Speech-to-touch sensory substitution: a 10-decibel improvement in speech-in-noise understanding after a short training

Katarzyna Ciesla (✉ kasia.j.ciesla@gmail.com)

The Baruch Ivcher Institute for Brain, Cognition & Technology, The Baruch Ivcher School of Psychology, Interdisciplinary Center Herzliya (IDC), Herzliya

T. Wolak

World Hearing Centre, Institute of Physiology and Pathology of Hearing, Warsaw

A. Lorens

World Hearing Centre, Institute of Physiology and Pathology of Hearing, Warsaw

H. Skarżyński

World Hearing Centre, Institute of Physiology and Pathology of Hearing, Warsaw

A. Amedi

The Baruch Ivcher Institute for Brain, Cognition & Technology, The Baruch Ivcher School of Psychology, Interdisciplinary Center Herzliya (IDC), Herzliya

Research Article

Keywords: Speech understanding in noise, sensory substitution device, vibrotactile stimulation, cochlear implants, multisensory training, hearing impairment, multisensory rehabilitation

Posted Date: April 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-429202/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Understanding speech in background noise is challenging. Wearing face-masks during COVID19-pandemics made it even harder. We developed a multi-sensory setup, including a sensory substitution device (SSD) that can deliver speech simultaneously through audition and as vibrations on fingertips. After a short training session, participants significantly improved (16 out of 17) in speech-in-noise understanding, when added vibrations corresponded to low-frequencies extracted from the sentence. The level of understanding was maintained after training, when the loudness of the background noise doubled (mean group improvement of ~ 10 decibels). This result indicates that our solution can be very useful for the hearing-impaired patients. Even more interestingly, the improvement was transferred to a post-training situation when the touch input was removed, showing that we can apply the setup for auditory rehabilitation in cochlear implant-users. Future wearable implementations of our SSD can also be used in real-life situations, when talking on the phone or learning a foreign language. We discuss the basic science implications of our findings, such as we show that even in adulthood a new pairing can be established between a neuronal computation (speech processing) and an atypical sensory modality (*tactile*). Speech is indeed a multisensory signal, but learned from birth in an audio-*visual* context. Interestingly, adding lip reading cues to speech in noise provides benefit of the same or lower magnitude as we report here for adding touch.

Introduction

The COVID19 pandemics imposed on us the obligation of social distancing and wearing masks covering the mouth. Both these restrictions reduce transmission of sounds and prevent access to visual cues from speech/lip reading. Meanwhile, access to speech reading and efficient integration of this visual information with the auditory input is very important for enhanced understanding of speech to people with hearing loss, who naturally receive degraded auditory cues from the environment [1, 2]. Healthy individuals also use lip reading in all face-to-face communication, and especially benefit from it when the acoustic context is ambiguous, such as when exposed to rapid speech, speech in non-native language, background noise and/or more speakers talking simultaneously [3, 4, 5]. With visual cues present, understanding speech against noise has consistently been found to improve, in both healthy individuals and in patients with hearing loss [1, 6, 7, 8, 9].

In today's modern society challenging acoustic conditions occur increasingly more often, including exposure to concurrent multiple auditory streams and almost constant exposure to noise. At the same time, the prevalence of hearing impairments is growing (WHO 2020), which disease if not treated, can accelerate cognitive decline and thereby social exclusion [10, 11, 12]. In addition, many users of modern hearing aids and cochlear implants complain that their devices fail to effectively compensate for hearing loss when they are exposed to ambiguous acoustic situations, as described above [13, 14, 15, 16].

All this indicates the importance of developing novel methods and devices that can be employed to improve communication, both among healthy individuals as well as those with hearing deficits.

Especially solutions combining multisensory inputs are appealing, as increasingly more experimental studies [17, 18, 19, 20] and conceptual works point to the superiority of multisensory over unisensory training for learning and sensory recovery [21, 22, 23, 24, 25]. For improved understanding of auditory signals that are either degraded or presented in suboptimal conditions, multisensory training regimes that complement audition with *vision* have been found successful [20, 25], including for rehabilitation of patients with hearing aids or cochlear implants, by adding speech reading or sign cues [1, 6].

Interestingly, several recent works showed benefits of adding *tactile* stimulation to improve degraded auditory speech comprehension, including our own findings [26, 27, 28]. Applying a combination of auditory and tactile stimulation in assistive communication devices is an interesting and novel approach, that can prove useful especially if access to visual cues during communication is limited. Audition and touch share some key similarities, such as both use mechanoreceptors to encode vibration in a shared range of frequencies (approx. 50Hz to 700Hz). Therefore, vibrotactile and auditory information can be naturally perceived as an interleaved signal, which is then also processed in partially shared brain regions, in both hearing participants as well as in congenitally deaf [29, 30, 31, 32, 33, 34, 35].

