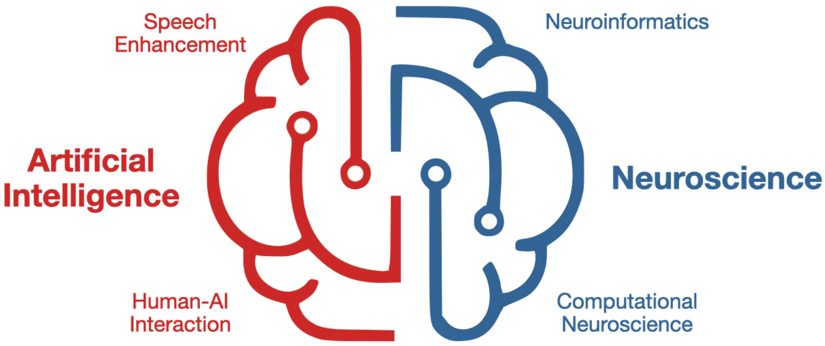
Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

UKE Universitätsklinikum Hamburg-Eppendorf

Antrag zur BMFTR-Bekanntmachung zu Pilotprojekten zum Thema „Neurobiologisch inspirierte Künstliche Intelligenz“



Crossmodal Human-AI Collaboration for Selective Hearing Assistance

Hamburg, den 5.3.2025

A) Project Outline

1. General Information

Acronym: CHAIS

Project Title: Crossmodal Human-AI Collaboration for Selective Hearing Assistance

Keywords: Selective Hearing, Crossmodal Interaction, Assistive Human-Computer Technology

Principle Investigator (PI): Prof. Dr. Frank Steinicke, Human-Computer Interaction, Department of Computer Science, Vogt-Kölln-Str. 30, 22527 Hamburg

Requested Project Budget: 549.288 €; **Funding Period:** 36 months

2. Project Partners

All PIs are members of the University of Hamburg (UHH): two PIs (UHH-AI) are from the Faculty of Mathematics, Informatics, and Natural Sciences (MIN) and two PIs (UKE-NEURO) are from the Faculty of Medicine as well as the Medical Center Hamburg-Eppendorf (UKE).

Name of PI		Institution / Department	Role / task in the project	Budget (€)
UHH-AI	Prof. Dr. Frank Steinicke	Faculty of Mathematics, Informatics, and Natural Sciences, Department of Computer Science	Coordinator, WP 1 and 4 leader; Expertise in Human-AI Interaction	316.964 €
	Prof. Dr.-Ing. Timo Gerkmann	Faculty of Mathematics, Informatics, and Natural Sciences, Department of Computer Science	WP 3 leader; Expertise in speech and multimodal signal processing	
UKE-NEURO	Prof. Dr. Stefano Panzeri	Faculty of Medicine, Institute for Neural Information Processing	WP 2 leader; Expertise in Neuroinformatics and neuronal information processing	232.323 €
	Prof. Dr. Claus C Hilgetag	Faculty of Medicine, Institute of Computational Neuroscience	WP 5 leader; Expertise in computational and network neuroscience	

3. Project Abstract

Recently, artificial intelligence (AI), especially large-language models (LLMs), as well as speech and natural language processing (NLP), have made tremendous improvements [1,2]. In particular, the combination of voice-based technologies and conversational agents has received considerable attention and might become ubiquitous, for example, in smart hearing aids [3,4] or intelligent virtual assistants [5]. Despite their impressive capabilities under perfect conditions, problems become apparent in many real-world situations. In noisy environments, factors such as the presence of multiple speakers, background noise, echoes, or reverberated sounds often lead to failures. In contrast to AI systems, human listeners can take advantage of the coordinational between the hierarchical oscillatory structure and sounds and visual signals to focus attention on a speaker and increase the perceived voice even when multiple sounds compete. Although this so-called *cocktail party phenomenon* has been studied for more than half a century [6,7], understanding its neural

mechanisms has only advanced recently [8,9]. Furthermore, today's AI-based voice technologies can by no means restore the full spectrum of such perceptual effects.

In a cross-disciplinary cooperation between core AI disciplines (i.e., speech and multimodal signal processing and human-AI communication) and neuroscience (i.e., neuroinformatics and computational neuroscience), we will develop neuro-inspired crossmodal AI-based speech signal processing algorithms to address the cocktail party phenomenon. To showcase the unique synergistic potential of this cross-disciplinary cooperation, we will integrate the algorithms into a prototype smart glasses application to support humans during selective hearing tasks.

4. Objectives and Relevance

The advancements of AI systems present significant research opportunities, particularly in the medicine and health domains [10,11]. Here, AI systems, e.g., those based on artificial neural networks (ANN), have achieved impressive successes in various application areas ranging from image recognition [12] to speech and text generation [1,12]. Furthermore, the increasing availability of intelligent assistants built into wearables, such as smartphones or smart glasses, enables novel assistive technologies [13]. However, AI systems are still far from the capabilities of the human brain and are not flexible and robust enough for many applications [14]. Current advancements in AI performance are primarily driven by training with large datasets requiring enormous resources. To further sustainably advance scientific progress through AI, the interface between neuroscience and AI offers a unique synergy potential. The principles of human information processing, the biological foundations, and cognitive processes can provide a rich source of inspiration for novel AI and machine learning (ML) algorithms and architectures. Several examples illustrate the benefits of the human brain compared to AI. However, one classical example particularly highlights the potential for a synergistic approach: The *cocktail party effect* [7], which denotes the crossmodal phenomenon that humans can focus on a specific speaker while filtering out interfering noises. Although this ability manifests to a limited extent through hearing alone, it is greatly enhanced when simultaneous information from the visual sense is available [15]. For instance, watching a speaker's lips can significantly help disambiguate speech in noise, while understanding the context of a sentence enables the listener to anticipate the potential next words of the speaker [6,16].

AI experts have been striving to mimic this human capability with computers for decades, yet it remains challenging [17]. AI systems frequently encounter difficulties in noisy environments, failing to isolate particular voices from surrounding sounds [17]. While hearing aids [3] can amplify sound, they often lack advanced filtering functionalities that enable users to focus on a singular conversation amidst competing auditory stimuli. Indeed, recent advances in AI have set the stage for a transformative shift in multimodal signal processing [17]. However, existing solutions still require large amounts of training data, consume enormous energy, and cannot be integrated into technologies such as hearing aids or intelligent virtual assistants [18]. This is rather concerning given the rising proportion of older adults suffering from age-related or noise-induced hearing loss due to demographic change or exposure to loud environments (through leisure activities, urban living, or work) [4,19]. Given the growing awareness of hearing loss and its impact on the quality of life and less stigma associated with wearing hearing aids, the number of people requiring those assistive technologies will most likely increase [20]. As workplaces strive for inclusivity, intelligent hearing aids can empower individuals with hearing impairments to actively participate in discussions, and intelligent virtual assistants can seamlessly be integrated into their lifestyles. The following user scenario illustrates in Figure 1 how such assistive technologies can empower individuals with hearing difficulties in combating social isolation and enhancing mental well-being:

User scenario: *Bettina, a 58-year-old project assistant with mild hearing loss, attends an informal team meeting. She wears her new smart glasses with integrated neuro-inspired AI-based speech signal processing. As the discussion begins in a moderately noisy room, the smart glasses capture Bettina's gaze and enhance the voice of the speaker she looks at, while background noises are filtered. In addition, the current speaker's name is highlighted, and real-time captions further improve her experience. Bettina can now actively participate in the discussion and feel more engaged in the conversation.*

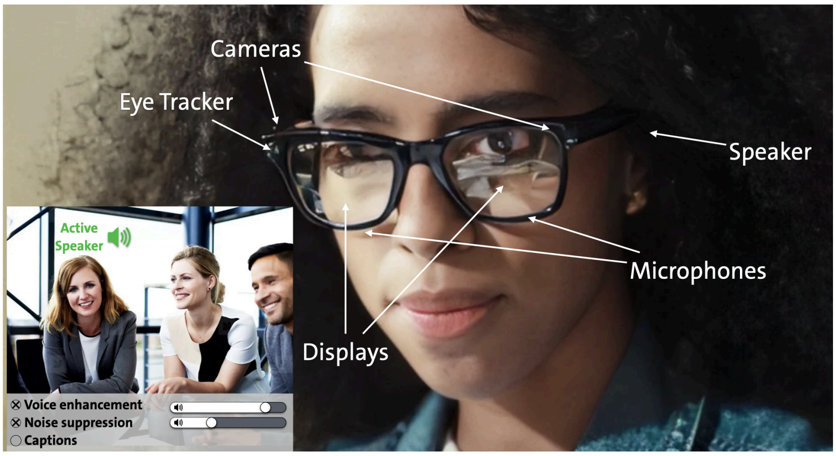


Figure 1: Illustration of the smart glasses prototype. The inset shows the user's view with example settings and a highlighted speaker whose voice is enhanced with the neuro-inspired crossmodal AI algorithms.

Our project aims to take concrete steps toward cross-disciplinary research at the intersection of AI and neuroscience. We will demonstrate for the specific example of the crossmodal cocktail party effect how insights from neuroscience can improve the developments of AI, while at the same time, AI models can be used to verify neuroscience computational models. Finally, we will show how innovations in AI and neuroscience can be transferred to societal impact in a human-centered design process. Therefore, we will address the following research objectives:

O1. Development of neuro-inspired crossmodal AI-based speech signal processing algorithms for the real-time Cocktail Party problem,

O2. Improving and verifying the understanding of the neural mechanisms of selective hearing,

O3. Design of a prototype smart glasses application to assist users in selective hearing tasks in a human-centered design process, and

O4. Enhancing usability, user experience, and performance of the smart glasses application and the neuro-inspired crossmodal AI-based speech signal processing algorithms.

Our pilot project exemplifies the synergistic potential of the collaboration of neuroscience and AI in an important societal challenge, both for the future development of AI and for understanding the fundamental principles of neurobiological systems.

5. Scientific Background and Previous Work

5.1 Human-AI Interaction and Collaboration

The interaction between humans and AI technology has received considerable attention recently [21], and AI chatbots, virtual assistants, and smart home devices assist humans during everyday tasks. Previous work has particularly concentrated on enhancing the transparency and explainability of AI systems [22–24]. In addition, substantial efforts have been made to address fairness and accountability in AI systems [25,26]. Furthermore, there is a growing body of work dedicated to designing AI systems that are responsive and adaptive [27], e.g., to adjust to individual user needs, preferences, goals, and feedback. Here, the integration of AI into smart glasses technology provides enormous promises for more personalized and user-centered experiences [28]. In this context, the group of PI Steinicke has developed intelligent virtual agents/assistants for multimodal human-AI communication in several projects [29,30] and, compared, for example, text- with voice-based [31] communication with AI. Furthermore, using multimodal communication, their AI assistants can express uncertainty in a multimodal way [32,33], they can interact with their smart surroundings [32], and assist humans in health applications [34–36]. While most studies and experiments have been conducted in controlled lab environments, the usage of voice-based interaction with AI in real-life scenarios has highlighted the need for novel AI-based signal processing algorithms to separate the current speaker from background noise; this is one of the major challenges in real-world applications [29,37].

Replicating human auditory perception in AI to solve the Cocktail Party Problem has emerged as a promising field of AI research [16,17]. Researchers have devised methodologies, including Blind Source Separation [21] and Independent Component Analysis [22], to separate a set of source signals from a set of mixed signals or to decompose a multivariate signal into additive subcomponents. Nevertheless, while these techniques exhibit effectiveness under controlled conditions—where sound sources are predictable and exhibit minimal frequency overlap—they encounter substantial difficulties with overlapping voices and sounds, particularly in dynamic and unpredictable contexts [7,15,17,22]. This challenge stems from the absence of sensory and contextual information that humans naturally leverage [23], and, hence, AI systems face significant hurdles in navigating the intricate auditory landscapes of everyday environments.

