Stuart Rosen, Christian Lorenzi

Vol. 14: *Rémi Decorsière*, Spectrogram inversion and potential applications for hearing research, 2013.
External examiners: Michael Stone, Oded Ghitza

Vol. 15: *Søren Jørgensen*, Modeling speech intelligibility based on the signal-to-noise envelope power ration, 2014.
External examiners: John Culling, Martin Cooke

Vol. 16: *Kasper Eskelund*, Electrophysiological assessment of audiovisual integration in speech perception, 2014.
External examiners: Lawrence Rosenblum, Matthias Gondan

Vol. 17: *Simon Krogholt Christiansen*, The role of temporal coherence in auditory stream segregation, 2014.
External examiners: Shihab Shamma, Guy Brown

Vol. 18: *Márton Marschall*, Capturing and reproducing realistic acoustic scenes for hearing research, 2014.
External examiners: Sascha Spors, Ville Pulkki

Vol. 19: *Jasmina Catic*, Human sound externalization in reverberant environments, 2014.
External examiners: Bernhard Seeber, Steven van de Par

Vol. 20: *Michał Feręczkowski*, Design and evaluation of individualized hearing-aid signal processing and fitting, 2015.
External examiners: Christopher Plack, Enrique Lopez-Poveda

Vol. 21: *Alexandre Chabot-Leclerc*, Computational modeling of speech intelligibility in adverse conditions, 2015.
External examiners: Steven van de Par, John Culling

Vol. 22: *Federica Bianchi*, Pitch representations in the impaired auditory system and implications for music perception, 2016.
External examiners: Ingrid Johnsrude, Christian Lorenzi

Vol. 23: *Johannes Zaar*, Measures and computational models of microscopic speech perception, 2016.
External examiners: Judy Dubno, Martin Cooke

Vol. 24: *Johannes Käsbach*, Characterizing apparent source width perception, 2016.
External examiners: William Whitmer, Jürgen Tchorz

Vol. 25: *Gusztáv Löcsei*, Lateralized speech perception with normal and impaired hearing, 2016.
External examiners: Thomas Brand, Armin Kohlrausch

Vol. 26: *Suyash Narendra Joshi*, Modelling auditory nerve responses to electrical stimulation, 2017.
External examiners: Laurel Carney, Bob Carlyon

Vol. 27: *Henrik Gerd Hassager*, Characterizing perceptual externalization in listeners with normal, impaired and aided-impaired hearing, 2017.
External examiners: Volker Hohmann, Piotr Majdak

Vol. 28: *Richard Ian McWalter*, Analysis of the auditory system via synthesis of natural sounds, speech and music, 2017.
External examiners: Maria Chait, Christian Lorenzi

Vol. 29: *Jens Cubick*, Characterizing the auditory cues for the processing and perception of spatial sounds, 2017.
External examiners: Ville Pulkki, Pavel Zahorik

Vol. 30: *Gerard Encina-Llamas*, Characterizing cochlear hearing impairment using advanced electrophysiological methods, 2017.
External examiners: Roland Schaette, Ian Bruce

Vol. 31: *Christoph Scheidiger*, Assessing speech intelligibility in hearing-impaired listeners, 2018.
External examiners: Enrique Lopez-Poveda, Tim Jürgens

Vol. 32: *Alan Wiinberg*, Perceptual effects of non-linear hearing aid amplification strategies, 2018.
External examiners: Armin Kohlrausch, James Kates

Vol. 33: *Thomas Bentsen*, Computational speech segregation inspired by principles of auditory processing, 2018.
External examiners: Stefan Bleeck, Jürgen Tchorz

Vol. 34: *François Guérit*, Temporal charge interactions in cochlear implant listeners, 2018.
External examiners: Julie Arenberg, Olivier Macherey

Vol. 35: *Andreu Paredes Gallardo*, Behavioral and objective measures of stream segregation in cochlear implant users, 2018.
External examiners: Christophe Micheyl, Monita Chatterjee