Given these similarities, we developed an in-house audio-to-touch (assistive) Sensory Substitution Device (SSD). SSDs convey information typically delivered by one sensory-modality (e.g., vision) through a different sensory modality (audition, touch) using specific translation algorithms that can be learned by the user [36, 37, 38]. A classic example is a chair developed by Prof Bach-Y-Rita, that delivered a visual image to the back of its blind users through patterns of vibration [39].

In the first work with our audio-to-touch SSD we showed immediate and robust improvement by mean 6dB in speech-in-noise understanding (Speech Reception Threshold), when auditory speech was complemented with low-frequency tactile vibration delivered on fingertips [28]. Importantly, in this previous experiment the improvement occurred without applying any training, as opposed to other numerous works using SSDs which required hours of training and/or prolonged use to yield benefits, probably due to the complexity of the applied algorithms [40, 41, 42, 43, 44].

The key goal of the current study was to investigate whether understanding speech delivered as a multisensory audio-tactile input can improve even further when a short training session is applied. This research is important for both practical reasons and for enhancing basic science understanding of multisensory processing and cross-modal integration. It is especially intriguing, as the audio-tactile multisensory context is utterly novel for speech perception and only learned in adulthood. This contrasts with audio-visual speech, i.e. listening to speech and speech/lip reading at the same time, which is a natural language input acquired in critical/sensitive periods of development and to which we are all exposed throughout lifetime. In terms of practical applications of our research, we believe that it can inform development of potentially rapid and successful rehabilitation protocols using touch to improve speech perception for patients with hearing loss, as well as the design of novel technologies to assist the general public in practical cases. This include speech understanding in noisy environments when wearing a mask, when trying to learn a new language or talking on the phone.

Material And Methods

Seventeen ($N = 17$) native Hebrew speakers (6 male/11 female; mean 27 years) participated in the study, mostly University students or their friends. Participants were all right-handed and reported no history of neurological/neurodevelopmental impairments. All were also fluent speakers of English and used English on every day basis. All research was performed in accordance with all relevant guidelines/regulations and the Declaration of Helsinki. An informed consent was obtained from all participants and they were compensated for participation. The experiment was approved by the Institutional Review Board of IDC Herzliya, the School of Psychology.

Subjects participated in a series of tasks. In all of them we used our in-house audio-to-touch Sensory Substitution Device (SSD) (designed in collaboration with the World Hearing Centre in Warsaw, Poland and <http://www.neurodevice.pl/en>), noise-cancelling headphones (BOSE QC35 IIA), a 5.1 soundcard (Creative Labs, SB1095) and a PC. The SSD was used to deliver speech signal as tactile vibration on the index and middle finger of the dominant hand. A dedicated MatLab (version R2016a, The MathWorks Inc., Natick, MA, USA) application with a user-friendly GUI was developed to run the study. Only a brief description of the study set-up is provided here, as all the details with accompanying figures can be found in our previous publication [28].

In the current study all participants took part in three tests of speech comprehension, once before and once after a short training session. Introducing one additional test condition and a short, dedicated training session (to see whether subjects will improve further in understanding of novel non-trained sentences) were the main differences in the experimental paradigm, as compared to our previous work [28].

Practice

The first test session was proceeded with a brief practice of listening to two different *vocoded* (the sound was modified to resemble stimulation through a cochlear implant system) and non-vocoded sentences from the English HINT sentence database [45] with and without accompanying vibration, to familiarize the participants with the study setting. The details of the vocoding procedure can be found in our previous work [28].

Speech Reception Threshold tests

In the actual tests, as well as in the training session, the task of the participants was to repeat sentences presented *via* noise-cancelling headphones. All the sentences were vocoded using an in-house algorithm and presented against background noise (IFFN; URL: <https://www.ehima.com/>). There were 3 test conditions, both before and after training, all with sentences presented in the headphones: a) with no concurrent vibration delivered on fingertips (audio only; hereafter A for Audio), b) together with low frequencies including the fundamental frequency (f_0) extracted from the heard sentence, i.e. matching f_0 , delivered as vibration on fingertips (hereafter AM, as in “Audio-Matching”), c) together with low

frequencies, including the fundamental frequency not corresponding, i.e. non-matching, to the heard sentence, delivered as vibration on fingertips (hereafter AnM). The A condition always came first, followed by the AM condition in 9 participants, and by the AnM condition in 8 participants. As described in detail in our previous work [28], f_0 had been earlier extracted from each original sentence derived from the English HINT database, using the STRAIGHT algorithm that was further improved. The outcome measure of each test was Signal to Noise Ratio (SNR) for 50% understanding, i.e. the Speech Reception Threshold (SRT) of the target (vocoded) lists of sentences against background noise. For each test 20 different sentences were used (2 lists of 10 HINT sentences) and the difficulty level was adapted based on individual performance (2dB up – 2dB down). The applied adaptive procedure is a typical procedure used in the clinical ENT setting when assessing speech understanding in hearing aid and/or cochlear implant users [46]. After the training session participants performed the second round of 3 test conditions (A, AM, AnM).