5.2 Neural Crossmodal Oscillatory Mechanisms of Selective Hearing

Selective hearing in humans allows us to distinguish, interpret, and focus on one sound although other competing sounds are present. The human brain processes sounds binaurally, detecting slight differences in timing and level in inputs from both ears to locate sounds [38]. This ability enables the orientation toward a desired sound source, even when other sounds compete for attention. Furthermore, selective attention [39], often guided by the crossmodal interaction between vocal and facial cues, allows filtering out irrelevant noise and an increased clarity of the perception of the speaker in focus [40]. Moreover, context and memory cues assist in separating speech from background noise [41]. This complex sensory and cognitive processing system is incredibly efficient, but replicating it into AI remains challenging [14]. To combine various visual and auditory stimuli into a coherent percept, the brain must solve the problem of binding stimulus features that belong together while separating them from potentially overlapping signals from a different source. This computational problem is commonly termed auditory scene analysis in audition [42] and image segmentation in vision [43]. Binding linked features from multisensory sources requires the seamless integration of auditory and visual signals into a coherent audio-visual percept [44,45]. Consequently, attentional selection is inherently multimodal, and the neural networks responsible for coordinating attentional control for speech predominantly overlap with those involved in visual processing, indicating the existence of a supramodal system that governs both volitional and reflexive attention allocation [46,47]. Neural processing in the brain is distributed across many specialised regions linked through global nerve fibre networks. These networks support a global neural workspace (GNW) architecture based on a characteristic topological organisation that combines a densely connected association core with more sparsely connected sensory and motor regions at the network periphery [48] (Fig. 2). Such a network organisation supports attentative and conscious signal processing [48] and also provides an ideal substrate for simultaneously achieving robust functional segregation of different sensory modalities represented at the periphery and multimodal integration in the central network core [49]. The GNW organisation is particularly pronounced [50], however, its implications for information processing in ANNs are currently not well understood [48]. Moreover, the functional potential of the GNW architecture in the context of oscillatory brain dynamics has not yet been explored.

Neural oscillations play a fundamental role in speech perception and segmentation, enabling the brain to synchronize auditory and visual inputs for effective communication. Speech is inherently rhythmic, with different linguistic features in different oscillatory bands, and internally generated brain oscillations align with it to process sounds. For example, phonemes and fine-grained acoustic details are expressed in speech and encoded in brain activity by fast gamma (30-80Hz) oscillations; syllables are expressed and encoded at slower brain theta (1-4 Hz) oscillations, which the speech envelope to segment continuous speech into syllabic units [51]. Theta-gamma coupling enhances speech intelligibility, ensuring that incoming speech information is optimally structured for processing and is encoded in a way that is robust to acoustic noise [52]. In particular, theta oscillations modulate cortical gain dynamically, allowing for selective amplification of attended speech while suppressing irrelevant background noise [53] and making it easier to perceive interesting sounds against noise when internal brain oscillations are well aligned to the interesting signal [54]. This gain modulation mechanism provides a flexible adaptation to changing auditory environments, making it crucial for solving real-world challenges in speech perception. Moreover, the Cocktail Party Problem has been shown to involve the synchronization of theta oscillations across sensory modalities [9]. When visual speech cues, such as lip movements, align with auditory speech, theta oscillations in the auditory cortex entrain to the rhythm of visual input, improving speech tracking and intelligibility [55]. This crossmodal synchronization enhances selective attention, reducing the cognitive load required to separate overlapping speech streams. These insights call for the development of a framework for developing AI-driven selective hearing assistance that closely replicates the brain’s natural ability to process in complex environments.

5.3 Crossmodal AI-based Speech Signal Processing

Multimodal AI-based systems for speaker separation have made remarkable progress. Especially in acoustically challenging environments, integrating audio signals and lip movements can provide important additional cues, allowing for better source separation [56]. In unimodal acoustic speech enhancement, significant successes have been achieved with predictive and generative methods [30,56]. In the areas of deep multi-microphone predictive approaches [57] and generative diffusion models [58,59], considerable success has been achieved. Recent developments in AI, particularly within the domain of deep neural networks, have markedly improved machines' proficiency in addressing the Cocktail Party Problem [60], e.g., BioCPPNet [60,61]. Similar to the binaural processing with to ears, spatial processing by employing the spatial diversity of a sound field using multiple microphones is a strong cue to solving the cocktail party problem also for AI-based approaches. Deep spatial filtering amplifies these capabilities by utilizing multiple microphones. While the first approaches used deep learning to estimate the parameters of linear spatial beamformers [62], it has been shown that replacing the linear beamformer with a deep nonlinear joint spatial-spectral filter overcomes, the fundamental limits of linear beamformer in non-Gaussian noise and when fewer microphones than sources are available [57].

The crossmodal integration of visual cues, such as lip movements, is another way to enhance AI's ability to accurately isolate and amplify specific vocalizations. Ephrat et al. [63] developed models capable of isolating speech in a noisy environment. They implemented a deep neural network system utilizing ensembles, convolutional neural networks, and bidirectional LSTM models trained on 2.000 hours of video footage with an individual speaking in front of the camera without any background noise. An elegant way of combining multiple modalities has been proposed in [64], which uses a multi-stream architecture that processes audio, visual, and text inputs separately before combining them. By using cross-attention, the approach can also cope with delays in the signals of the different modalities, which is of high practical relevance. An approach to combine generative audio diffusion models with visual information has been proposed in [59].

In AI speech and multimodal signal processing, often the architecture design is signal-driven and designed to exploit correlations between tempo-spectral features or modalities; often taken from other like computer vision [65]. As described in Sec. 5.2, human perception is intrinsically oscillatory, and, thus, certain signal features are easier to perceive than others is not taken into explicitly account. Also the network topology is often still a drastic simplification as compared to the topology in the human brain. The hypothesis is that existing AI architectures may be either inefficient as they may optimize signal features that are less relevant for human perception. Also we hypothesize that taking inspiration from the core-periphery organisation of the GNW, more efficient ANNs can be developed that enable unimodal and multimodal speaker separation methods that maximise the processing of perceptually-relevant signal features.

6. Impact, Innovation and Exploitation Strategy

The expected results of our project have enormous potential to gain knowledge in neuro-inspired crossmodal AI-based speech signal processing algorithms. Understanding the neural oscillatory mechanisms underlying selective attention to speech provides a strong foundation for designing neuro-inspired AI systems for speech enhancement. By incorporating oscillatory phase synchronization and multimodal integration, future assistive technologies could (i) improve speech tracking in noisy environments by dynamically aligning theta oscillations to speech envelopes, (ii) enhance audio-visual speech perception by leveraging theta synchronization between auditory and visual modalities, and (iii) optimize computational efficiency through biologically-inspired gain control mechanisms. To this end, the results will reveal and showcase a concrete example of how AI can benefit from findings in neuroscience research and vice versa.

Besides the scientific outcomes of the project, the involvement of the UKE and Hospital zum Heiligen Geist (HzHG) (see LoIs) allows the inclusion of end-users and, thereby, enables a human-centered design approach to develop and evaluate assistive smart glasses applications. Hence, the project can have a significantly impact on societal challenges. Notably, the expected advancements of our project addressing the Cocktail Party Problem will catalyze impactful developments within audio technology in general while we focus here on the development of (i) intelligent hearing aids built inside smart glasses to automatically adjust sound settings in real-time to enhance hearing based on the user's gaze, and (ii) intelligent virtual assistants, such as Amazon's Alexa and Apple's Siri, with improved operational efficiency in noisy settings. However, the neuro-inspired crossmodal AI-based speech signal processing algorithms can further stimulate future real-world applications in multiple areas, such as

- creating target speech hearing for noise-canceling headphones to retain the audibility of a specific individual’s voice while eliminating extraneous sounds,
- enhancing call clarity in videoconferencing by filtering background noise and emphasizing the speaker's voice, or
- advancing audio post-production processes by isolating individual sound sources within recorded materials, leading to cleaner audio tracks and more efficient editing workflows.

Therefore, the project has enormous potential to stimulate future collaborative projects, including small and medium enterprises in sound technology (see LoI). Furthermore, the consortium is currently drafting a DFG proposal for a Research Training Group (RTG) to educate future doctoral students at the intersection of AI, neuroscience, and psychology.[1]

7. Communication And Dissemination

The topic of this project is of major interest for both academic and non-academic communities since it addresses cross-disciplinary research at the intersection of neuroscience and AI, as well as health and well-being (UN Sustainable Development Goal). Hence, our communication and dissemination plan has a two-way strategy:

1. We will present our research in top-tier journals and conferences in which the PIs frequently publish. Advancements in the AI-based signal processing algorithms will be submitted, e.g., to Interspeech, ICASSP, the IEEE Transactions on Audio, Speech and Language Processing, as well as general ML venues such as NeurIPS and ICLR. Findings in the neural mechanisms of the audio-visual interactions will be submitted journals such as Nature, Nature Neurosci, Current Biology, Science Adv, PNAS, and PLoS Biology. Finally, we plan to publish results related to assistive technology, e.g., ACM CHI, INTERACT, or HHAI. Furthermore, we plan to organize cross-disciplinary workshops to bring together experts from neuroscience and AI.
2. To exchange with the general public and other disciplines outside of neuroscience and AI, we plan to cooperate with the Academy of Science and Humanities in Hamburg (AdWHH). In 2024, the AdWHH founded a working group "Artificial Cognitive Systems", co-led by PI Steinicke. We will organize workshops, an Academy lecture series on natural and artificial cognitive systems, and AI Promptathons targeted to a broader audience. Those events will be co-organized and financed by the AdWHH.

To constantly inform about our project and the progress, we will set up a website and post intermediate results and news via several social media platforms (e.g., Youtube, BlueSky, or LinkedIn) and in traditional media, i.e., television (e.g., ARD, ZDF) and radio (e.g., NDR).

8. Cooperation and Networking

The planned project builds on a long-term cooperation at UHH, which started with a DFG graduate school (CINACS) followed by the international collaborative research center DFG-TRR 169 on „Crossmodal Learning", in which computer science, neuroscience, and psychology have cooperated. PI Steinicke, Gerkmann, and Hilgetag have been involved in the TRR 169. PI Panzeri is an affiliate member of the Department of Computer Science. Hence, the team has extensive experience in cross-disciplinary research collaboration and subversion of PhD students. We have established effective communication and shared terminology among collaborators from AI and neuroscience. This is of particular importance so that the doctoral students can develop a common language to bridge any potential jargon barriers, complemented by interdisciplinary training

[1] The CHAIS project could be associated with the RTG to provide optimal networking and exchange possibilities.

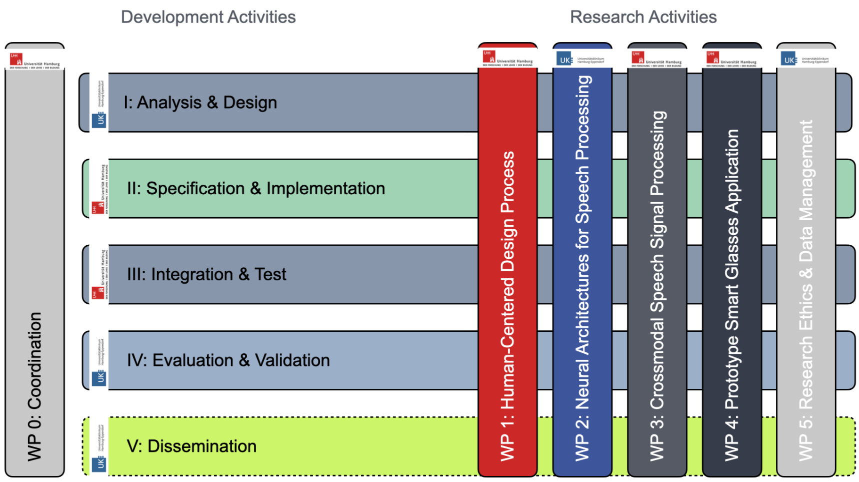
sessions that facilitate mutual understanding of each discipline's concepts and methodologies. Two doctoral students will be supervised by a cross-disciplinary pair of PIs (PI Steinicke/Hilgetag and PI Gerkmann/Panzeri). The supervision team will have bi-weekly meetings, and the entire consortium will meet in monthly meetings to ensure close collaborations. Furthermore, we will provide offices and lab rotation models to allow doctoral students to constantly exchange with experts in AI as well as neuroscience. In addition, doctoral students will be enrolled in the Hamburg Brain School at the Hamburg Center of Neuroscience and will have have access to training and doctoral programs provided by European Laboratory for Learning and Intelligent Systems (ELLIS).[2]

Finally, our project perfectly fits into UHH's long-term research strategy "Cognitive Systems and Neurosciences", which is one of six core research areas. The UHH supports various activities in this field such as the DFG-RTG initiative entitled Human-AI Synergy in which all PIs are involved.