Vol. 36: *Søren Fuglsang*, Characterizing neural mechanisms of attention-driven speech processing, 2019.
External examiners: Shihab Shamma, Maarten de Vos

Vol. 37: *Borys Kowalewski*, Assessing the effects of hearing-aid dynamic-range compression on auditory signal processing and perception, 2019.
External examiners: Brian Moore, Graham Naylor

Vol. 38: *Helia Relaño Iborra*, Predicting speech perception of normal-hearing and hearing-impaired listeners, 2019.
External examiners: Ian Bruce, Armin Kohlrausch

Vol. 39: *Axel Ahrens*, Characterizing auditory and audio-visual perception in virtual environments, 2019.
External examiners: Pavel Zahorik, Piotr Majdak

Vol. 40: *Niclas A. Janssen*, Binaural streaming in cochlear implant patients, 2019.
External examiners: Tim Jürgens, Hamish Innes-Brown

Vol. 41: *Wiebke Lamping*, Improving cochlear implant performance through psychophysical measures, 2019.
External examiners: Can Gnasia, David Landsberger

Vol. 42: *Antoine Favre-Félix*, Controlling a hearing aid with electrically assessed eye gaze, 2020.
External examiners: Jürgen Tchorz, Graham Naylor

Vol. 43: *Raul Sanchez Lopez*, Clinical auditory profiling and profile-based hearing-aid fitting, 2020.
External examiners: Judy R. Dubno, Pamela E. Souza

Vol. 44: *Juan Camilo Gil Carvajal*, Modeling audiovisual speech perception, 2020.
External examiners: Salvador Soto-Faraco, Kaisa Maria Tippana

Vol. 45: *Charlotte Amalie Emdal Navntoft*, Improving cochlear implant performance with new pulse shapes: a multidisciplinary approach, 2020.
External examiners: Andrew Kral, Johannes Frijns

Vol. 46: *Naim Mansour*, Assessing hearing device benefit using virtual sound environments, 2021.
External examiners: Virginia Best, Pavel Zahorik

Vol. 47: *Anna Josefine Munch Sørensen*, The effects of noise and hearing loss on conversational dynamics, 2021.
External examiners: William McAllister Whitmer, Martin Cooke

Vol. 48: *Thirsa Huisman*, The influence of vision on spatial localization in normal-hearing and hearing-impaired listeners, 2021.
External examiners: Steven van de Par, Christopher Stecker

Vol. 49: *Florine Lena Bachmann*, Subcortical electrophysiological measures of running speech, 2021.
External examiners: Samira Anderson, Tobias Reichenbach

Vol. 50: *Nicolai Pedersen*, Audiovisual speech analysis with deep learning, 2021.
External examiners: Zheng-Hua Tan, Hani Camille Yehia

The end.

To be continued…

Age-related hearing loss causes substantial deficits in speech perception and can turn everyday conversations into a challenging and exhausting process. Hearing aids often improve speech perception in quiet but difficulty understanding speech in noise remains one of the most common problems reported by hearing-aid users. Therefore, other ways of improving speech-in-noise perception in individuals with hearing impairment are of interest to researchers, clinicians, and the hearing-aid industry. Auditory training is one of the possible strategies to complement hearing aids by enhancing auditory and/or cognitive skills relevant for speech perception in challenging acoustic scenarios. Research on auditory training can help identify the most beneficial auditory training strategies and determine the conditions under which the highest effectiveness of the training can be achieved. The present work investigated the efficacy of two auditory training paradigms in improving speech-in-noise perception in hearing-aid users, namely phoneme-in-noise recognition training and rhythm-based musical training. The results presented in this thesis provide insights into the potential benefits and limitations of these two methods and can be used for further development and refinement of auditory training approaches based on speech and musical stimuli. In the future, auditory training might be incorporated into standard hearing healthcare practice, potentially leading to better outcomes of auditory rehabilitation, especially in individuals with limited hearing-aid benefit.

DTU Health Tech
Department of Health Technology

Ørsteds Plads
Building 352
DK-2800 Kgs. Lyngby
Denmark
Tel: (+45) 45 25 39 50
www.dtu.dk