Training

After the initial series of tests each person participated in a short training session. The training consisted of listening to and repeating 148 vocoded sentences, each accompanied with matching low frequencies delivered on fingertips of two fingers *via* our SSD. The SRT calculated for every person in the test condition involving auditory and matching tactile input (AM) prior to the training was used as the SNR throughout the whole training session. Each sentence was presented up to 3 times as a combined audio-tactile input, and if the person was unable to repeat it correctly, the sentence was presented as text on the PC screen in front of the participant (black font, middle of the screen). The sentence that required such visual feedback remained in the database of the training sentences and was presented again at the end of the session. The training continued until all 148 sentences were repeated correctly without visual feedback. In all tested participants the training was approx. 35–50 min in duration. The feedback was decided to be visual as opposed to auditory, as in the future the authors wish to apply the same testing (and training) procedures in participants with hearing deficits.

Results

Most importantly, we found that a short multisensory training session provides robust statistically significant improvement of almost 10dB in SRT for understanding speech in noise, when accompanied with matching tactile frequencies delivered on fingertips. The 10 dB difference in SRT indicates that after training the same performance (50 % understanding) was achieved with the noise level twice louder. For the trained AM condition the mean improvement was 9.8 ± 6.8 dB (from 12.1 ± 5 dB to 2.4 ± 5.4 dB, $p = 0.001$, Wilcoxon Signed Rank test, two-tailed asymptomatic significance), with 16 out of the 17 tested participants obtaining a better SRT after training for novel untrained lists of sentences. This result is shown in Figs. 1A and 1B.

After training the obtained SRT for the AM test condition was significantly lower than the SRT reported for the AnM condition, indicating better performance ($p = 0.007$). Before the training scores in these two conditions were not statistically different ($p = 0.3$). The SRT for the AnM condition also improved after training, by 3.8 ± 6.5 dB (from 11.2 ± 6.7 dB to 7.3 ± 4.8 dB), with 13 out of the 17 tested subjects

showing improvement (see Fig. 2A and 2B). The improvement was of a lesser extent and at a much lower level of statistical significance ($p = 0.02$, Wilcoxon Signed Rank test), when compared to the trained AM condition.

The third tested condition, audio only (A), was always presented as first, both before and after training. Therefore, we refrain from comparing the SRT values obtained for this test condition with the AM and AnM conditions, as the differences may reflect solely the effect of order. Nevertheless, it should be noted that the SRT for the A condition also improved significantly after the AM training session, from 17.7 ± 7.7 dB to 7 ± 5 dB ($p = 0.000$), with 14 out of 17 participants showing improvement. We will discuss in the *Discussion* section why this outcome is important with respect to the development of multisensory training regimens for the hearing-impaired population.

An additional analysis that we performed revealed positive statistically significant correlations between some of the SRT test outcomes, and namely, between: (a) the AM and the AnM SRT values before training ($r = 0.5$, $p = 0.04$), as well as (b) the A and the AM SRT values after training ($r = 0.51$, $p = 0.035$).

Discussion

We show here that understanding of speech in noise improves significantly after a short and simple training session when auditory sentences are combined with a tactile input that corresponds to their low frequencies, including the fundamental frequency (f_0). The effect that we report for the trained condition (though using novel sentences), a mean decrease of 10 dB in the speech recognition threshold, is profound, as it represents maintaining the same performance with background noise perceived twice louder [47]. The current work expands our previous findings [28] which showed that understanding degraded speech against background noise improves immediately by mean 6 dB in SRT without any training, when accompanied with corresponding and synchronized low frequencies delivered as vibrotactile input on fingertips. Interestingly, in both our studies we reported a very similar mean score in the audio-tactile condition with matching vibration (AM before training in the current study) with the mean group SRT of approx. 12 dB. The sentence database, the applied vocoding algorithm and the background noise were identical in both experiments. Obtaining similar scores in a total of almost 30 participants strengthens the interpretations we propose. Interestingly, and as we expected, the AM training applied in the current study also led to improved speech in noise understanding in the unisensory, auditory only, test condition.

These main outcomes of the study have several implications, both for basic science as well as in terms of their potential practical applications. Intriguingly, we show here that the improvement of speech comprehension through audio-tactile training was in the order of magnitude (~ 10 dB in SRT) similar or higher than that found for audio-visual speech-in-noise testing settings, when the performance is compared to only auditory [5, 48, 49, 50, 51]. The reported improvement in SRT was by 3 to 11 dB, although the applied methodology and the language content varied. The implications of this finding for basic science are discussed further in text. In addition, showing improvement in speech understanding,

both in the multisensory and the unisensory test setting is especially important with respect to the development of novel technologies for the general public and rehabilitation programs for patients with hearing problems. We discuss that in some more detail below.