9. Research Plan

9.1. Methodological Approach

To structure the research and development activities of the overall project both chronologically and thematically into various work packages (WPs), a two-dimensional approach was chosen:



The integrative human-centered design process (WP 1) together with an ethical perspective and data management (WP 5) provide the framework for the research activities. The research tasks (T) in the WPs will progress through the development phases from I. Analysis & Design to V. Dissemination. Prototypes will be iteratively developed and evaluated together with potential end-users, who will be recruited via UKE and HzHG (see LoI).

To address objects O1 and O2, we will first individuate the computational mechanisms found in brain experiments as being key to solving the cocktail party problem and to robustly perceive speech. Next, we will implement and study them in biophysically credible models of neural networks to better understand their origin, fundamental features, and computational function. We then can streamline these algorithms into versions that can be implemented in fast online processing. Figure 2 illustrates our approaches, which are described in detail in WP 2 and WP 3.

[2] PI Panzeri is a member of ELLIS.

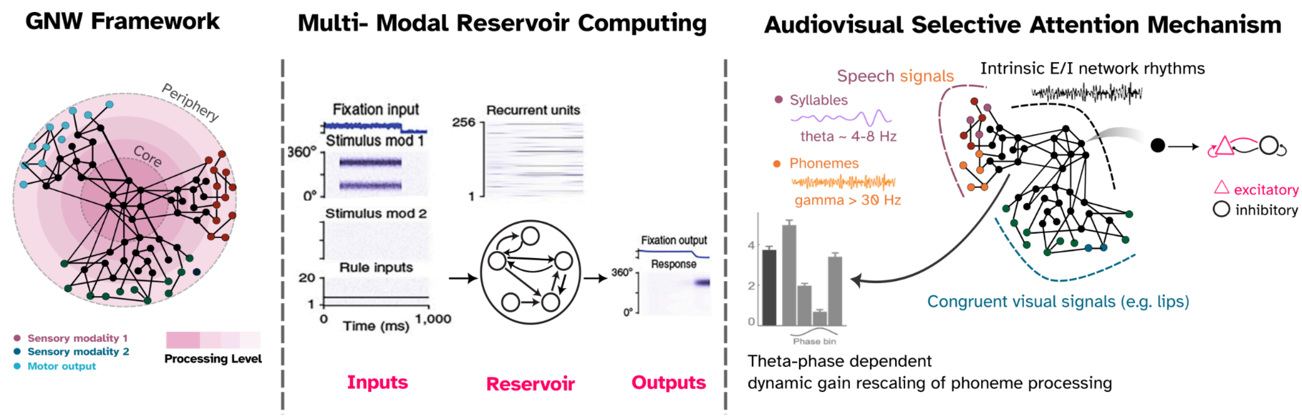


Figure 2: (left) Core-periphery network architecture as connectomic backbone of the GNW, where the core aligns with the "neuronal workspace" while the periphery modules encompass the sensory and motor territories. (Middle) Reservoir computing (RC) architecture and examplar inputs, activity traces in the recurrent layer, and outputs [66], in which the core-periphery network architecture is integrated into the reservoir layer. (Right) Scheme of an excitatory and inhibitory (E/I) neurons network [67] that uses audio-visual selective attention by synchronization of congruent theta rhythms [68] and theta-phase dependent dynamic gain rescaling of phoneme processing [52] modeling human audio-visual perception. Visual input (in blue) focuses attention by individuating theta sound frequencies for dynamic processing.

During the development phase, we will monitor the achieved quality using informal experiments and perceptually motivated instrumental metrics for quality and intelligibility as well as algorithmic latency and compute time. We aim for an algorithmic latency below 10ms, while the compute time should allow for real-time processing in consumer graphical processing units. For the instrumental metrics we aim at reaching competitive numbers on the voicebank demand benchmark[3] and aim at improving PESQ and POLQA by 1 point on the MOS scale, while reaching ESTOI intelligibility prediction above 0.85. By testing which features of these algorithms benefit speech perception, we improve our understanding which key brain algorithmic features for sound perception are, and we will feed them back into an improved understanding of the neural mechanisms of selective hearing.

To address objects O3 and O4, we will integrate the findings into assistive technology in an iterative human-centered design approach. We will begin by analyzing the users' requirements through interviews and focus groups. This understanding informs the ideation phase, where creative solutions for the smart glasses applications are prototyped, allowing users to interact with them and provide feedback. The process of testing and refining ensures that the final design of the smart glasses application effectively addresses users' needs and enhances their overall experience. With this process, we aim for high usability (SUS[4] > 80), user experience (UEQ[5]>1.5 in all dimensions), and acceptability (TAM[6]>80%).

[3] https://paperswithcode.com/dataset/voice-bank-demand
[4] System Usability Scale (SUS) [70]
[5] User Experience Questionaire [71]
[6] Technology Acceptance Model [72]

9.2. Work Programme

A detailed Gantt chart with timeline, milestones, and personal months (PMs) per task, can be found in the appendix. Here, we summarize the individual WPs, their tasks, periods, and deliverables:

WP 0	**Project Coordination**	**PM (total): 4**	**Period: M01 - M36**
Responsible: PI Steinicke; **Other participants:** all			
Description: Project coordination includes budget planning, administrative tasks, project documentation, reports, and public relations. PI Steinicke organizes monthly consortium meetings as well as bi-weekly WP meetings. The R&D progress of the WPs is accompanied by appropriate quality control and documented publicly on the project website. ● T0.1 Administration and Communication: Planning a kick-off workshop and subsequent project meetings for synchronization and coordination of tasks; progress of the WPs will be monitored, and any necessary adjustments will be managed; establishing an efficient communication infrastructure. ● T0.2 Documentation and Dissemination: Documentation through reports; presentation of project results (e.g., TV, radio, newspaper) and maintenance of websites and social media channels.			
Deliverables: D0.1 Dissemination of (interim) results; D0.2 Project website; D0.3 Communication and development platforms; D0.4 Project reports			

WP 1	**Human-Centered Design Process**	**PM (total): 9**	**Period: M01 - M36**
Responsible: PI Steinicke; **Other participants:** all			
Description: In this WP, the smart glasses applications with the integration of the neuro-inspired crossmodal AI-based signal processing algorithms will be implemented following a human-centered design process according to ISO EN 9214-210. Requirement analyses will capture the needs of end users through qualitative and quantitative methods, which will inform WP 4. We plan to recruit potential end users through our network with patient and potential customer access via UKE and HzHG (see LoI). ● T1.1 Requirement Analysis: We will iteratively collect functional and non-functional requirements from potential end users through focus groups, surveys, and interviews. ● T1.2 Documentation and Verification of Requirements: Structured documentation with scenarios and use cases from WP 1.1 will be created. ● T1.3 Formative Evaluations: To evaluate the algorithms (WP 4) we will conduct user studies to measure the perceived quality and user experience; all evaluations are conducted iteratively while results are fed back into the developments in WP 4. ● T1.4 Summative Evaluations: The final evaluation of the smart glasses application will be performed.			
Deliverables: D1.1 Requirement analysis documents; D1.2 Scientific papers about evaluations			

WP 2	**Neural Architectures for Speech Processing**	**PM (total): 18**	**Period: M03 - M32**
Responsible: PI Panzeri; **Other participants:** PI Hilgetag, PI Gerkmann			

Description: The goal of this WP is to develop a crossmodal speech signal processing neural network (see Fig. 2), which improves understanding of the neural mechanisms of selective hearing.

- T2.1 Development of Oscillation-based Model: We will develop an oscillation-based [69,70] network of E/I neurons with separate but interacting visual and auditory "cortical" modules that take as inputs neurally-processed sound and visual spectral features, and generates frequency in the gamma and theta range processing phonemes and syllables, respectively. Synchronized theta-range envelope modulations between theta signals across modules will focus auditory attention using visual signals of interest [9]. The organization of fast gamma dynamics within slower theta cycles will implement predictive coding and help stabilize phoneme and fine acoustic signals against background noise [52]. It will also implement the dynamic gain rescaling as a function of phase of slow cycles that helps syllable segmentation. We will study how credible neural computations affect audio-visual speech perception and which signal features are better perceived, comparing these models and by humans.
- T2.2 Effect of Large Scale Network Structures for Audio-Visual Integration: We will investigate network topologies inspired by the core-periphery organisation of the GNW with respect to their potential of audio-visual segregation and integration. We will systematically generate network architectures and vary key features such as link (a-)symmetry, network density, and topology in terms of link degree distribution and (hierarchical) network modularity. Performance of these networks in audio-visual integration tasks will be evaluated in the RC framework, which allows for a straightforward functional assessment without training-related weight changes. Initially standard rate or excitable dynamics will be used for reservoir nodes. Promising candidate networks will be iteratively optimised.
- T2.3 Interaction of Network Topology and Oscillatory Dynamics: We will use the oscillatory networks in conjunction with variants of the core-periphery GMW architecture to investigate uni-modal and cross-modal functional coupling as well as central integration in the networks. Standard measures such as coherence or Phase-Locking Value (PLV) will be used to assess phase and amplitude coupling in different frequency bands, as the basis of auditory signal processing.
- T2.4 Testing and Fine Tuning: Candidate network architectures will be applied to real-world signals and iteratively optimized with respect to detection accuracy, processing speed and robustness.

Deliverables: D3.1 Realistic oscillatory models; D3.2 Efficient brain-inspired large scale audio-visual integration architecture; D3.3 Understanding of computations afforded by combining temporal and spatial structural organization; D3.4 Fine tuned bio-network architectures

WP 3	Crossmodal Speech Signal Processing	PM (total): 16	Period: M06 - M32

Responsible: PI Gerkmann; **Other participants:** PI Panzeri, PI Hilgetag

Description: Here, we plan to develop a multimodal ML model with brain-inspired network architectures that are memory and computationally efficient. In particular, we will take into account the brain-inspired neural network structures (cf. T2.1) and the oscillatory time structure and synchronization (T2.2).

- T3.1 Development of Brain-inspired Neural Network: We will translate findings from WP2 to an ANN model for speech and audio-visual data. For this, the target signal cost functions should emphasize signal features weighted according to their relevance for human perception or neurally plausible computations. We will use the oscillation-based model (T2.1) to extract relevant features, which are subsequently used as input for neural network based training.

- T3.2 Brain-inspired Topology: We will modify the employed network architecture of the ANN to take into account a brain-inspired topology (cf. T2.2 and 2.3). We will change the network topology of existing audio and audio-visual ANNs to consider the neuroscientifically motivated key features, e.g., the link symmetry or asymmetry, network density, and network modularity. In audio-only and audio-visual speech extraction tasks, we will change the network topology step-by-step in collaboration with T2.2 and T2.3, and feedback the impact of the changes to speech quality, computational demands and processing latency to neuroscience.
- T3.3 Distillation of Neuroscientific Principles: Distillation of audio-visual processing into simpler and differentiable ANN that can implement simulated computations more lightly than the bio-physical models. At the same time, differentiable models allow us to easily integrate the newly developed model of neuroscientific principles into a common ANN to address the cocktail party effect.
- T3.4 Comparison: We will compare the performance of prototypes with and without the developed neurally plausible computations regarding speech quality and computational demands to feed back results about how strongly these putative computations actually impact perception.

Deliverables: D3.1 Brain-inspired neural network model; D3.2 Distillation of the developed model to a differentiable and efficient ANNs; D3.4 Comparative results for the developed brain-inspired ANN

WP 4	**Prototype Smart Glasses Application**	**PM (total): 9**	**Period: M12 - M36**

Responsible: PI Steinicke; **Other participants:** all

Description: We will iteratively prototype the smart glasses application in this WP, which will be provided as an early demonstrator (mid-project) and final demonstrator (end of project). Therefore, we will exploit commercially available smart glasses (e.g., Apple Vision Pro or Meta Ray-Ban glasses). The smart glasses have to be equipped with camera(s), eye tracker(s), speakers, and microphones (cf. Figure 1).

