

# 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 Behor  
Renew Senses  
Tel Aviv, Israel  
tom@renewsenses.com

Amir Amedi  
Faculties of Medicine (IMRIC),  
& 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 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

How can we enrich the amount of information that visually impaired people can gather from the environment? To provide you with an idea of the limits of spatial exploration without vision, imagine you are having dinner with friends and one of them asks you: “Can you please pass me the salt? It is behind the bottle”. As a sighted person, this action can be performed very quickly, because through vision, we are constantly aware of our surrounding and of the spatial relations between objects around us. However, for a visually impaired or blind person, this action may be very hard, if not impossible. To help the visually impaired population to gather more information from their surroundings, ultimately increasing their independent interactions with the environment, various apps successfully convey objects' identity through speech (e.g., TapTapSee).

Such approaches are very intuitive, but do not provide any information regarding the spatial locations of objects in a scene. 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 [1]. And indeed when trained with SSDs blind and visually impaired people succeed in many “visual” tasks [2]. 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 combines advantages of both aforementioned types of devices, conveying objects identity through speech and objects locations through manipulations of sounds features (e.g., pitch variations), similarly to SSD algorithms. We conducted a pilot study with blindfolded sighted participants conceived as a proof-of-concept for the efficacy of this algorithm, testing both its intuitivity and the generalization of learning to untrained spatial locations.

Positive results will suggest new promising frontiers for visual and spatial rehabilitation as well as for sensory augmentation.

## II. METHODS

### A. The Topo-Speech Algorithm

The final goal of the Topo-Speech algorithm is to allow the visually impaired population to fully experience their spatial surrounding. The aim is to develop an app that uses cutting-edge deep learning methods for object-recognition in real-time, allowing both to recognize the objects in a scene and to name them, and then to combine this output with an SSD-like algorithm which modifies the speech signal to convey spatial locations. Specifically, using the camera on the phone, the Topo-Speech app scans a visual scene from left-to-right using

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

a sweep-line algorithm. The location of objects in the x-axis is mapped to time, such that if one hears first the word “book” and then the word “apple”, it means that the book is located more to the left of the visual field. The location of objects in the y-axis is mapped to pitch variations, such that the lower one hears the word “book” in terms of pitch, the more the book will be located lower in the visual field.

For the purpose of this proof-of concept study, the words were recorded by a professional singer and further modified with Audacity software (see Equipment section for details). In each trial only one word was presented.

### B. 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. Stimuli consisted of a pull of 60 Hebrew words: 20 words were used for training and 30 words were used for the two experiments (3x3 and 5x5 grids). All words were highly frequent in the language, two syllables long, represented graspable objects and had little to no spatial position bias (e.g., “matzit” (lighter), “gezer” (carrot)). Each word was modified as follows: we added two beeps signaling the beginning and the end of each word presentation, thus defining the limits of the space in the x-axis. Therefore, for instance, a word presented after a short delay from the first beep, would be located more to the left of the board. We recorded each word with 5 different pitches to convey the spatial location in the y-axis: 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 and consisted of a beep signaling the start of the trial, a word played with a certain pitch and a certain delay from the first beep conveying the corresponding object’s spatial location for the y and the x axes respectively, and a final beep signaling the end of the trial.

### C. Participants

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

### D. Experimental design

Participants started with training on the 3x3 grid, then moved to the testing on the 3x3 grid, and finally to the testing on the 5x5 grid, with no additional specific training. Participants were always instructed to hear carefully to the stimuli, and to touch the cell on the board they thought corresponded to the 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. If they answered incorrectly, they heard the same word again.

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.

## III. RESULTS

During the short training, we observed that, over time, participants required fewer trial repetitions before giving 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  $75.33\% \pm 24\%$  (mean $\pm$ STD) and  $31.47\% \pm 12\%$ , respectively. Two Wilcoxon tests showed that both these results were significantly above the respective chance level ( $p = 0.0313$  for both experiments). We also 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 either 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, even though with a lower success rate than the trained locations. This latter result may prompt to the need for additional training.

Our next step is to test the efficacy of this algorithm in blind and visually-impaired people. We are also implementing new features in the Topo-Speech algorithm, e.g., conveying depth through volume manipulations or the x-axis positions through stereophonic sounds rather than through sounds delay. Our dedicated app, integrating our algorithm with deep learning methods for object-recognition is almost ready to be tested. We are also working towards embedding our algorithm in virtual environments to allow the visually impaired population to train in real life scenarios before venturing in the real world. We envision that all these developments together will allow the visually impaired population to experience their surrounding quickly, intuitively and comprehensively, ultimately increasing their independency in their interactions with the external world.

## REFERENCES

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