The Topo-Speech Algorithm: An intuitive Sensory Substitution for Spatial Information

Benedetta Heimler
Faculty of Medicine (IMRIC)
Hebrew University of Jerusalem
Jerusalem, Israel
benedettaheimler@gmail.com

Amir Shur
Faculties of Cognition & Computer Sciences
Hebrew University of Jerusalem
Jerusalem, Israel
amirshur1@gmail.com

Ophir Netzer
Faculties of Medicine (IMRIC) & Cognition
Hebrew University of Jerusalem
Jerusalem, Israel
ophir.netzer@gmail.com

Tomer Bechor
Renew Senses
Jerusalem, Israel
tomer@renewsenses.com

Amir Amedi
Faculties of Medicine (IMRIC), Brain (ELSC) & Cognition
Hebrew University of Jerusalem
Jerusalem, Israel
amira@ekmd.huji.ac.il

***Abstract*—Is it possible to quickly and reliably understand the position of objects in space without relying on vision? This is one of the biggest challenge blind people face daily. We developed a novel algorithm called the Topo-Speech which conveys the spatial position of objects via speech manipulations. We ran a pilot study on blindfolded sighted adults (n=5) to test the extent to which users can locate objects' spatial positions after short training with the Topo-Speech, as well as their ability to locate untrained spatial positions. Participants were trained for ~30 minutes on the detection of objects' positions on a 3x3 grid. Then they were tested on the same spatial locations (though using different stimuli). Finally, participants were tested on identifying the positions of objects on a 5x5 grid (i.e. additional spatial locations) without any specific training. Our results showed that participants performed significantly above chance for both trained and untrained spatial positions. This in turn suggests the feasibility of the Topo-Speech to convey spatial related information via a non-visual channel and prompt to quickly test such approach with people who are visually impaired.**

Keywords— sensory substitution, vision rehabilitation, spatial processing, visual impairment, blindness, accessibility

I. INTRODUCTION

Sighted people are constantly aware of all objects around them when exploring the external world through vision. Visually impaired people however, constantly miss plenty of information from their spatial surrounding, as they can perceive the spatial world mainly through touch, thus being limited to perceive only objects reachable with their hands. Audition can only help to localize objects that produce a distinct sound, which are not that many, beside the special case of echolocation that is anyway used only by a small percentage of blind and visually impaired people (i.e., producing mouth clicks and listening to the returning echoes to infer objects' locations and identities) [1]. Can we make the spatial locations of objects in a scene available to visually impaired people in a reliable way on a day-to-day basis? There are various apps that successfully convey objects' identitiy through speech (e.g., TapTapSee), but do not provide any information regarding their spatial locations. Another family of devices, called Sensory Substitution Devices (SSDs), aim at providing whole visual-scene information by transforming visual input into what are termed auditory soundscapes which preserve the shape of objects together with their spatial location [2]. And indeed when trained with SSDs blind and visually impaired people succeed in many "visual" tasks [3]. However, the learning of the SSDs algorithms take long and most training programs are focused on shape recognition rather than object location. Here we present a novel built-in algorithm called the Topo-Speech which conveys object identity through speech adding spatial information through sound manipulations, similarly to SSDs algorithms (e.g., vertical spatial locations are conveyed through pitch variations such that the lower the pitch, the lower the corresponding spatial position). We conducted a pilot study with blindfolded sighted participants conceived as a proof-of-concept for this algorithm efficacy. Our aims were to 1- test the Topo-Speech intuitivity by asking participants to localize different objects relying solely on our algorithm after short training (~30 minutes) and 2- test the generalizability of the learning to untrained spatial positions.

Positive results will proove the feasibility of the Topo-Speech to reliably convey spatial information through audition, thus suggesting new promising frontiers for visual and spatial rehabilitation.

II. METHODS

A. *Equipment*

We created two grids, a 3x3 and a 5x5 grid, on a whiteboard hanging on a wall. The two grids were identical in their overall size, even though their cell size differed. The borders of each cell were made of duct tape to enable blindfolded participants to feel the borders between cells.

Note: This is not the official copyright released version of the IEEE proceedings paper. When citing this paper, use the following format: Heimler B, Shur A, Netzer O, Bechor T and Amedi A, "The Topo-Speech Algorithm: An intuitive Sensory Substitution for Spatial Information", Proc. 13th Int'l Conf. on Virtual Rehab., WG Wright, S Subramanian, G Fluet, M Agmon, RM Proffitt, M Roberts (Eds), Tel Aviv, Israel, 21-24 July 2019.