Multisensory perceptual learning

Findings of our study are an example of rapid perceptual learning, which has already been shown for various acoustic contexts and distortions of the auditory signal, including natural speech presented in background noise, as well synthetically manipulated vocoded/time-compressed speech [25, 52, 53, 54]. Furthermore, our results are in agreement with the emerging scientific literature demonstrating that learning and memory of sensory/cognitive experiences is more efficient when the applied inputs are multisensory [22, 23]. The experimental procedure was also specifically designed to benefit from the fundamental rule of multisensory integration, and namely the inverse effectiveness rule [1, 55, 56, 57]. Although mainly shown for more basic sensory inputs, the rule predicts that multisensory enhancement, i.e. the benefit of adding information through an additional sensory channel, is especially profound for low signal-to-noise conditions and when learning novel tasks [56]. In our study the auditory speech signal was new to the study participants, degraded (vocoded), presented against background noise and in their non-native language. All these manipulations lead to a low signal to noise context and deemed the task of understanding the sentences challenging, thereby increasing the chance of improvement in performance *via* adding a specifically designed tactile input.

At the same time, the study conditions were ecologically valid, in that we aimed to recreate an every-day challenging acoustic situation encountered by both hearing impaired patients and healthy people, of being exposed to two auditory streams at the same time. We showed here that auditory stream segregation, and specifically focusing on one speaker can be facilitated by adding vibrotactile stimulation which is congruent with the target speaker (see [5] discussing similar benefits of audio-visual binding).

Design of the SSD – considerations for the current study

To deliver tactile inputs, we developed our own audio-to-touch SSD. In the ever growing body of literature it has been suggested that SSDs can advance the benefits of a multisensory training even further than mentioned above, since they convey one modality input *via* another one in a way that is specifically tailored to the neuronal computations characteristic for the original sensory modality [39, 38, 37, 36, 40, 58, 59, 60, 61]. In our study we used the SSD to deliver an input complementary to the auditory speech through touch which, nevertheless, maintained features typical for the auditory modality, as vibrations are also a periodic signal that fluctuates in frequency and intensity.

At the same time the applied frequency range of the inputs was detectable by the tactile Pacinian cells that are most densely represented on the fingertips and most sensitive for coding frequencies in the range of 150-300Hz (and up to 700-1000Hz) [62]. Specifically, the tactile vibration that was provided on fingertips of our participants was part of the signal below (and including) the extracted fundamental

frequency of the speech signal. Access to this low-frequency aspect of the temporal fine structure of speech is specifically reduced in patients with sensorineural hearing loss, including those using a cochlear implant. It has been shown that lack of this input profoundly impairs speech understanding in challenging acoustic situations, especially with several competing speakers [63, 64, 65].

Apart from the current and the previous work of our lab, also one other research group showed that adding low-frequency input delivered on fingertips can improve comprehension of auditory speech presented in background noise in normal hearing individuals [26]. The main advantage of our approach was, however, that to estimate the SNR for various task conditions we applied an adaptive procedure, as opposed to using fixed SNRs, thereby avoiding floor and ceiling effects [46]. In addition, we showed a more significant benefit of adding the specifically designed vibrotactile input than the other group, possibly due the fact that the stimuli we used were in the non-native language of the participants, which further enhanced the effect of the inverse effectiveness rule. Using the native language of the participants, and thus yielding the task easier was maybe also the reason why Fletcher and colleagues failed to show benefit from adding the vibrotactile input before training (which we did in our previous work, [28]).

The audio-tactile interplay

Interestingly, in the current study we also saw improvement in speech understanding in the control test condition, and namely when the auditory sentences were paired with vibro-tactile inputs corresponding to a different sentence than the one presented through audition, i.e., paired with non-matching vibration (AnM). The degree of improvement was, however, far less robust (mean of 4dB in SRT) and far less statistically significant than the improvement reported for the trained AM condition (mean of almost 10dB in SRT). The reason why we introduced this control test condition, was to acquire more information about the mechanisms of multisensory enhancement. Interestingly, we found that the group scores before training for both conditions combining an auditory and a tactile input (matching and non-matching) were almost the same; and they also correlated ($r = 0.5$). We hypothesize that at this point the participants were still becoming familiarized with the study set-up and the vibrotactile input, and only after a short, dedicated training session they “realized” the true benefit of the matching tactile f0. It is also possible that presenting these two conditions next to one another in the pre-training session might have caused confusion.

In our future studies we would like to investigate the role of the vibrotactile input delivered on two fingertips and the audio-tactile binding effects further. It remains to be studied whether it is of crucial importance that the amplitude and frequency fluctuations delivered as vibrations follow the auditory signal precisely. Alternatively, non-matching vibrations whose spectro-temporal characteristics are nevertheless still very different from the background speech signal will provide the same benefit, if trained. To answer this question, in our future work we will include two additional training groups, one exposed to the degraded speech input with concurrent vibrotactile stimulation that represents the fundamental frequency that does not match the auditory sentence (an alternative multisensory training), and one receiving training that is only auditory (a unisensory training).

