Auditory letter-name processing elicits crossmodal representations in blind listeners

Santani Teng (santani@mit.edu)
Smith-Kettlewell Eye Research Institute, San Francisco, CA 94115, USA
CSAIL, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Verena Sommer (vsommer@mpib-berlin.mpg.de)
Center for Lifespan Psychology, Max Planck Institute for Human Development
Lentzeallee 94, 14195 Berlin, Germany

Radoslaw Cichy (rmcichy@zedat.fu-berlin.de)
Department of Education and Psychology, Free University of Berlin
Habelschwerdter Allee 45, 14195 Berlin, Germany

Dimitrios Pantazis (pantazis@mit.edu)
McGovern Institute for Brain Research, Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139, USA

Aude Oliva (oliva@mit.edu)
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139, USA

Abstract:

As incoming stimuli travel through our sensory pipelines, they are processed in multiple formats, spanning levels of abstraction as well as sensory domains. The spatiotemporal and representational dynamics of this processing cascade remain especially unclear in nonvisual modalities and nontypical populations; in particular, a stimulus representation may either facilitate or suppress its multisensory analogue. Here, we presented auditory alphabetic letter names to blind listeners with no visual letter experience, as well as sighted listeners with no braille (tactile) reading experience. Blind and sighted groups' brain responses distinguished letter identity along similar time courses and accuracies. However, only blind listeners' brain signals correlated with a low-level model of braille characters, while no such correlation was found between sighted listeners' brain signals and a neural network model of low-level visual letter features. The results illustrate that visual experience modulates both the extent and nature of multisensory processing and object representation generally.

Keywords: audition; blindness; braille; MEG; MVPA; RSA

Background and Introduction

Alphabetic letters, a foundational element of literacy, constitute highly trained stimulus sets for their readers. Within a given language, letters also have robust mappings across modalities, with standardized visual shapes, auditory (spoken) names, and among blind braille readers, tactile forms. Presenting letters in one modality may induce representations that generalize across modalities and are contingent on the reader's experience (Rothlein & Rapp, 2014). A reader with sensory loss such as blindness would be isolated from visual letter representations, while a sighted reader would lack a tactile representation exclusive to braille readers, yet both would retain a similar auditory representation. In this way, understanding the neural dynamics of letter perception illuminates not only language-related processing, but also the multisensory representation of objects.

Here we combined multivariate pattern analysis (MVPA) and representational similarity analysis (RSA) to interrogate the neural response to heard letters. We show that while the overall time course of spoken letter processing is similar in blind and sighted listeners, representational dynamics suggest that crossmodal

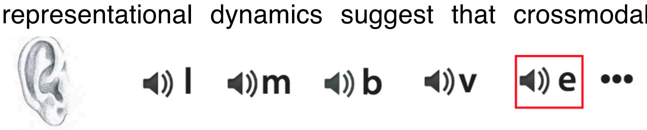

Figure 1. Auditory letter listening task during MEG recording. Spoken letter names were presented in random order, with two letters ('e' and 'o', indicated by red outline) as targets for button presses.

processing induced by the auditory stimuli is experience-dependent. The results suggest a critical role for visual experience in gating suppression vs. facilitation of a multisensory representation.

Methods

Participants

We presented blind (N=7) and sighted (N=7) participants with auditory letter-name stimuli during magnetoencephalography (MEG) recording. All participants reported normal hearing, were native/primary English speakers, and gave informed consent in accordance with MIT's Committee On the Use of Human Experimental Subjects. Blind participants were all proficient braille readers and, from birth, were either totally blind or only had some light perception; none had any experience with visual print reading or spatial vision generally. See Table I for details.

Table 1: Blind participants.

ID	Age (y)	M/F	Vis. impairment
1	34	F	Total
2	26	F	LP
3	29	F	Total
4	28	M	Total
5	25	M	Total
6	24	F	LP
7	27	M	LP

Table 1. Clinical details of blind participants. N=7, 4 females, mean age 27.6 y (SD 3.3 y). Blindness onset was congenital (age 0) for all participants. LP = light perception; no spatial vision.

Experimental Design

Stimuli. Stimuli were audio recordings of twelve spoken letter names of the American English alphabet (B, C, D, E, L, M, N, O, V, X, Y, Z), with E and O as target letters for an oddball detection task. We used a subset of the alphabet to increase power by allowing more trials per letter. The mean duration of letter presentations was 432 ms, not including the target letters, with durations of 342 and 360 ms, respectively. Stimuli were presented through earphones at a comfortable volume, approximately 70 dB SPL.