- T4.1 Companion App: To install and change parameters for the smart glasses application, we will provide a simple companion app, which allows simulating the smart glasses application on a mobile phone or desktop computer and finally eases the debugging process.
- T4.2 Gaze Tracking: The target speaker will be identidied through the utilization of multimodal features including the user's gaze and video streams captured by integrated cameras. For the aggregation of gaze events [71], we will build upon our previous research, specifically by predicting the time-to-event of potential shifts in user focus [72,73]. To further augment robustness, we also plan to investigate if window-slicing techniques can effectively identify transitions in focus. Following the local processing of audio and video streams, all data will be promptly discarded to ensure user privacy.
- T4.3 User Interface Settings: We will develop widgets for changing parameters for the signal processing algorithms and the smart glasses application, e.g., to change the volume of the speaker or background, turning captions on or off etc.; those can be changed on the glasses, e.g., via gaze-and-pinch, touchpad at the glasses, voice commands, or via the companion app (T4.1).

● T4.4 Multimodal Feedback: In addition to the current settings, we will provide multimodal, in particular visual feedback about the current speaker in the smart glasses as well as the companion app; the preferred feedback will be determined in the formative evaluation in T1.3.
Deliverables: D4.1 Protoypes of Smart Glasses Apps; D4.2 Early demonstrator; D4.3 Final demonstrator

WP 5	**Research Ethics & Data Management**	**PM (total): 7**	**Period: M03 - M36**
Responsible: PI Hilgetag; **Other participants:** all			
Description: The ethical acceptability of data collection (in particular, users' gaze and people in surroundings with the smart glasses) and the sustainable implementation of the data management system will be ensured according to ISO/IEC 27001 UKE and UHH have established ethics committees as well as data management routines. ● T5.1 Informed Consents: The legal and ethical basis for data collection is given by the consent declaration of participants of the user evaluations (T1.3 and 1.4). All data collection will follow the GDPR (see Sec. 10). ● T5.2 Ethics Applications: We will apply for ethics approval for the planned evaluations (see WP1); depending on the type of evaluation, either the local ethics committees from the Department of Computer Science or the UKE will be involved. ● T5.3 Data Management: Finalization of the data management plan (DMP) (see Appendix 1). Collected data from user studies will initially be pseudonymized. After initial use (publication), data will be stored and made available in anonymized form in the research data repository (RDR). ● T5.4 Open Source Software: We will make the smart glasses application and the neuro-inspired crossmodal AI-based signal processing algorithms publicly available via open source in the RDR using the GNU GPL v. 3.0 licence. ● T5.5 ELSA Workshops: We plan to conduct workshops with experts (e.g., data protection, security, acceptance, cybersickness, EU General Data Protection Regulation, etc.) to reflect on normative design, taking into account medical-ethical and legal regulations.			
Deliverables: D5.1. Informed consents; D5.2 Ethic applications; D5.3 Final DMP; D5.4 Open access to RDR; D5.5 Reports about ELSA workshops			

10. Data Management Plan (DMP)

UKE is member of the medical computer science initiative[7]. We have initiated a DMP based on the Horizon 2020 FAIR DMP template (see Appendix). To ensure that our planned data gathering, storage, and sharing, as well as open source provision meet legal requirements, we are in close contact with UHH's Chief Digital Officer, Data Protection Officer, Open Access Officer, and the Center for Sustainable Research Data Management (RDM Center).[8] The RDM Center provides long-term storage for research and meta-data, as well as sustainable quality services to RDM in

[7] https://www.medizininformatik-initiative.de/
[8] https://www.fdm.uni-hamburg.de

strict compliance with UHH's open access policy, the Hamburg Open Science strategy[9] and according to the FAIR as well as CARE principles to address both the technical and ethical dimensions of data management. We have established routines for data handling to ensure confidentiality, secure collection and storage, and, eventually, secure deletion of all personal and behavioral data following the guidelines for ``Good Scientific Practice'' of the DFG and UHH. Our data gathering and storage procedure will be approved by the Local Ethics Committee of the Department of Computer Science and UKE. All parameters that we collect and our planned analyses will be pre-registered. When participating in user studies, the consent form will inform users about which data will be gathered and for which purposes. Given the behavioral recordings and comments collected when end-users test the smart glasses application, the collected data will be treated confidentially by a pseudonymization procedure using coding lists. This respects certain measures, e.g., Art. 17 GDPR – Right to erasure. We plan to publish only extracted parameters at the UHH's research data repository (RDR) of the RDM Center instead of the raw data to ensure that no personal data will be available. Furthermore, the developed software (cf. WP 3 and 4) will be published in the RDR and our GitHub repositories[10]; publications will be provided with preferably open access in the Research Information System (RIS).

11. Financial Plan

INSTITUTION	MIN Faculty	Medical Faculty	SUM
Personnel	252.137 €	181.603 €	433.740 €
Travel	10.000 €	10.000 €	20.000 €
Consumables	2.000 €	2.000 €	4.000 €
Equipment	0 €	0 €	0 €
Subcontracts	0 €	0 €	0 €
Other	0 €	0 €	0 €
Projektpauschale (PP)	52.827 €	38.721 €	91.548 €
Funding rate (%)	100 %	100 %	100 %
TOTAL BUDGET (€)	323,564€	232.323 €	549.288 €

All other required ressources (such as access to GPU clusters, high-end acoustic lab, XR lab etc.) will be provided by UHH and UKE resources and are already available in the labs of the PIs.

Other Funding

A request for funding this project has not been submitted to any other addressee. In case we submit such a request, we will inform the Federal Ministry of Education and Research immediately.

[9] https://www.hamburg.de/bwfgb/openscience/
[10] https://github.com/orgs/uhhhci/repositories

B) ANNEX

1. CV/Biosketch of Each Partner

1.1 Prof. Dr. Claus Hilgetag

Claus C. Hilgetag is a professor and director of the Institute of Computational Neuroscience at the UKE. A pioneer in network neuroscience, his work combines computational modeling and empirical data analysis to study brain network organization and function in health and disease. Hilgetag is a member of the Academy of Sciences in Hamburg.

Academic qualifications and professional experiences

Study	Biophysics (Diplom), 1988-1991, 1992-1994: Humboldt University Berlin, Germany; Neuroscience: 1991-1992, Edinburgh University, UK Graduate studies in Neuroscience, 1994: Oxford University, UK; 1994-1999: Newcastle University, UK
Doctoral degree	1999: University of Newcastle upon Tyne, UK
Professional experiences	Since 2011: Full Professor (W3) Computational Neuroscience, University Medical Center Eppendorf (UKE), Hamburg University 2006-2011: Associate Professor, Jacobs University Bremen 2001-2006: Assistant Professor, International University Bremen 1999-2001: Postdoctoral Fellow, Boston University School of Medicine

Publications

1. Hilgetag C.C., Goulas A, Changeux J.P. (2022). A natural cortical axis connecting the outside and inside of the human brain. Netw Neurosci, 6(4):950-959. doi: 10.1162/netn_a_00256.
2. Damicelli F., Hilgetag C.C., Goulas A. (2022). Brain connectivity meets reservoir computing. PLoS Comput Biol., 16;18(11):e1010639. doi: 10.1371/journal.pcbi.1010639.
3. Goulas A., Damicelli F., Hilgetag C.C. (2021). Bio-instantiated recurrent neural networks: Integrating neurobiology-based network topology in artificial networks. Neural Netw.,142:608-618. doi: 10.1016/j.neunet.2021.07.011.
4. Changeux J.P., Goulas A., Hilgetag C.C.. (2021). A Connectomic Hypothesis for the Hominization of the Brain, Cereb Cortex, 31(5):2425-2449. doi: 10.1093/cercor/bhaa365.
5. Goulas A., Changeux J.P., Wagstyl K., Amunts K., Palomero-Gallagher N., Hilgetag C.C. (2021) The natural axis of transmitter receptor distribution in the human cerebral cortex. PNAS, 118(3):e2020574118. doi: 10.1073/pnas.2020574118.

1.2 Prof. Dr.-Ing. Timo Gerkmann

Timo Gerkmann is a professor of Signal Processing at the Department of Informatics at the Universität Hamburg. He combines signal processing and machine learning with domain knowledge in speech and acoustics to derive novel, robust and real-time capable algorithms for speech processing in acoustically challenging environments. He received the VDE ITG award 2022

Academic qualifications and professional experiences

Study	1999-2001 Vordiplom, Elektro- und Informationstechnik, Universität Bremen 2001-2004 Diplom, Elektro- / Informationstechnik, Ruhr-Universität Bochum
Doctoral degree	2005-2010 Dr.-Ing., Elektro- und Informationstechnik, Ruhr-Universität Bochum
Habilitation	2014 Positive Zwischenevaluation Juniorprofessur
Professional experiences	since 2016 (W2) Professor of Signal Processing, Department of Computer Science, University of Hamburg, Germany 2015-2016 Principal Scientist Audio & Acoustics, Technicolor R/I, Hannover 2011-2015 Juniorprofessor Speech Signal Processing, Dep. Medical Physics and Acoustics, Universität Oldenburg 2010-2011 Postdoctoral Researcher, Sound & Image Proc. Lab, KTH Royal Institute of Technology, Stockholm, Sweden 2005-2010 PhD Student, Inst. Com. Acoust., Ruhr-Universität Bochum

Publications

1. Simon Welker, Matthew Le, Ricky T. Q. Chen, Wei-Ning Hsu, Timo Gerkmann, Alexander Richard, Yi-Chiao Wu, "FlowDec: A flow-based full-band general audio codec with high perceptual quality," *Int. Conf. Learning Representations (ICLR)*, Singapore, Apr. 2025.
2. Jean-Marie Lemercier, Julius Richter, Simon Welker, Eloi Moliner, Vesa Välimäki, Timo Gerkmann, "Diffusion Models for Audio Restoration", IEEE Signal Processing Magazine, Nov 2024, Vol. 41, No. 6, pp. 72-84, Nov. 2024.
3. Kristina Tesch, Timo Gerkmann, "Multi-channel Speech Separation Using Spatially Selective Deep Non-linear Filters", *IEEE/ACM Trans. Audio, Speech, Language Proc.*, Vol 32, pp. 542-553, 2024.
4. Julius Richter, Simon Welker, Jean-Marie Lemercier, Bunlong Lay, Timo Gerkmann, "Speech Enhancement and Dereverberation with Diffusion-based Generative Models", *IEEE/ACM Trans. Audio, Speech, Language Proc.*, Vol. 31, pp. 2351 - 2364, 2023.
5. Julius Richter, Simone Frintrop, Timo Gerkmann, "Audio-Visual Speech Enhancement with Score-Based Generative Models", *ITG Speech Communication*, Aachen, Germany, Sep. 2023.

1.3 Prof. Dr. rer. nat. habil. Frank Steinicke

Frank Steinicke is a professor of Human-Computer Interaction at the Department of Informatics at the Universität Hamburg. He received the IEEE VGTC Virtual Reality Technical Achievement Award and was inducted into the prestigious IEEE VR Academy in 2023. Furthermore, he has been a member of the Academy of Science and Humanities in Hamburg since 2024.

Academic qualifications and professional experiences

Study	1997-2001 Diploma in Mathematics (Minor: Computer Science), Westfälische Wilhelms University (WWU) Münster, Germany
Doctoral	2006 Doctoral deegree in Computer Science, WWU Münster, Germany
Habilitation	2009 Venia Legendi in Computer Science, Westfälische Wilhelms University Münster, Germany
Professional experiences	Since 2014 (W3) Professor of Human-Computer Interaction, Department of Computer Science, University of Hamburg, Germany 2011-2014 (W2) Professor of Computer Science in Media, Department of Computer Science, University of Würzburg, Germany 2007-2011 Postdoctoral Researcher, Department of Computer Science, Westfälische Wilhelms University Münster, Germany 2002-2006 Research Associate, Department of Computer Science, Westfälische Wilhelms University Münster, Germany

Publications

1. Mostajeran, F., Steinicke, F., Reinhart, S., Stuerzlinger, W., Riecke, B.E. & Kühn, S. (2023). Adding virtual plants leads to higher cognitive performance and psychological well-being in virtual reality, Nature Scientific Reports 13 (1), 8053.
2. Kruse L., Mostajeran F., Steinicke F. (2023). High Levels of Visual Anthropomorphic Representations of Virtual Agents Increase the Social Presence of Users. IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).
3. Schmidt S., Ariza O., Steinicke F. (2020). Intelligent Blended Agents: Reality–Virtuality Interaction with Artificially Intelligent Embodied Virtual Humans. Multimodal Technologies and Interaction, 4 (4), 85.
4. Steinicke, F. (2016). *Being really virtual*. Immersive natives and the future of virtual reality. Springer.
5. Valkov, D., Steinicke, F., Bruder, G., Hinrichs, K. (2011). 2D Touching of 3D Stereoscopic Objects. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), pp. 1353-1362.

