112

AUDITORY IMPLANTS

Using artificial intelligence for the enhancement of speech in noise perception in typical-hearing adults

Anne van Alphen, Feyisayo E. Olalere, H. Christiaan Stronks, Kiki A. van der Heijden, Yağmur Güçlütürk, Jeroen J. Briaire, Marcel A.J. van Gerven, Johan H.M. Frijns

Introduction

A cochlear implant (CI) is a device placed in the inner ear to treat severe-to-profound sensorineural hearing loss (Figure 2). CI users have great difficulty to understand speech in background noise, particularly when multiple talkers are present. To address this, a custom-designed, artificial intelligence driven algorithm was developed to separate two talkers using bilateral CI microphone input (Figure 1). The bilateral CIs allow for the retrieval of both phase and intensity differences, which can be used by the algorithm to locate and separate the talkers. Fundamental frequency differences between target and interfering talker could also a potential cue for the separation of the talkers by the algorithm.

The algorithm was trained based on supervised learning, i.e., using labeled datasets of male and female talkers with known patterns to recognize for the algorithm. This knowledge is then applied to newly presented two-talker speech to separate the talkers from each other, regardless of gender pairings.

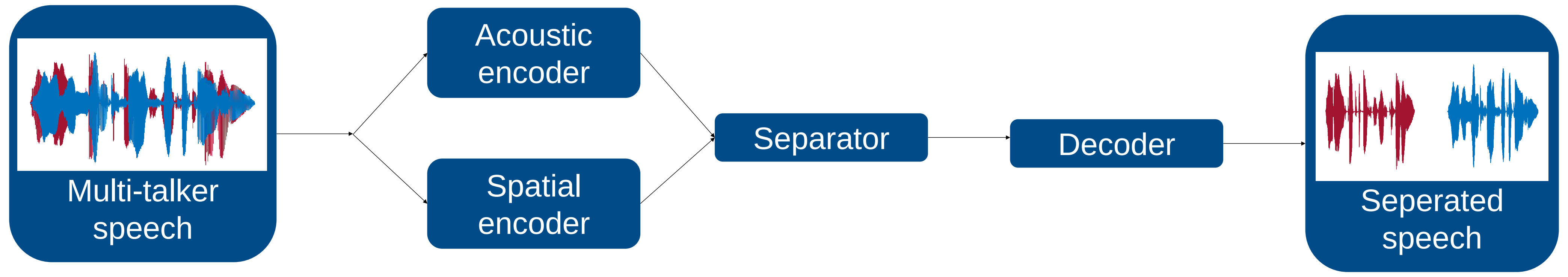

Figure 1. Functionality of the algorithm. Two-speaker speech is provided to the algorithm. The speech of both talkers are seperated.

Hypothesis

The algorithm performs best when the angle between the target and competing talker is 90º, because both the phase and the intensity cues are strongest. At 0º, performance is expected to be low.

Objectives

Determine the effectivity of the custom-designed, artificial intelligence driven algorithm to enhance speech recognition in the presence of competing speech in typical-hearing adults.

Methods and Materials

Different gender pairings were presented to the algorithm to measure its efficiency based on objective intelligibility measures, namely the Short-Time Objectivity Intelligibility(1) (STOI) and Scale Invariant Signal-to-Distortion-Ratio(2) (SI-SDR).

To determine the baseline SNR without the algorithm for use in the clinical trial, 10 typical-hearing individuals were tested using two angles between target and competing talker (0º and 90º). The baseline SNR was determined at 50% speech recognition (i.e., at the speech recognition threshold; SRT). The participants were presented with a female target and a male competing talker through headphones. All speech material was preprocessed offline using the head-related transfer function for CI-users and appropriate room acoustics. The female target (FT, LIST sentences(3)) was presented at 0º and 90º, and the male competing talker (MC, VU98 sentences(4)) always came from the front, i.e., at 0º (Figure 3). Signal-to-noise ratios (SNRs) were varied to fit a psychometric curve and calculate the SRT, with SRT defined as the signal-to-noise ratio where speech recognition is 50%. Word correct scores were used as outcome measure. Sound level was 60 dB(A) on average, with an additional broadband (LTSS) noise of 70 dB(A) to increase difficulty.