Implications for rehabilitation

The findings of our study have implications for auditory rehabilitation programs for the hearing impaired (including the elderly). Besides providing great improvement in speech recognition in noise, our set-up is also intuitive to the user and thus requires minimal cognitive effort. In addition, it is relatively minimal in terms of the applied technical and time resources. We are working on reducing the size of the device even further. This contrasts with the more cumbersome solutions using tactile inputs that are available on the market [43, 66].

Interestingly, the participants of our study after audio-tactile training also improved when speech understanding was tested only through the auditory channel with the tactile stimulation removed. These findings represent a transfer of a short multisensory training not only to the novel multisensory stimuli but also to the unisensory modality (audio only). The reported scores for the auditory only and the AM conditions after training were also strongly correlated (not the case before training) which suggests common learning mechanisms. We were hoping to see such an effect, as this outcome has crucial implications for the development of rehabilitation programs for the hearing-impaired patients, including actual and future users of hearing aids and cochlear implants. Our in-house SSD and the multisensory training procedure can be potentially applied to a population of HA/CI users to help them progress in their auditory (i.e. unisensory) performance. A similar idea of a multisensory training boosting unisensory auditory and visual speech perception, was shown by Bernstein and colleagues [20] and by Eberhard and colleagues [25], respectively, although the applied language tasks were more basic than repeating whole sentences. At the same time, with future hard-of-hearing candidates for cochlear implantation, we believe that both unisensory tactile and multisensory audio-tactile training can be applied using our set-up, with the aim to “prepare” the auditory cortex for future processing of its natural sensory input [22, 67]. We believe that this can be achieved, based on a number of works from our lab which show that specialization of brain sensory regions, such as the (classically termed as) visual or the auditory cortex, can emerge also following a specifically designed training with inputs from another modality, which however preserve the computational features specific for a given sensory brain area. We have been referring to this type of brain organization in our works as Task-Selective and Sensory-Independent (TSSI) [21, 60].

A revised critical period theory

We show here, with our device, significant multisensory enhancement of speech-in-noise comprehension, at a level comparable or higher to that reported when auditory speech in noise is complemented with cues from lip/speech reading [5; 48, 49, 50, 51]. This is definitely an interesting finding, as synchronous audio-visual speech information is what we as humans are exposed to from the very early years of development and throughout lifetime. The brain networks for speech processing combining higher-order and sensory brain structures, often involving auditory and visual cortices, are also well established. At the same time, exposure to an audio-tactile speech input is an utterly novel experience. We argue, therefore, that one can establish in adulthood a new coupling between a given neuronal computation and an

atypical sensory-modality which was never used for encoding that type of information before. We also show that this coupling can be leveraged for improving performance through a tailored SSD-training. Quite interestingly, this can even be achieved for the very complex and dynamic signal, such as speech. The current study thus provides further evidence supporting our new conceptualization of critical/sensitive periods in development, as presented in our recent review paper [22]. Although, in line with the classical assumptions, we agree that brain plasticity spontaneously decreases with age, we also believe that it can be reignited across the lifespan, even with no exposure to certain unisensory or multisensory experiences during childhood. Several studies from our lab and other research groups, mainly involving patients with congenital blindness, as well as the current study, point to that direction [22, 68, 69, 70].

Summary, future applications and research directions

In summary, we show that understanding of speech in noise greatly improves after a short multisensory audio-tactile training with our in-house SSD. The results of the current experiment expand our previous findings where we showed a clear and immediate benefit of complementing an auditory speech signal with tactile vibrations on fingertips, with no training at all [28]. Our research and the specifically developed experimental set-up are indeed novel and contributes to the very scarce literature on audio-tactile speech comprehension.

We believe that development of assistive communication devices involving tactile cues is especially needed in the time of the COVID19 pandemics, which imposes numerous restrictions on live communication, including the limited access to visual speech cues from lip reading. Besides the discussed rehabilitation regimes for the hearing impaired and the elderly, our technology and the tactile feedback can also assist normal hearing individuals, in second language acquisition, improving appreciation of music, as well as when talking on the phone. Furthermore, our lab already started developing new minimal tactile devices that can provide vibrotactile stimulation on other body parts, beyond the fingertips. Our aim is to design a set-up that will be wearable and therefore will assist with speech-in-noise comprehension (and sound source localization) in real-life scenarios.

In addition, our current SSD is compatible with a 3T MRI scanner. We have already collected functional magnetic resonance (fMRI) data in a group of participants performing the same tasks of unisensory and multisensory speech comprehension. To our knowledge this study would be the first to look into the neural correlates of understanding speech presented as combined auditory and vibrotactile stimulation. Our results can help uncover the brain mechanisms of speech-related audio-tactile binding, and may elucidate the neuronal sources of inter-subject variability for what it concerns the benefits of multisensory learning. This latter aspect, in turn, could further direct rehabilitation and training programs. Future fMRI studies are foreseen in the deaf population in Israel that will investigate the neural correlates of perceiving a closed set of trained speech stimuli solely through vibration.