1.4 Prof. Dr. Stefano Panzeri

Stefano Panzeri is Professor and Director of UKE's Institute for Neural Information Processing. He was elected Fellow of the ELLIS (European Laboratory for Learning and Intelligent Systems). Moreover, he serves as Area Chair of NeurIPS and ICML and is Visiting Scholar at Harvard Medical School.

Academic qualifications and professional experiences

Study	1988-1993 University of Turin, MSc in Physics, Summa cum Laude
Doctoral degree	1991-1996 PhD in Computational Neuroscience, International School for Advanced Studies (SISSA), Trieste, Italy
Habilitation	2016: Italian Full Professor National Habilitation in: Bioengineering, Neurophysiology, General Psychology, and Theoretical Physics
Professional experiences	Since 2021, Nucleus Professor and Institute Director, Institute for Neural Information Processing, UKE, Hamburg, Germany 2013- 2020: Senior Scientist and Research Centre Director, Centre for Neuroscience and Cognitive Systems, Italian Institute of Technology 2012 – 2013: Full Professor University of Glasgow, UK 2008 – 2012: Senior Scientist, Italian Institute of Technology 2003 - 2009: Reader, University of Manchester; Course Director, MSc Programme in Computational Neuroscience and Neuroinformatics 1999 –2003: MRC Fellow in Neuroinformatics, University of Newcastle 1996 –1999: Marie Curie Research Fellow , University of Oxford

Publications

1. Kuan AT, Bondanelli G, Driscoll L, Han J, Kim M, Hildebrand D, Graham B, Wilson D, Thomas L, Panzeri S, Harvey CD, Lee W-CA (2024). Synaptic wiring motifs in posterior parietal cortex support decision-making. *Nature* 627, 367–373 (Senior co-corresponding)
2. Celotto M, Bím J, Tlaie A, De Feo V, Hanganu I, Donner T, Brovelli A, Panzeri S (2023) An information-theoretic quantification of the content of communication between brain regions. A*dvances in Neural Information Processing Systems (NeurIPS)* 36, 64213–64265
3. Valente M, Pica G, Bondanelli G, Moroni M, Runyan C, Morcos A, Harvey C, Panzeri S (2021) Correlations enhance the behavioral readout of neural population activity in association cortex. *Nature Neuroscience*, 24, 975–986
4. Giordano B, Ince R, Gross J, Panzeri S, Kayser C (2017) Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLIFE* 6, e24763
5. Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S (2015) Rhythmic Auditory Cortex Activity at Multiple Timescales Shapes Stimulus–Response Gain and Background Firing. *J Neurosci* 35: 7750 –7762

2. References

1. Thakur K, Barker HG, Pathan A-SK. Artificial Intelligence and Large Language Models: An Introduction to the Technological Future. CRC Press; 2024.
2. Balasubramaniam S, Kadry S, Prasanth A, Dhanaraj RK. Generative AI and LLMs: Natural Language Processing and Generative Adversarial Networks. Walter de Gruyter GmbH & Co KG; 2024.
3. Hameed H, Lubna, Usman M, Kazim JUR, Assaleh K, Arshad K, et al. Artificial intelligence enabled smart mask for speech recognition for future hearing devices. Sci Rep. 2024;14: 30112.
4. Bouton K. Smart Hearing: Strategies, Skills, and Resources for Living Better with Hearing Loss. Riverwest Press; 2018.
5. Al Mubarak M. Innovative and Intelligent Digital Technologies; Towards an Increased Efficiency: Volume 2. Springer Nature;
6. Lu L, Wu D, Zhang X, Chen L. Linguistic Processing of Unattended Speech Under a Cocktail Party Listening Scenario. J Speech Lang Hear Res. 2025;68: 69–78.
7. Cherry EC. Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am. 1953;25: 975–979.
8. Ostrand AE, Johnson V, Gazzaley A, Zanto TP. Neural Correlates of the Musicianship Advantage to the Cocktail Party Effect. J Cogn Neurosci. 2025; 1–11.
9. Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science. 2008;320: 110–113.
10. Cohen TA, Patel VL, Shortliffe EH. Intelligent Systems in Medicine and Health: The Role of AI. Springer Nature; 2022.
11. Shaban-Nejad A, Michalowski M, Buckeridge DL. Explainable AI in Healthcare and Medicine: Building a Culture of Transparency and Accountability. Springer Nature; 2020.
12. Zgallai WA, Ozsahin DU. Artificial Intelligence and Image Processing in Medical Imaging. Elsevier; 2024.
13. Pieraccini R. AI Assistants. MIT Press; 2021.
14. Jiao L, Ma M, He P, Geng X, Liu X, Liu F, et al. Brain-Inspired Learning, Perception, and Cognition: A Comprehensive Review. IEEE Trans Neural Netw Learn Syst. 2024;PP. doi:10.1109/TNNLS.2024.3401711
15. Best V, Boyd AD, Sen K. An Effect of Gaze Direction in Cocktail Party Listening. Trends Hear. 2023;27: 23312165231152356.
16. Ephrat A, Mosseri I, Lang O, Dekel T, Wilson K, Hassidim A, et al. Looking to listen at the cocktail party: A speaker-independent audio-visual model for speech separation. arXiv [cs.SD]. 2018. Available: http://arxiv.org/abs/1804.03619
17. Zia T. How AI Solves the “Cocktail Party Problem” and Its Impact on Future Audio Technologies. In: Unite.AI [Internet]. 2024. Available: https://www.unite.ai
18. Kraaijeveld SR, van Heijster H, Bol N, Bevelander KE. The ethics of using virtual assistants to help people in vulnerable positions access care. J Med Ethics. 2025.

doi:10.1136/jme-2024-110464

19. Dobie RA. Medical-Legal Evaluation of Hearing Loss, Third Edition. Plural Publishing; 2015.
20. Burkey JM. The Hearing-Loss Guide: Useful Information and Advice for Patients and Families. Yale University Press; 2015.
21. Khan SB, Namasudra S, Chandna S, Mashat A. Innovations in Artificial Intelligence and Human-Computer Interaction in the Digital Era. Elsevier; 2023.
22. Tripathy BK, Seetha H. Explainable, Interpretable, and Transparent AI Systems. 2024.
23. Salvi M, Seoni S, Campagner A, Gertych A, Acharya UR, Molinari F, et al. Explainability and uncertainty: Two sides of the same coin for enhancing the interpretability of deep learning models in healthcare. Int J Med Inform. 2025;197: 105846.
24. Yeong DJ, Panduru K, Walsh J. Exploring the Unseen: A Survey of Multi-Sensor Fusion and the Role of Explainable AI (XAI) in Autonomous Vehicles. Sensors (Basel). 2025;25. doi:10.3390/s25030856
25. Yılmaz C. The Case for Open-Source AI: Ensuring Fairness, Accountability, and Ethical Innovation. Independently Published; 2024.
26. Reed C. AI Fairness and Beyond: Law, Regulation, and Technology. Bloomsbury Publishing; 2024.
27. Dignum V. Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer Nature; 2019.
28. Steinicke F. B(l)ending Realities. 2021 IEEE Virtual Reality and 3D User Interfaces (VR). IEEE; 2021. doi:10.1109/vr50410.2021.00015
29. Schmidt S, Ariza O, Steinicke F. Intelligent Blended Agents: Reality–Virtuality Interaction with Artificially Intelligent Embodied Virtual Humans. Multimodal Technologies and Interaction. 2020;4: 85.
30. Gao W. Speech Synthesis and Personalization under Unimodal and Multimodal Conditions. Transactions on Computer Science and Intelligent Systems Research. 2024;7: 126–137.
31. Korkmaz L, Schmidt S, Kruse L, Steinicke F. I talk - you write. Exploring asymmetric text- and voice-based communication between humans and virtual agents. Proceedings of the 12th International Conference on Human-Agent Interaction. New York, NY, USA: ACM; 2024. pp. 25–33.
32. Schmidt S, Nunez OJA, Steinicke F. Blended Agents: Manipulation of Physical Objects within Mixed Reality Environments and Beyond. Symposium on Spatial User Interaction. New York, NY, USA: Association for Computing Machinery; 2019. pp. 1–10.
33. Schmidt S, Rolff T, Voigt H, Offe M, Steinicke F. Natural expression of a machine learning model's uncertainty through verbal and non-verbal behavior of intelligent virtual agents. Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology. New York, NY, USA: ACM; 2024. pp. 1–15.
34. Keller JK, Kusari A, Czok S, Simgen B, Steinicke F, Diekhof EK. ACHOO - Bless you! Sense of Presence can provoke Proactive Mucosal Immune Responses in Immersive Human-Agent Interactions. 2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR). IEEE; 2024. pp. 557–567.

35. Kruse L, Hertel J, Mostajeran F, Schmidt S, Steinicke F. Would you go to a virtual doctor? A systematic literature review on user preferences for embodied virtual agents in healthcare. 2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE; 2023. doi:10.1109/ismar59233.2023.00082

36. Kruse L, Mostajeran F, Steinicke F. The influence of virtual agent visibility in virtual reality cognitive training. Proceedings of the 2023 ACM Symposium on Spatial User Interaction. New York, NY, USA: ACM; 2023. pp. 1–9.

37. Schmidt S, Steinicke F. Intelligent Blended Agents: Reality-Virtuality Interaction with Artificially Intelligent Embodied Virtual Humans. Journal of Multimodal Technologies and Interaction. : (Invited Paper).

38. Litovsky RY, Goupell MJ, Fay RR, Popper AN. Binaural Hearing. Springer; 2021.

39. Lister JJ, Carmenate-Nichols R, Hudak EM, O'Brien JL, Edwards JD. Auditory Processing but Not Peripheral Hearing Differs Between Older Adults With and Without Mild Cognitive Impairment. Am J Audiol. 2024;33: 718–727.

40. Moray N. Attention: Selective Processes in Vision and Hearing. Routledge; 2017.

41. Li D, Gao Y, Zhu C, Wang Q, Wang R. Improving speech recognition performance in noisy environments by enhancing lip reading accuracy. Sensors (Basel). 2023;23. doi:10.3390/s23042053

42. Bregman AS. Auditory scene analysis and the role of phenomenology in experimental psychology. Can Psychol. 2005;46: 32–40.

43. Driver J, Davis G, Russell C, Turatto M, Freeman E. Segmentation, attention and phenomenal visual objects. Cognition. 2001;80: 61–95.

44. Klatt L-I, Begau A, Schneider D, Wascher E, Getzmann S. Cross-modal interactions at the audiovisual cocktail-party revealed by behavior, ERPs, and neural oscillations. Neuroimage. 2023;271: 120022.

45. Bizley JK, Jones GP, Town SM. Where are multisensory signals combined for perceptual decision-making? Curr Opin Neurobiol. 2016;40: 31–37.

46. Krumbholz K, Nobis EA, Weatheritt RJ, Fink GR. Executive control of spatial attention shifts in the auditory compared to the visual modality. Hum Brain Mapp. 2009;30: 1457–1469.

47. Renvall H, Seol J, Tuominen R, Sorger B, Riecke L, Salmelin R. Selective auditory attention within naturalistic scenes modulates reactivity to speech sounds. Eur J Neurosci. 2021;54: 7626–7641.

48. Dehaene S, Kerszberg M, Changeux JP. A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci U S A. 1998;95: 14529–14534.

49. Changeux J-P, Goulas A, Hilgetag CC. A connectomic hypothesis for the hominization of the brain. Cereb Cortex. 2021;31: 2425–2449.

50. Hilgetag CC, Goulas A, Changeux J-P. A natural cortical axis connecting the outside and inside of the human brain. Netw Neurosci. 2022;6: 950–959.

51. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, et al. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 2013;11: e1001752.