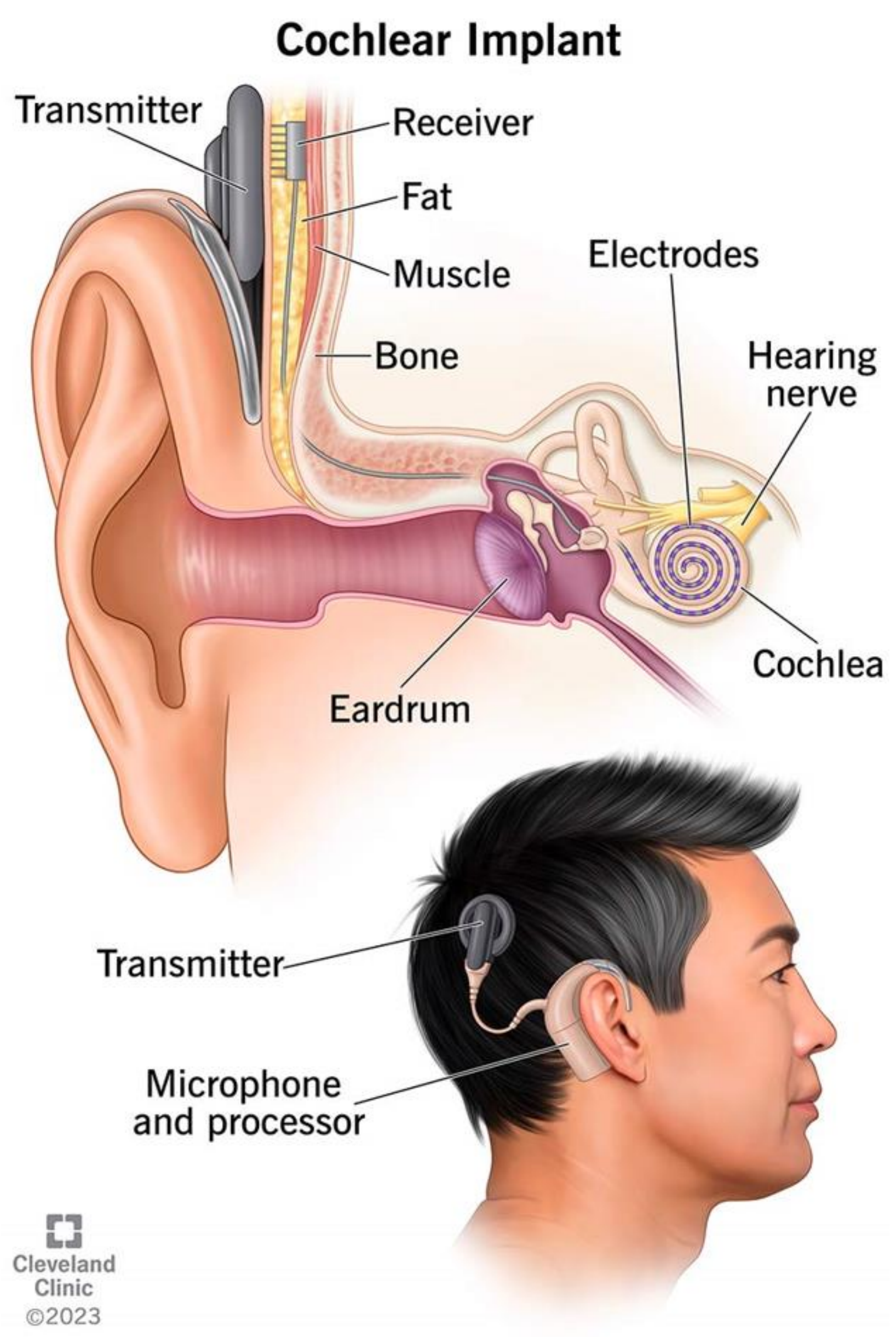

Figure 2. Overview cochlear implant
Retrieved from Cleveland Clinic
(Ohio, United States)

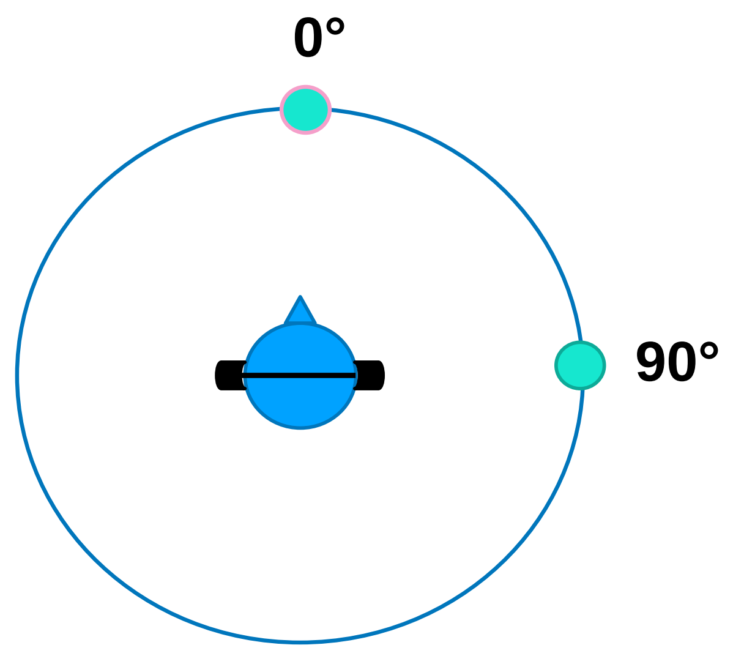

Figure 3. Study setup.
Speech material is played through headphones, with the male speaker at 0º and female changing location.