Declarations

Acknowledgements

This work was supported by the Polish National Science Center (grant MOBILNOSC PLUS, 1642/MOB/V/2017/0) awarded to K.C., the European Research Council Consolidator-Grant (773121) to A.A., the Joy Ventures grant awarded to A.A.

Contribution statements:

KC performed literature review, designed the study, collected the data, analysed the results, prepared the manuscript

TW designed the study, developed methods for data collection and presentation, reviewed the manuscript

AL designed the study, reviewed the manuscript

HS designed the study, reviewed the manuscript

AA designed the study, prepared the manuscript

Competing interests: The author(s) declare no competing interests.

References

1. Stevenson, R. *et al.* M. Multisensory integration in cochlear implant recipients. *Ear Hear.***38** (5), 521–538 (2017).
2. Bayard, C. *et al.* Cued Speech Enhances Speech-in-Noise Perception. *J Deaf Stud Deaf.***24** (3), 223–233 (2019).
3. Jaha, N., Shen, S., Kerlin, J. R. & Shahin, A. J. Visual Enhancement of Relevant Speech in a 'Cocktail Party'. *Multisens. Res.***33** (3), 277–294 (2020).)
4. Peele, J. E. & Sommers, M. S. Prediction and constraint in audiovisual speech perception. *Cortex.***68**, 169–181 (2015).
5. van de Rijt, L. P. H., Roye, A., Mylanus, E. A. M., van Opstal, A. J. & van Wanrooij, M. M. The principle of inverse effectiveness in audiovisual speech perception. *Front. Hum. Neurosci.* (2019).
6. Leybaert, J. & LaSasso, C. Cued Speech for Enhancing Speech Perception and First Language Development of Children with Cochlear Implants. *Trends Amp.***14** (2), 96–112 (2010).)
7. Knowland, V. C. P., Evans, S., Snell, C. & Rosen, S. Visual Speech Perception in Children with Language Learning Impairments. *J Speech Lang Hear Res.***59** (1), 1–14 (2016).
8. Bidelman, G. M., Brown, B., Mankel, K. & Nelms, C. Psychobiological Responses Reveal Audiovisual Noise Differentially Challenges Speech Recognition. *Ear Hear.***41** (2), 268–277 (2020).
9. Lalonde, K. & Werner, L. A. Development of the Mechanisms Underlying Audiovisual Speech Perception Benefit. *Brain Sci.***11** (1), 49 (2021).

10. Huber, M. *et al.* Cognition in older adults with severe to profound sensorineural hearing loss compared to peers with normal hearing for age. *Int J Audiol.***59** (4), 254–262 (2020).
11. Livingston, G. *et al.* Dementia prevention, intervention, and care. *Lancet.***390** (10113), 2673–2734 (2017).
12. Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S. & Lawlor, B. A. Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis. *JAMA Otolaryngol Head Neck Surg.***144** (2), 115–126 (2018).
13. Schneider, B., Pichora-Fuller, M. K. & Daneman, M. Effects of senescent changes in audition and cognition on spoken language comprehension. In *The aging auditory system* (eds. Gordon-Salant, S., Frisina, R.D., Fay, R. & Popper, A.) 167–210, (New York:Springer-Verlag2010).
14. Agrawal, Y., Platz, E. A. & Niparko, J. K. Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999–2004. *Arch Intern Med.***168** (14), 1522–1530 (2008).
15. Imam, L. & Hannan, S. A. Noise-induced hearing loss: a modern epidemic? *Br J Hosp Med (Lond).***78** (5), 286–290 (2017).
16. Davis, A. *et al.* Aging and Hearing Health: The Life-course Approach. *Gerontologist.***56** (2), S256–67 (2016).
17. Tinga, A. M. *et al.* Multisensory stimulation to improve low-and higher-level sensory deficits after stroke: a systematic review. *Neuropsych. Rev.***26** (1), 73–91 (2016).
18. Keller, I. & Lefin-Rank, G. Improvement of visual search after audiovisual exploration training in hemianopic patients. *Neurorehabil. Neural Repair.***24** (7), 666–673 (2010).
19. Bolognini, N. *et al.* Multisensory integration in hemianopia and unilateral spatial neglect: Evidence from the sound induced flash illusion. *Neuroreport.***18** (10), 1077–1081 (2017).
20. Eberhardt, S., Auer, E. T. & Bernstein, L. E. Multisensory Training can Promote or Impede Visual Perceptual Learning of Speech Stimuli: Visual-Tactile versus Visual-Auditory Training. *Front. Hum. Neurosci.***8**, 829 (2014).
21. Heimler, B., Striem-Amit, E. & Amedi, A. Origins of task-specific sensory-independent organization in the visual and auditory brain: Neuroscience evidence, open questions and clinical implications. *Cur. Opin. Neurobiol.***35**, 169–172 (2015).
22. Heimler, B. & Amedi, A. Task-selectivity in the sensory deprived brain and sensory substitution approaches for clinical practice; In: *Multis. Percept.* 321–342(2020).
23. Shams, L. & Seitz, A. Benefits of multisensory learning. *Trends Cogn Sci.***12** (11), 411–417 (2008).
24. Murray, M.M., Wallace, M.T. (editors). *The Neural Bases of Multisensory Processes* (Boca Raton (FL):CRC Press/Taylor & Francis2012).
25. Bernstein, L. E., Auer, E. T., Jr, Eberhardt, S. P. & Jiang, J. Auditory perceptual learning for speech perception can be enhanced by audiovisual training. *Front. Neurosci.***7**, 34 (2013).