52. Kayser C, Montemurro MA, Logothetis NK, Panzeri S. Spike-phase coding boosts and

stabilizes information carried by spatial and temporal spike patterns. Neuron. 2009;61: 597–608.

53. Articles, Systems/Circuits Rhythmic Auditory Cortex Activity Multiple Timescales Shapes Stimulus-Response Gain Background Firing.

54. Ng BSW, Schroeder T, Kayser C. A precluding but not ensuring role of entrained low-frequency oscillations for auditory perception. J Neurosci. 2012;32: 12268–12276.

55. Giordano BL, Ince RAA, Gross J, Schyns PG, Panzeri S, Kayser C. Contributions of local speech encoding and functional connectivity to audio-visual speech perception. Elife. 2017;6. doi:10.7554/eLife.24763

56. Lu R, Duan Z, Zhang C. Listen and look: Audio–visual matching assisted speech source separation. IEEE Signal Process Lett. 2018;25: 1315–1319.

57. Tesch K, Gerkmann T. Insights into deep non-linear filters for improved multi-channel speech enhancement. IEEE ACM Trans Audio Speech Lang Process. 2023;31: 563–575.

58. Lemercier J-M, Richter J, Welker S, Moliner E, Välimäki V, Gerkmann T. Diffusion Models for Audio Restoration: A review. IEEE Signal Process Mag. 2024;41: 72–84.

59. Richter J, Frintrop S, Gerkmann T. Audio-visual speech enhancement with score-based generative models. arXiv [eess.AS]. 2023. Available: http://arxiv.org/abs/2306.01432

60. Simpson AJR, Roma G, Plumbley MD. Deep karaoke: Extracting vocals from musical mixtures using a convolutional deep neural network. Latent Variable Analysis and Signal Separation. Cham: Springer International Publishing; 2015. pp. 429–436.

61. Bermant PC. BioCPPNet: automatic bioacoustic source separation with deep neural networks. Sci Rep. 2021;11: 23502.

62. Heymann J, Drude L, Haeb-Umbach R. Neural network based spectral mask estimation for acoustic beamforming. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE; 2016. doi:10.1109/icassp.2016.7471664

63. Ephrat A, Mosseri I, Lang O, Dekel T, Wilson K, Hassidim A, et al. Looking to listen at the cocktail party. ACM Trans Graph. 2018;37: 1–11.

64. Rahimi A, Afouras T, Zisserman A. Reading to listen at the cocktail party: Multi-modal speech separation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2022. doi:10.1109/cvpr52688.2022.01024

65. Richter J, Welker S, Lemercier J-M, Lay B, Gerkmann T. Speech enhancement and dereverberation with diffusion-based generative models. IEEE ACM Trans Audio Speech Lang Process. 2023;31: 2351–2364.

66. Yang GR, Joglekar MR, Song HF, Newsome WT, Wang X-J. Task representations in neural networks trained to perform many cognitive tasks. Nat Neurosci. 2019;22: 297–306.

67. Mazzoni A, Whittingstall K, Brunel N, Logothetis NK, Panzeri S. Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. Neuroimage. 2010;52: 956–972.

68. Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE. The leading sense: supramodal control of neurophysiological context by attention. Neuron. 2009;64: 419–430.

69. Ghitza O, Greenberg S. On the possible role of brain rhythms in speech perception:

intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. Phonetica. 2009;66: 113–126.

70. Dogonasheva O, Zakharov D, Giraud A-L, Gutkin B. Neuro-oscillatory models of cortical speech processing. arXiv [q-bio.NC]. 2025. Available: http://arxiv.org/abs/2502.12935

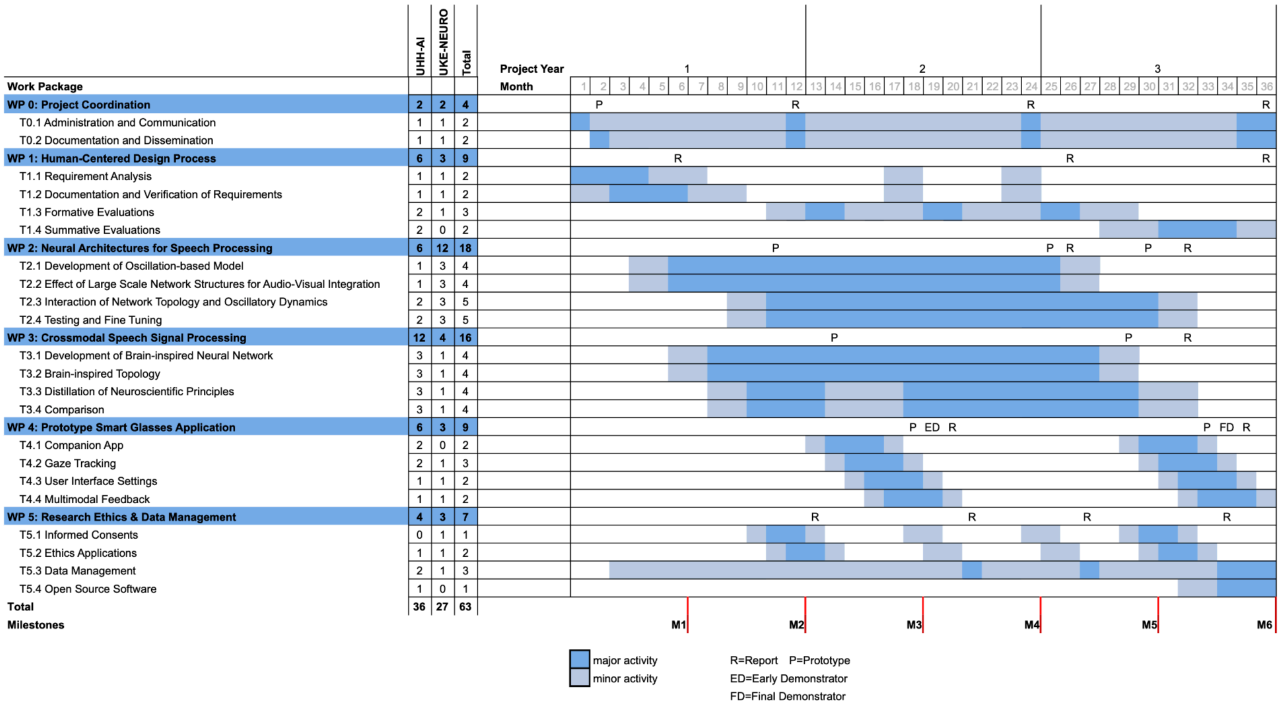
71. Suzanne A, Raschia G, Martinez J, Tassetti D. Slicing techniques for temporal aggregation in spanning event streams. Inf Comput. 2021;281: 104807.

72. Rolff T, Schmidt S, Steinicke F, Frintrop S. Deep Learning Architecture Egocentric Time-to-Saccade Prediction using Weibull Mixture-Models Historic Priors. ETRA ’23: Proceedings of the 2023 Symposium on Eye Tracking Research and Applications. 2023. pp. 1–8.

73. Rolff T, Steinicke F, Frintrop S. When do Saccades begin? Prediction of Saccades as a Time-to-Event Problem. 2022 Symposium on Eye Tracking Research and Applications. New York, NY, USA: ACM; 2022. doi:10.1145/3517031.3529627

2. Gantt Chart & Milestones

The following Gantt chart shows an estimate of the person months (PMs) per work package and task for the planned three years. Please note that the doctoral student of the UHH-AI team will have a 100% E13 position, and the UKE-NEURO doctoral student will assume a 75% E13 position. Therefore, the PMs add up to 36 PM (UHH-AI) + 27 PM (UKE-NEURO) = 63 PM for the entire project.



Milestone	Month	Description
M1	6	First iteration of requirements analysis document is finished
M2	12	Ethics application for first formative evaluations is finished
M3	18	Early demonstrator has been finalized
M4	24	Neural architecture are integrated in crossmodal speech signal processing
M5	30	Summative evalaution is finished
M6	36	Final demonstrator is completed and project is finished

3. Letters of Interest

HOSPITAL ZUM HEILIGEN GEIST
Lebensvielfalt für Senioren

HOSPITAL ZUM HEILIGEN GEIST • HINSBLEEK 11 • 22391 HAMBURG

Prof. Dr. Frank Steinicke
Human-Computer Interaction
Department of Informatics
Universität Hamburg
Vogt-Kölln-Str. 30
22527 Hamburg

Ihr Ansprechpartner:
Nizar Müller
Fundraising, Innovations- und Projektmanagement
Telefon 040 60601-115
Telefax 040 60601- 22 115
muellernizar@hzhg.de

04.03.2025

Letter of intent zum Forschungsvorhaben CHAIS

Sehr geehrte Frau Dr. Hüttner,
sehr geehrte Frau Dr. Kälin,
sehr geehrte Herr Prof. Dr. Steinicke,

hiermit möchten wir als Hospital zum Heiligen Geist ausdrücklich unsere Unterstützung des Forschungsvorhabens „Crossmodal Human-AI Collaboration for Selective Hearing Assistance“ (CHAIS) ausdrücken, welches im Rahmen des BMBF-Forschungsprogramms zur BMBF-Bekanntmachung zu Pilotprojekten zum Thema „Neurobiologisch inspirierte Künstliche Intelligenz" durchgeführt werden soll.

Das übergeordnete Ziel von CHAIS ist es, dass Synergiepotential von KI und den Neurowissenschaften zum Beispiel bei der Entwicklung von intelligenten Hörgeräten aufzuzeigen. Somit leistet das Projekt einen wichtigen Beitrag beim Aufzeigen der Möglichkeiten und Chancen für eine gegenseitige Inspiration und Integration von Neurowissenschaften und KI in Pilotprojekten an konkreten gesellschaftlich relevanten Beispielen.

Das Hospital zum Heiligen Geist mit Sitz im Stadtteil Poppenbüttel ist mit rund 1100 Bewohner:innen und 900 Mitarbeitenden eine der größte Alten- und Pflegeheime in Hamburg. Gerne unterstützen wir das Projekt darin, die Technologie unseren Bewohner:innen für Tests zur Verfügung zu stellen.

Wir würden uns über eine erfolgreiche Förderung des Vorhabens freuen und sehen der Zusammenarbeit zuversichtlich entgegen.

Wir wünschen viel Erfolg bei der Antragstellung.

Mit freundlichen Grüßen

M. Kröger

Michael Kröger
Vorsitzender des Vorstandes

HOSPITAL ZUM HEILIGEN GEIST mit Oberalten-Stift, Marien-Magdalenen-Kloster und Altendank
Hinsbleek 11 · 22391 Hamburg · Telefon 040 60601-111 · Telefax 040 60601-129 · info@hzhg.de · www.hzhg.de
Vorstand: Michael Kröger (Vorsitz) · N. N. · Steuernummer 17/409/01318
Vorsitzender des Verwaltungsrates: Dr. Michael Labe · Präses des Kollegiums der Oberalten
Bankverbindung: Hamburger Sparkasse · BIC: HASPDEHHXXX · IBAN: DE95 2005 0550 1299 1237 84

ADVANCED BIONICS
POWERFUL CONNECTIONS

ADVANCED BIONICS GmbH | Feodor-Lynen-Str. 35 | DE-30625 Hannover
Prof. Dr.-Ing. Timo Gerkmann
Universität Hamburg
Department of Informatics
SP Research Group

Hannover, 04.03.2025

Letter of intent for support of the proposed project "Crossmodal Human-AI Collaboration for Selective Hearing Assistance (CHAIS)" in the framework of the BMBF initiative "Neurobiologisch inspirierte Künstliche Intelligenz"

Dear Dr. Hüttner, Dr. Kälin, Prof. Dr. Steinicke,
Prof. Dr.-Ing. Gerkmann, Prof. Dr. Hilgetag and Prof. Dr. Panzeri,

this letter of intent indicates our interest and willingness to support and contribute to the project called "Crossmodol Human-AI Collaboration for Selective Hearing Assistance (CHAIS)" in the framework of the BMBF call "Neurobiologisch inspirierte Künstliche Intelligenz".