Results I – in silico

The algorithm showed better results when trained with both spatial and spectral than only spectral cues (Table 1). Also, a significant relation between SI-SDR and separation angle was found for female-male speech pairings (p=0.002).

Table 1. In silico results algorithm.

Training materials for algorithm	STOI
Two channel audio with spectral and spatial cues	0.78
Two channel audio with spectral cues	0.77
Reference: single channel audio with spatial and reverberant aspect	0.70

Results II – Baseline performance in participants

Figure 4 shows the psychometric curves of all participant data. The red cross markers show the mean score associated to the measured SNR level, with the red line fitted afterwards to determine the SRT level (blue diamond). The dashed line represents the steepness of the slope of the fitted curve. A mean SRT -8 dB(A) SNR (re. competing talker level) was calculated for the 90º condition (Figure 4A), and -2 dB SNR for 0º (Figure 4B). A statistically significant difference was found between the SRTs of the two test conditions (p=0.002). Data collection is still ongoing.

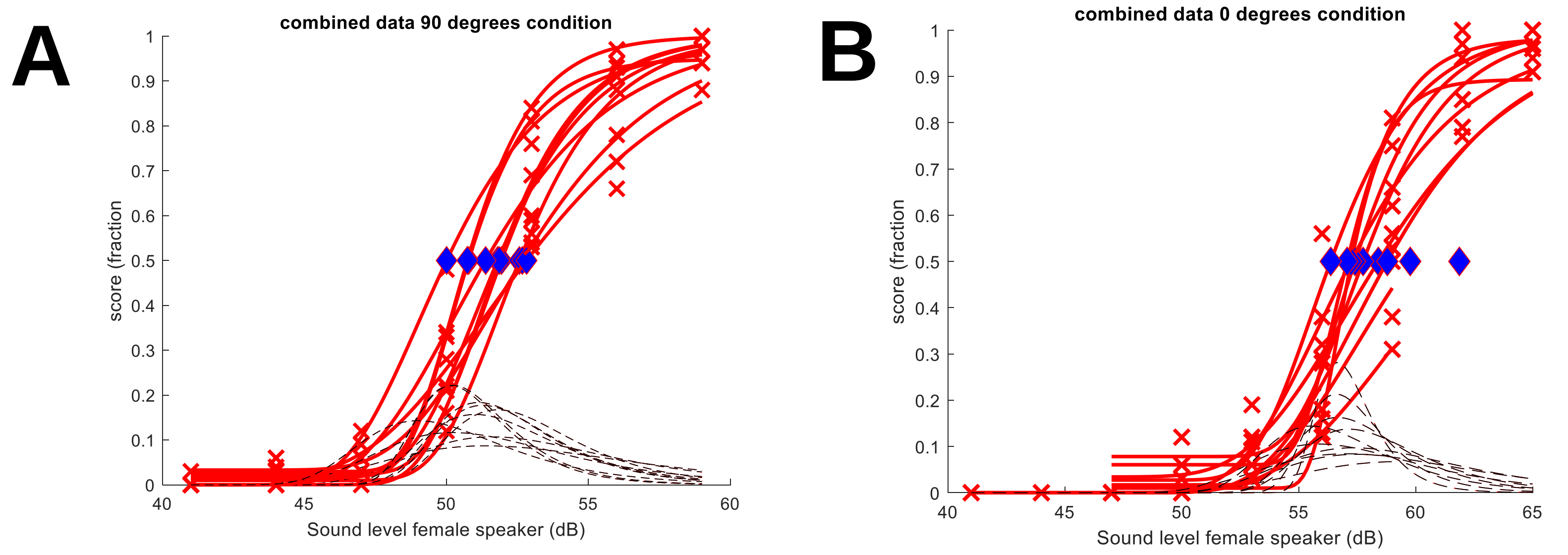

Figure 4. Study results.
A: 90º condition.
B: 0º condition.

Conclusion

The algorithm shows a predicted increase in speech intelligibility (STOI) of 8% in two-talker situations. Speech recognition is significantly better in the 90º condition. Test conditions will be adjusted based on the baseline performance found.

Future work

Ongoing work will reveal the effectiveness of the algorithm in typical hearing people and CI users. The conditions that will be tested are FT-only, FT-MC, and FT-MC with algorithm. Additional angles will be tested, namely 30º and 165º.

References

(1) Taal et al. (2010). 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. https://doi.org/10.1109/icassp.2010.5495701

(2) Roux et al. (2018). arXiv (Cornell University). https://doi.org/10.48550/arxiv.1811.02508

(3) van Wieringen et al. (2008). Int J Audiol, 47(6), 348–355. https://doi.org/10.1080/14992020801895144

(4) Versfeld, N. J., Daalder, L., Festen, J. M., & Houtgast, T. (2000). JASA, 107(3), 1671–1684. https://doi.org/10.1121/1.428451

Leiden University Medical Center
Advanced Bionics
NWO Dutch Organization for Scientific Research
INTENSE INNOVATIVE NEUROTECHNOLOGY FOR SOCIETY
DONDERS INSTITUTE
19›22 September 2024
Paris, France