26. Fletcher, M. D., Mills, S. R. & Goehring, T. Vibro-Tactile Enhancement of Speech Intelligibility in Multi-talker Noise for Simulated Cochlear Implant Listening. *Trends Hear.*,22(2018).
27. Huang, J., Sheffield, B., Lin, P. & Zeng, F. Electro-Tactile Stimulation Enhances Cochlear Implant Speech Recognition in Noise. *Sci Rep.***7** (1), 2196 (2017).
28. Cieřła, K. *et al.* Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution. *Restor. Neurol. Neurosci.***37**, 155–166 (2019).
29. Levänen, S., Jousmäki, V. & Hari, R. Vibration-induced auditory-cortex activation in a congenitally deaf adult. *Curr Biol.***8** (15), 869–872 (1998).
30. Auer, E. T., Bernstein, L. E., Sungkarat, W. & Sing, M. Vibrotactile Activation of the Auditory Cortices in Deaf versus Hearing Adults. *Neuroreport.***18** (7), 645–648 (2007).)
31. Good, A., Reed, M. J. & Russo, F. A. Compensatory Plasticity in the Deaf Brain: Effects on Perception of Music. *Brain Sci.***4**, 560–574 (2014).
32. Soto-Faraco, S. & Deco, G. Multisensory contributions to the perception of vibrotactile events. *Behav. Brain Res.***196** (2), 145–154 (2009).
33. Young, G. W., Murphy, D. & Weeter, J. Haptics in Music: The Effects of Vibrotactile Stimulus in Low Frequency Auditory Difference Detection Tasks. *IEEE Trans on Haptics.***99**, 1 (2016).
34. Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L. & De Volder, A. G. Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. *Eur J Neurosci.***45** (11), 1439–1450 (2017).
35. Caetano, G. & Jousmaki, V. Evidence of vibrotactile input to human auditory cortex. *Neuroimage.***29** (1), 15–28 (2006).
36. Abboud, S. *et al.* Introducing a “visual” colorful experience for the blind using auditory sensory substitution. *Restor. Neurol. Neurosci.***32** (2), 247–257 (2014).
37. Bach-y-Rita, P. Tactile sensory substitution studies. *Ann N Y Acad Sci.***1013**, 83–91 (2004).
38. Meijer, P. An experimental system for auditory image representations. *IEEE Trans Biomed Eng.***39** (2), 112–121 (1992).)
39. Bach, Y., Rita, P., Collins, C. C., Saunders, F. A. & White, B. Scaddem, L. Vision Substitution by Tactile Image Projection. *Nature.***221**, 963–964 (1969).)
40. Striem-Amit, E., Guendelman, M. & Amedi, A. ‘Visual’ acuity of the congenitally blind using visual-to-auditory sensory substitution. *PloS one.***7** (3), (2012). e33136
41. Bubic, A., Striem-Amit, E. & Amedi, A. Large-Scale Brain Plasticity Following Blindness and the Use of Sensory Substitution Devices. In: Multisensory Object Perception in the Primate Brain, 351–380(2010).
42. Maidenbaum, S., Chebat, D. R., Levy-Tzedek, S., Furstenberg, R. & Amedi, A. **The effect of expanded sensory range via the EyeCane sensory substitution device on the characteristics of visionless virtual navigation.***Multisens. Res.* **27(5–6)** (2014).