Advanced Bionics (www.advancedbionics.com) is part of the Sonova group which is a leading provider of innovative hearing care solutions ranging from hearing aids to cochlear implants. Speech understanding in challenging situations with background noise and reverberation (the so-called cocktail party phenomena) is very difficult for all our users with state-of-the-art technology. To solve the problem of speech understanding in noise, the proposed project wants to combine the knowledge of AI based signal processing with the mechanisms known from neurosciences in e.g. hearing aids. To benefit from the knowledge of both domains and to leverage possible synergies for the development of hearing aids has the potential to make an impact for users that are severely impacted by their hearing loss in their daily life.

To maintain our leading position, we rely on systematic research and development in the fields addressed by the envisaged research project, especially in the fields of AI-based signal processing. In view of this, we consider the proposed research project as an outstanding opportunity to work with our academic partners on a topic that can impact users' performance and, therefore, their quality of life significantly. We strongly believe that the proposed research will have the potential to improve technology, e.g. new signal processing algorithms improving sound perception in adverse listening conditions such as noise, reverberation and with interfering speakers.

We intend to actively participate in and contribute to the envisaged research program. Please note that Advanced Bionics' actual commitment to support the CHAIS project will be subject to further internal compliance, legal and financial review and approval, and the conclusion of the appropriate agreement. No rights can be conferred from this letter.

Sincerely,

Dr. Volkmar Hamacher

Managing Director AB GmbH

A Sonova brand

Advanced Bionics GmbH, Feodor-Lynen-Str. 35, 30625 Hannover, Germany
+49.511.524.87.50 – info.de@advancedbionics.com
Amtsgericht Hannover Nr. HRB 203338 – Geschäftsführer: Volkmar Hamacher – Deike Markgraf

4. Data Management Plan

Data management plan:V0.1 – 28.02.2025

DATA MANAGEMENT PLAN

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Universitätsklinikum Hamburg-Eppendorf

Speech Enhancement
Neuroinformatics
Artificial Intelligence
Neuroscience
Human-AI Interaction
Computational Neuroscience

Crossmodal Human-AI Collaboration for Selective Hearing Assistance

PROJECT	
Project number:	[project number]
Project acronym:	CHAIS
Project name:	Crossmodal Human-AI Collaboration for Selective Hearing Assistance

DATA MANAGEMENT PLAN	
Date:	28.02.2025
Version:	0.1

1. Introduction

1.1. Purpose of this document

This document serves as the initial iteration of the Ethics Framework and Data Management Plan (DMP) for the CHAIS research project. It includes definitions of the ethical and legal background and offers guidance on data management policies within the scope of the project. The DMP in its infal version will be based on the project's Grant and Consortium Agreement (GA/CA), adheres to the FAIR (findable, accessible, interoperable, reusable) data principles, and complies with the General Data Protection Regulation (GDPR) of the European Union (EU) 2016/679.

1.2. Scope of this document

The DMP covers mainly the following aspects:

- Presenting a comprehensive overview of the data that the CHAIS project partners will generate, collect, or re-use during the project period.

- Providing guidelines for handling data in compliance with the FAIR data principles, ensuring that the data remains findable, accessible, interoperable, and reusable.
- Establishing guidelines for handling potential Intellectual Property Rights (IPR) on data that will be generated, collected, or re-used during the project period.
- Implementing strict policies for promoting physical and logical data security, especially regarding personal data governed by the GDPR.
- Providing an ethical framework for all data collection activities involving human participants, emphasising the principles of respect, privacy, and informed consent.

1.3. Status of this document

This document presents a preliminary version of the DMP delivered prior project submission. It comprises all information on the origin, nature, and management of data as well as ethical foundations known or defined prior the project. The DMP is a living document that will become more detailed in upcoming versions and face possible changes due to the design and execution of the planned research

2. Data Summary

The following sections offer an overview of the data that the CHAIS project partners will generate, collect, or re-use throughout the project's duration. They provide a summary of user and technical studies that will be or have been conducted within the scope of the project to keep track of all collected data.

Expected purpose of the data and their relations to the objectives of the project:
UHH/UKE will mainly generate data by conducting user studies. The data will be used to evaluate the solutions and user experience aspects and will consist of participant information and Human-AI interaction related measures (e.g., usability, user experience). Hence, the data will include demographics (e.g., age, gender, experience), subjective questionnaires (e.g., SUS, UEQ, TAM), and objective interaction performance measures (e.g., required time for specific tasks, number of change of parameters, learning time). In addition, it is necessary to perform audio or video recordings from interactions during user studies used for post-hoc analysis. During the software development process realizing the technical solutions, text-based scripts, image and audio files will be generated or re-used. In addition, text-based data will be generated for producing different deliverables or research outputs like scientific publications.
Expected re-use of existing data and its purpose:
Priorly collected experimental data from internal or external sources might be re-used for between-subject evaluations or benchmark comparisons. Publicly available assets might be used within software developments. Deliverables of other projects with public dissemination can serve as exemplary input for the deliverables to be generated.
Expected data types and formats:
Most data generated will be stored as ASCII-based data files that can be converted to proprietary file formats. We plan to store collected or re-used data in the following formats: • Text: TXT, RTF, MD, DOCX • Tables: CSV JSON, XLSX • Graphics: PNG, JPG, SVG, • Audio: WAV, MP3 • Video: MP4 • Others: PDF, CS
Expected size of the data:
Collected experimental data is usually text-based and will not exceed a total size of 1GB. For the purpose of recording experiments, it might also be necessary to record audio, video, or movement data usually leading to a larger amount of data with a total size of up to 1 TB. Software development repositories will also be smaller than 1 GB when containing sample

applications.
Expected origin/provenance of the data:
Regardless of its purpose, the generated or re-used data will originate from several sources, which include internal or external preexisting data, data from the scientific literature, public data repositories, and data collected during user studies.
Expected use of the data outside the project:
Our generated anonymized data originating from user studies might be useful to other researchers also performing user studies related to our domain or beyond. It can serve as comparative or benchmark data and increase the reproducibility of our results. Software developed within the project that will be published might be useful for other researchers.

3. FAIR data

The FAIR data guidelines presented in the next sections aim to improve the findability, accessibility, interoperability, and re-usability of digital data generated, collected, and processed within the CHAIS project. These principles are mandatory for each partner to follow to make publishable data and other research output publicly accessible and generally usable without technical, legal, or semantic barriers and thus maximise the impact and benefits of project investments. FAIR data not only supports the re-use of data by individuals such as scientists but is also intended to enhance data collaboration across different fields and support computer systems like large language models to find and use the data automatically. Within the project, each partner generating or collecting data is responsible for publishing data following the principle of “as open as possible and as closed as necessary” while adhering to the FAIR data guidelines.

Following Wilkinson et al., 2016, we consider the following aspects to be the critical requirements for implementing the FAIR data principles in the CHAIS project.

Requirements for data **to be findable:**

F1. (Meta)data need to be assigned to a globally unique and eternally persistent identifier

F2. Data need to be described with rich metadata

F3. (Meta)data need to be registered or indexed in a searchable resource

F4. Metadata needs to specify the data identifier

Requirements for data to **be accessible:**

A1. (Meta)data need to be retrievable by their identifier using a standardised communications protocol

A1.1. The protocol needs to be open, free, and universally implementable.

A1.2. The protocol needs to allow for an authentication and authorisation procedure where necessary

A2. Metadata needs to be accessible, even when the data are no longer available

Requirements for data to **be interoperable:**

I1. (Meta)data needs to use a formal, accessible, shared, and broadly applicable language for knowledge representation

I2. (Meta)data needs to use vocabularies that follow FAIR principles

I3. (Meta)data needs to include qualified references to other (meta)data

Requirements for data to **be re-usable:**

R1. (Meta)data needs to have a plurality of accurate and relevant attributes

R1.1. (Meta)data needs to be released with a clear and accessible data usage license

R1.2. (Meta)data needs to be associated with their provenance

R1.3. (Meta)data needs to meet domain-relevant community standards

3.1. Making data findable, including provisions for metadata

Data generated and processed within the CHAIS project should be easily locatable by both humans and computers. To fulfil this, each partner generating or collecting data is responsible for ensuring that all appropriate data will be made openly available through an institutional (e.g., RDM) or open data repository (e.g., Zenodo, EOSC). To ensure that the data is findable easily, rapidly, and identically, exact and standard parameters need to be used to identify the datasets. Each published dataset is required to receive a globally unique and persistent identifier (e.g., DOI or Handle) and each contributing individual must have a personal identifier (e.g., ORCID) to make both easily findable. In addition, all datasets need to receive rich and indexable metadata, ideally according to a commonly used metadata schema (e.g., Dublin Core), to make them easy to discover. The metadata needs to provide information at least about datasets (description, date of deposit, author(s), venue, and embargo), grant project name, acronym and number, licensing terms, persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs. The added metadata should further contain information about methods used to capture the data, the measuring instruments used, the individual who performed the collection or generation, the data formats and units, and the date and location when the data was created. Additionally, the metadata should include appropriate keywords describing the content of the dataset and offer information on how it can be re-used and how the included data needs to be interpreted.

3.2. Making data accessible

Once the data published by the CHAIS project has been found, it should be easily accessible. Therefore, the datasets need to be stored openly available in an institutional (e.g., FDM for UHAM) or open data repository (e.g., Zenodo, EOSC), where they can be retrieved using open, free, and universally implementable protocols (e.g. HTTPS) provided by web services that also allow downloading the datasets. The chosen repository must ensure that uploaded datasets receive a globally unique and persistent identifier (e.g., DOI or Handle) by which they can be accessed. If additional software is needed to access the data, documentation on how the data can be accessed should be included in the dataset. In general, all datasets associated with research output will be made accessible as default unless there is a specific reason not to publish the data. Other data may be made available on a case-by-case basis if it is relevant for third parties. The following data will not be made publicly available:

- Data that contains personal or sensitive information.
- Data was obtained with the permission of third parties, but the third parties have not agreed to make the data publicly available.
- Data that discloses the identity of a manufacturer.
- Data that has been declared confidential by the CA.
- Data for which a partner lodges an objection to publication within a 15-day objection period.
- Data that compromises the protection of a partner(s) IPR or goes against their legitimate interests or other constraints as per the GA.

If public access to the data needs to be limited for any legitimate reason, the dataset needs to be listed under restricted access but will not be made available directly. Third parties and partners of CHAIS may request access by contacting the issuing partner. In this case, the access rights are managed via the data repository. In addition, data may need an embargo period to allow time to publish the research output, protect intellectual property, or adhere to publisher conditions. For this reason, an embargo period of a maximum of two years can be applied to retain the data before it must be published. The metadata of the datasets must be made available openly and licensed using the Creative Common Public Domain Dedication (CC 0) license or equivalent that allows for redistribution. Information about the accessibility of the dataset needs to be included in the metadata. All published data and metadata will remain available as long as the repository remains active. The metadata should remain available if the data needs to be removed for any specific reason.

3.3. Making data interoperable

The CHAIS project aims to make published data as interoperable as possible. However, as the exact nature of all data has not been completely defined at this project stage, it is not yet feasible to derive a comprehensive set of guidelines that fully guarantees data interoperability. In general, all partners should aim to publish their data using common interoperable data formats like TXT, MD, CSV, JSON, XML, PNG, JPG, or PDF. If it becomes apparent that either proprietary or non-interoperable data formats need to be used, it must be described how interoperability can be achieved. For data where no agreed metadata standards exist, we aim to ensure interoperability by providing appropriate documentation on how to handle the data, including suggestions for compatible open-source software to handle the data. When naming and describing data, CHAIS project partners are encouraged to use controlled vocabulary as far as available, adhere to the common terminology used in the respective field, and avoid ontologies that are not commonly used. If non-common vocabulary or ontologies will be used, they must be described in the data. To implement controlled vocabulary, it is recommended to consult the various controlled vocabulary entries on the Publications Office of the European Union website. Furthermore, it is recommended that research output identifiers be linked to published datasets and vice versa.

3.4. Increase data re-use

The CHAIS project aims to promote the re-usability of data that has been made publicly accessible. For this reason, the published datasets must contain comprehensive interoperable documentation addressing the data structure, the definition of variables, and the units of measurement. Ideally, the documentation is included in the corresponding research output. Where possible, datasets should be made accessible as early as possible after production following the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a license with equivalent rights. Experimental and test data might only become available after the publication of the results. We expect most of the data generated to be made available without restrictions and only datasets subject to IPR and other confidentiality issues will be restricted. Where this will be the case, agreements will be made based on the individual datasets. Requests for using the data by externals will be approved by the partner issuing the dataset.