Stimuli consisted of a pull of 60 Hebrew words: 20 words were used for training and 30 words were used for the two experiments with the 3x3 and the 5x5 grids. All words were highly frequent in the language, two syllables long, represented graspable objects and had little to none spatial bias regarding spatial position.

B. The Topo-Speech Algorithm

All the words were adapted to fit the Topo-Speech algorithm: Two different beeps were signaling the beginning and the end of each object presentation. The y-axis was conveyed through pitch manipulations (e.g., higher the pitch, higher the spatial position) and the x-axis was conveyed through time in a left-to-right direction (e.g., a word heard closer to the first beep would be located more to the left of the visual field). All the words were recorded by a professional singer and the pitch and delay were adjusted using Audacity software. The pitch was changed to one of 5 different pitches: Low C, Low F, Low A#, Middle D# and Middle G#. Pitches Low C, Low A# and Middle G# were used for the 3x3 grid experiment, while all the pitches were used for the 5x5 grid experiment. Each trial lasted 2 seconds in both experiments. However, the delays signaling the different stimuli locations in the x-axis were longer in the 3x3 than in the 5x5 grid.

C. Participants

Five sighted individuals, aged 28± 2 years (mean ± SD), participated in this study. All participants were naïve to the Topo-Speech algorithm as well as to any Sensory Substitution Device. Participants were blindfolded across the experimental procedure. This experiment was conducted in accordance with the Helsinki declaration.

D. Experimental design

The experiment was composed of three parts: training on the 3x3 grid, testing on the 3x3 grid, and finally testing generalization on the 5x5 grid, with no additional specific training. For all parts, participants were instructed to hear the stimuli and touch the cell on the board corresponding to stimulus location.

During training with the 3x3 grid (27 trials), participants could hear each word as many times as they liked and received a feedback after each trial.

During both experiments (3x3 with 90 trials and 5x5 with 150 trials), each word was played twice and no feedback on the responses was provided.

There were short breaks between each of the three parts.

III. Results

During the short training, we observed that, over time, participants needed less word repetitions to provide the correct response, ultimately assuring that the training program was effective. In the two tests with the 3x3 and 5x5 grids, the participants' average success rate was 75.33%±24% (mean±STD) and 31.47%±12%, respectively. Two Wilcoxon tests showed that both these results were significantly above the respective chance level ($p = 0.0313$ for both experiments). We also analyzed the pattern of errors to investigate whether participants improved their understanding of the algorithm over experimental time. For both tests, we divided the experiment in bins of 10 trials each, and we calculated the number of errors for each bin. From a visualization of the results, we observed that in the 3x3 experiment participants made most errors in the first bin, with the number of errors decreasing over bins. We did not observe such a pattern for the 5x5 grid experiment, where participants made comparable number of errors across the experiment. Finally, we checked whether the learning of the Topo-Speech algorithm was easier for one of its two dimensions: x-axis (i.e., delay variations) or y-axis (i.e., pitch variations). Wilcoxon tests showed that there was no significant difference between the learning of the algorithm dimensions in etiehr the 3x3 ($p=0.44$) nor the 5x5 grids ($p=1$).

IV. Discussion

Our findings suggest that with short training, the Topo-Speech algorithm can successfully convey the spatial position of objects and that users can generalize their learning to untrained spatial positions. Additional data will indicate whether the lower success rate in the 5x5 experiment compared to the 3x3 experiment is due to general tiredness of participants at the end of the experimental procedure, or to some limits in the generalization process, thus prompting to the need for additional specific training.

Our next step is to test the efficacy of this algorithm in blind and visually-impaired people. We are also working towards adding features to the Topo-Speech algorithm, e.g., conveying depth through volume manipulations. Finally, to promote the daily-use of this algorithm, we are developing a dedicated app, integrating our algorithm with cutting-edge deep learning methods for object-recognition.

Acknowledgment

This work was supported by a European Research Council Consolidator-Grant (773121), a James S. McDonnel Foundation scholar award (no. 652 220020284).

References

[1] L. Thaler, S. R. Arnott, and M. A. Goodale, "Neural correlates of natural human echolocation in early and late blind echolocation experts," *PLoS One*, vol. 6, no. 5, 2011.

[2] P. B. L. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, pp. 112–121, 1992.

[3] S. Maidenbaum, S. Abboud, and A. Amedi, "Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation," *Neurosci. Biobehav. Rev.*, vol. 41, pp. 3–15, 2014.