43. Kupers, R. & Ptito, M. Cross-Modal Brain Plasticity in Congenital Blindness: Lessons from the Tongue Display Unit. *i-Perception*. 2, 748(2011).
44. Yamanaka, T., Hosoi, H. & Skinner, K. & Bach-y-Rita, P. Clinical application of sensory substitution for balance control. *Practica Oto-Rhino-Laryngol*.**102**, 527–538 (2009).
45. Nilsson, M., Soli, S. D. & Sullivan, J. A. Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J. Acoust. Soc. Am*.**95** (2), 1085–1099 (1994).
46. Levitt, H. Adaptive testing in audiology. *Scand. Audiol. Suppl*.**6** (6), 241–291 (1978).
47. Stevens, S. S. On the psychophysical law. *Psychol Review*.**64** (3), 153–181 (1957).
48. van Hoesel, R. J. M. Audio-visual speech intelligibility benefits with bilateral cochlear implants when talker location varies. *J Assoc Res Otolaryngol*.**16** (2), 309–315 (2015).
49. Blackburn, C. L., Kitterick, P. T., Jones, G., Summer, C. J. & Stacey, P. C. Visual Speech Benefit in Clear and Degraded Speech Depends on the Auditory Intelligibility of the Talker and the Number of Background Talkers. *Trends Hear*.**23**, (2019).
50. MacLeod, A. & Summerfield, Q. Quantifying the contribution of vision to speech perception in noise. *Br J Audiol*.**21** (2), 131–141 (1987).
51. Grange, J. A. & Culling, J. F. Head orientation benefit to speech intelligibility in noise for cochlear implant users and in realistic listening conditions. *J Acoust Soc A*.**140** (6), 4061 (2016).
52. Guediche, S., Blumstein, S. E. & Fiez, J. A. Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Front Syst Neurosci*.**7**, 126 (2013).
53. Fu, Q-J. & Galvin, J. J. 3 Perceptual learning and auditory training in cochlear implant recipients. *Trends Amplif*.**11** (3), 193–205 (2007).
54. Hervais-Adelman, A. G., Davis, M. H., Johnsrude, I. S. & Taylor, K. J. Generalization of Perceptual Learning of Vcoded Speech November. *J Exp Psychol Hum Percept Perform*.**37** (1), 283–295 (2010).)
55. Otto, T. U., Dassy, B. & Mamassian, P. Principles of Multisensory. *Behavior J Neurosci*.**33** (17), 7463–7474 (2013).
56. Meredith, M. A. & Stein, B. E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J Neurophysiol*.**56** (3), 640–662 (1986).
57. Holmes, N. P. The law of inverse effectiveness in neurons and behaviour: multisensory integration versus normal variability. *Neuropsych*.**45** (14), 3340–3345 (2007).
58. Chebat, D. R., Schneider, F. C., Kupers, R. & Ptito, M. Navigation with a sensory substitution device in congenitally blind individuals. *Neuroreport*.**22** (7), 342–347 (2011).
59. Chebat, D. R., Maidenbaum, S. & Amedi, A. Navigation using sensory substitution in real and virtual mazes. *PloS one*.**10** (6), (2015). e0126307
60. Amedi, A., Hofstetter, S., Maidenbaum, S. & Heimler, B. Task Selectivity as a Comprehensive Principle for Brain Organization. *Trends Cogn. Sci*.**21** (5), 307–310 (2017).

61. Hamilton-Fletcher, G., Wright, T. D. & Ward, J. Cross-Modal Correspondences Enhance Performance on a Colour-to-Sound Sensory Substitution Device. *Multis. Res.***29** (4–5), 337–363 (2014).
62. Johnson, K. O. The roles and functions of cutaneous mechanoreceptors. *Cur Opin Neurobiol.***11**, 455–461 (2001).
63. Ding, N., Chatterjee, M. & Simon, J. Z. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure.. *Neuroimage***88**, 41–46 (2014).
64. Moore, B. The Role of Temporal Fine Structure Processing in Pitch Perception, Masking, and Speech Perception for Normal-Hearing and Hearing-Impaired People. *JARO***9**, 399–406 (2008).
65. Moon, J. & Hong, S. H. What Is Temporal Fine Structure. Why Is It Important? *Korean J Audiol.***18** (1), 1–7 (2014).
66. Novich, S. D. & Eagleman, D. M. Using space and time to encode vibrotactile information: toward an estimate of the skin's achievable throughput. *Exp. Brain Res.***233** (10), 2777–2788 (2015).
67. Heimler, B., Pavani, F. & Amedi, A. Implications of cross-modal and intra-modal plasticity for the education and rehabilitation of deaf children and adults. In: Evidence-Based Practices in Deaf Education (eds. Knoors, H. & Marschark, M) (Oxford University Press 2018).
68. Chebat, D. R., Schneider, F. C. & Ptito, M. S. Competence and Brain Plasticity in Congenital Blindness via Sensory Substitution Devices. *Front Neurosci.***14**, 8–15 (2020).
69. Jicol, C. *et al.* Efficiency of Sensory Substitution Devices Alone and in Combination With Self-Motion for Spatial Navigation in Sighted and Visually Impaired. *Front Psychol.***11**, 1443 (2020).
70. Amedi, A. *et al.* Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nat. Neurosci.***10**, 687–689 (2007).

Figures

Figure 1

The AM test condition - outcomes before and after AM training, (A) at the group level, (B) in single participants; SRT – Speech Reception Threshold; PRE – test performed before training, POST – test performed after training; the bars are error bars; participant no 15 had an SRT = 0 dB in the POST training AM test.

Figure 2

The AnM test condition - outcomes before and after AM training, (A) at the group level, (B) in single participants; SRT – Speech Reception Threshold; PRE – test performed before training, POST – test performed after training; the bars are error bars.