To further support data re-usability, the CHAIS project partners must ensure high data quality. To this end, the quality control of the data needs to happen at various stages during the quality assurance process. As the quality of the data collection methods used strongly influences data quality, an initial quality check must be carried out at the beginning of the data collection by the partners who were responsible for planning and implementing the measures. The initial check should include control of the data integrity and validation of the collected measures. Further controls should be performed by other involved project partners when available. A final quality check regarding the integrity of the dataset and metadata needs to be performed by the issuing partner before publishing the data.

4. Other research outputs

The CHAIS project aims to uphold the FAIR data principles not only for data generated, collected, or re-used during performed studies but also for any other generated research output. Therefore, the following sections define the data handling procedures concerning any other expected research outputs.

4.1. Academic Publications

Scientific publications generated within the context of the CHAIS project are considered data that need to be covered by the FAIR and CARE data principles. All partners must guarantee open access to peer-reviewed scientific publications relating to their results. In particular, they must ensure that a machine-readable electronic copy of the published version, or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications no later than at the time of publication. Immediate open access needs to be provided to the deposited publication via the repository under the latest available version of the Creative Commons Attribution International Public Licence (CC BY) or a license with equivalent rights. For monographs and other long-text formats, the license may exclude commercial uses and derivative works (e.g., CC BY-NC or CC BY-ND). Further information needs to be given via the repository about any research output or any other tools and instruments that are needed to validate the conclusions of the scientific publication. To this end, partners and authors must retain sufficient IPR over their publications to ensure compliance with the access requirements. Before publishing any academic output, the partners involved need to contact the technical coordinator for revision/validation.

4.2. Software

Software developed within the context of the CHAIS project should also adhere to the FAIR and CARE data principles and be made publicly available under the conditions of open-source software licenses (e.g., MIT, GNU, Apache) using either an open software repository (e.g., GitHub) or institutional solutions (e.g., GitLab). To this end, each partner acting as a licensor of software developed as part of the project, which is not subject to IPR, is responsible for making the software publicly available. However, it can be expected that parts of the software will be subject to IPR preventing it from being published openly. Software being distributed within the project should be stored using institutional software repositories that permit access only to authorised project partners, as further outlined in the section “Data Security”.

5. Allocation of resources

PI Steinicke is responsible for overseeing the overall data management within the CHAIS project in its position as task leader of WP 5.3 (Project Ethics, IPR, and Data Management) and as project coordinator. However, partners who generate or collect data are responsible for ensuring that the respective data adheres to the FAIR data principles. No further resources have been explicitly allocated within the GA for the purpose of managing the data of the project.

In general, the consortium aims to make all data accessible using either institutional or publicly available data repositories. These are free of charge to use and, to our knowledge, will remain so also in the future. Accordingly, we expect that the costs for making data accessible will remain free of charge, ensuring long-term preservation. Regarding making scientific publications open access, publishers that offer open access free of charge should be favoured. Otherwise, it is up to the respective partners to decide whether and how they can cover the costs. Indirect workload incurred because of the effort required to comply with the FAIR data principles needs to be reported within the corresponding WPs. Hence, it is not expected that the PRESENCE project will have additional costs to implement the FAIR data principles. Any changes to this policy or exceptions will be included in future versions of the DMP.

6. Data security

All data management activities within the CHAIS project need to adhere to the data security policies described in this section. To this end, each partner of the consortium is responsible for implementing the policies, including the logical security, physical security, network security, and security of computer systems, to ensure the security of data and prevention of unauthorised access, changes to data, disclosure, or destruction of data.

6.1. Personal Data Security

The CHAIS project partners must fully comply with the GDPR to ensure lawful and ethical use and avoid data breaches of personal data. For this purpose, the ethical handling of personal data beyond the general policies for logical and physical data security is further specified in the section about ethics of this document.

6.2. Personal Data Security

To protect unpublished data generated, collected, or re-used within the CHAIS project from unauthorised access, secure user authentication methods need to be applied to restrict the access of data. To this end, each member of the project must use unique credentials to access project data, including consistently strong passwords and two-factor authentication (2FA) where possible.

Project partners are not allowed to store any confidential data, such as those containing personal information, on servers or computers connected to an external or free network. An exception to this is the sharing of data between partners of the CHAIS consortium, which must be realised through secure institutional solutions within the individual WPs and adhere to the here-described data security policies. In this case, strict access levels must be assigned based on the user's role within the project. For instance, project or technical coordinators or task leads should have full access to all data, while executing partners should have restricted access based on their tasks. A version control system (e.g., Git) should be used to manage changes to data and ensure that accurate records of modifications are maintained.

To maintain network security, firewall protection and security-related upgrades and patches for the operating systems of PCs and servers must be put in place to prevent attacks through breaches, viruses, and malicious code. Data transmitted between systems using an internet connection need to be encrypted using safe transport layer security protocols (e.g., HTTPS) to protect against interception and eavesdropping. Automated full backups of all data should be conducted weekly, with incremental backups performed daily, also including PCs and workstations used by project members. Backup copies should be stored in secure off-site locations to protect against data loss due to hardware failure or other disasters as further described in the next section about physical data security. In the event of data loss or corruption, the responsible partner needs to initiate data recovery procedures, aiming to restore data within 24 hours.

6.3. Physical Data Security

To assure physical data security, all CHAIS partners need to implement highly restricted access to the buildings where data, computers, or servers are held. Transporting of data storage media is restricted to exceptional circumstances (i.e., repair of physical systems). Ideally, the servers are protected against data loss and downtime due to power failures by surge protection systems and uninterruptible power supply systems.

To ensure the security of local computer systems and files, all members of the CHAIS partners must always lock their computer systems with a strong password when currently not in use. To avoid breaches or loss of data in case of theft or loss of an end device, data should be stored in encrypted form (e.g., AES-256) on the hard drives of the end devices and backups should be performed regularly as described above.

7. Ethics

7.1. CARE Principles

The CARE Principles (Collective Benefit, Authority to Control, Responsibility, Ethics) are a set of principles intended to guide open data projects in engaging rights and interests of indigenous people. CARE was created in 2019 by the International Indigenous Data Sovereignty Interest Group, a group that is a part of the Research Data Alliance. It outlines collective rights related to open data in the context of the United Nations Declaration on the Rights of Indigenous Peiple and Indigenous data sovereignty.

The CARE Principles are 'people and purpose-oriented, reflecting the crucial role of data in advancing Indigenous innovation and self determination', and intended as a complement to the data-oriented perspective of other standards such as FAIR data (findable, accessible, interoperable, reusable).

- **Collective Benefit**: Data ecosystems shall be designed and function in ways that enable Indigenous Peoples to derive benefit from the data.
- **Authority to Control**: Indigenous Peoples' rights and interests in Indigenous data must be recognised and their authority to control such data be empowered. Indigenous data governance enables Indigenous Peoples and governing bodies to determine how Indigenous Peoples, as well as Indigenous lands, territories, resources, knowledges and geographical indicators, are represented and identified within data.
- **Responsibility**: Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous Peoples' selfdetermination and collective

benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.

- **Ethics**: Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

7.2. Ethical Considerations

The CHAIS project consortium aims to maintain the highest ethical standards recommended by professional bodies, institutions, and governments both during user studies and when generating, collecting, or re-using data. In general, all planned user studies need to be performed in accordance with the Declaration of Helsinki. Particular attention shall be paid to:

- The principle of proportionality.
- The right to privacy.
- The right to the protection of personal data.
- The right to the physical and mental integrity of a person.
- The right to non-discrimination.
- The need to ensure the protection of the environment.
- The need to ensure high levels of human health protection.

Before conducting a study involving human participants, participating partners should critically scrutinise whether their planed evaluation activities will contradict the above-mentioned principles in any way. To this end, an ethics self-assessment form is being developed and will be appended to future versions of this document. In case the critical self-assessment raises ethical concerns, it is imperative that studies involving human participants are not conducted unless approved by the research ethics committee (REC) of a participating partner.

Partner Research Ethics Committee (REC)

- Ethics Commission from the Department of Computer Science at Universität
- Clinical Ethics at the University Medical Center Hamburg-Eppendorf!

Partners who do not have access to a REC need to cooperate with a partner who has access to a REC or obtain external ethics approval. Partners conducting user studies without official ethical approval must be aware that publishing scientific results and data in high-impact publication targets often requires the approval of a REC or at least proof of compliance with local ethical guidelines. To avoid data breaches leading to ethical objections, any generated, collected, or re-used data, particularly data originating from user studies, must always strictly adhere to the guidelines for data security described above.

7.3. Recruitment, information, and consent for data collection

When recruiting participants for user studies as part of the CHAIS project, particular care must be taken to ensure a diverse selection in terms of gender, ethnicity, identification, age, orientation, education, beliefs, and more. Recruitment of adult healthy volunteers should be favored whenever possible, with exceptions requiring ethical approval. All human participants who intend to take part in a user study performed by any project partners must be informed about the purpose of the study and the collected data, the procedure they are undergoing, and the potential risks related to it. They must explicitly agree to participate in the study and confirm that they have been comprehensively informed about it. Where studies might involve some level of discomfort, participants must be explicitly informed of this point in the written information and verbal briefing before participation. All participants will be free to withdraw at any stage without reason. They may receive compensation for their participation depending on the guidelines and best practices of the institution concerned. Data from user studies with human participants may only be made publicly accessible if the participants consent to the publication of their data and the data is properly depersonalised. Sample documents for information and consent will be provided, which includes a comprehensive description of the foundations for conducting user studies involving human participants within the project.

benefit. Accountability requires meaningful and openly available evidence of these efforts and the benefits accruing to Indigenous Peoples.

- **Ethics**: Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

7.2. Ethical Considerations

The CHAIS project consortium aims to maintain the highest ethical standards recommended by professional bodies, institutions, and governments both during user studies and when generating, collecting, or re-using data. In general, all planned user studies need to be performed in accordance with the Declaration of Helsinki. Particular attention shall be paid to:

- The principle of proportionality.
- The right to privacy.
- The right to the protection of personal data.
- The right to the physical and mental integrity of a person.
- The right to non-discrimination.
- The need to ensure the protection of the environment.
- The need to ensure high levels of human health protection.

Before conducting a study involving human participants, participating partners should critically scrutinise whether their planed evaluation activities will contradict the above-mentioned principles in any way. To this end, an ethics self-assessment form is being developed and will be appended to future versions of this document. In case the critical self-assessment raises ethical concerns, it is imperative that studies involving human participants are not conducted unless approved by the research ethics committee (REC) of a participating partner.

Partner Research Ethics Committee (REC)

- Ethics Commission from the Department of Computer Science at Universität
- Clinical Ethics at the University Medical Center Hamburg-Eppendorf!

Partners who do not have access to a REC need to cooperate with a partner who has access to a REC or obtain external ethics approval. Partners conducting user studies without official ethical approval must be aware that publishing scientific results and data in high-impact publication targets often requires the approval of a REC or at least proof of compliance with local ethical guidelines. To avoid data breaches leading to ethical objections, any generated, collected, or re-used data, particularly data originating from user studies, must always strictly adhere to the guidelines for data security described above.

7.3. Recruitment, information, and consent for data collection

When recruiting participants for user studies as part of the CHAIS project, particular care must be taken to ensure a diverse selection in terms of gender, ethnicity, identification, age, orientation, education, beliefs, and more. Recruitment of adult healthy volunteers should be favored whenever possible, with exceptions requiring ethical approval. All human participants who intend to take part in a user study performed by any project partners must be informed about the purpose of the study and the collected data, the procedure they are undergoing, and the potential risks related to it. They must explicitly agree to participate in the study and confirm that they have been comprehensively informed about it. Where studies might involve some level of discomfort, participants must be explicitly informed of this point in the written information and verbal briefing before participation. All participants will be free to withdraw at any stage without reason. They may receive compensation for their participation depending on the guidelines and best practices of the institution concerned. Data from user studies with human participants may only be made publicly accessible if the participants consent to the publication of their data and the data is properly depersonalised. Sample documents for information and consent will be provided, which includes a comprehensive description of the foundations for conducting user studies involving human participants within the project.