Procedure. Stimuli were presented in random order, with targets occurring every four trials on average. The task was to respond to each target letter with a button press; target stimulus trials were then excluded from further analyses (Fig. 1). Each of the ten non-target letters appeared a total of ~120 times per experiment. Stimulus onset asynchrony (SOA) was 1000 to 1100 ms after a non-target letter, and 2000 ms after a target letter.

MEG data acquisition and preprocessing. We acquired continuous MEG signals from 306 channels at 1000 Hz, filtered between 0.03 and 330 Hz. Brainstorm and custom MATLAB code were used to extract MEG trials from 200 ms before to 1000 ms after stimulus onset; we removed the baseline mean of each channel and applied a low-pass filter at 30 Hz.

Data Analysis

Multivariate pattern analysis. MVPA was conducted to distinguish between each pair of letter conditions. At each time point in the trial epoch, MEG pattern vectors were subaveraged across trials, whitened with multivariate noise normalization (Guggenmos et al., 2018), and then used to train a linear support vector machine to decode letter identity. Accuracies were computed pairwise via leave-one-out cross-validation, then averaged to assign an overall letter decoding accuracy at each time point (Fig. 2) or to populate a representational dissimilarity matrix (RDM) for model fitting (Fig. 3).

Model fitting. Within the RSA framework, we operationalized computational models of letter representation as RDMs of hypothesized similarity structures. For each group we hypothesized low-level representations of letters in the respective "primary" reading modality, i.e. vision and touch.

We modeled visual representations by extracting feature vectors of visual printed letters from the first two layers of a pretrained deep neural network (DNN; AlexNet; Krizhevsky et al., 2012), an architecture whose early layers are known to mimic early visual brain responses. To model tactile braille representations, we extracted simulated response rate vectors to braille character stimulation of D2d (the index finger pad) using the TouchSim MATLAB toolbox, based on modeled afferent nerve recordings (Saal et al., 2017). Note that these models only capture sensory-specific, and not higher-level linguistic, letter information.

Visual and tactile model RDMs were computed by determining the pairwise correlation distances (1–*r*) of the resulting feature or rate vectors. Finally, at each time point of the MEG decoding time course, we evaluated each model fit via Spearman correlation.

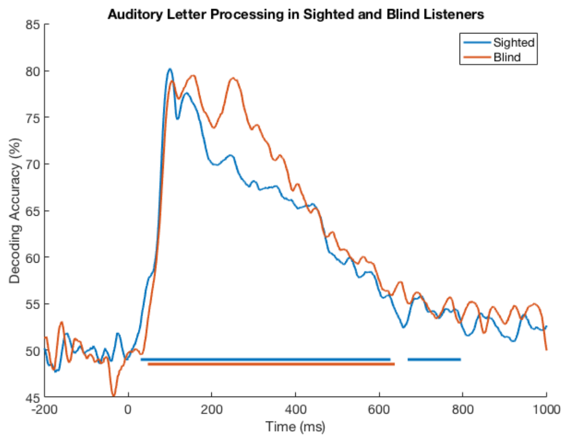

Figure 2. Time course of auditory letter identity decoding in sighted (blue) and blind (orange) listeners. Significant letter identity decoding emerged at 30 (10-74) ms for sighted readers and 47 (37-60) ms for blind readers.

Sensorwise analyses. The decoding and model fitting analyses above were repeated in sensor space, using 3-element MEG pattern vectors (from sensor triplets in 102 positions) rather than whole-brain data from all 306 sensors. The 102 time courses are thus roughly localized to their sensor triplet locations.

Statistical inference. To evaluate statistical significance of decoding and model correlation time courses, we used nonparametric statistical tests, avoiding assumptions of normality. Time series were subjected to permutation-based cluster-size inference (1000 permutations, 0.05 cluster definition threshold and 0.05 cluster threshold) against null hypotheses of 50% decoding accuracy or zero correlation. Significant clusters are indicated by color-matched bars in the respective time courses of Figs. 2 and 3. Intervals reported in parentheses after onsets and peaks represent 95% confidence intervals computed via bootstrapping the sample 1000 times.

Results

Both groups performed well on the behavioral task, and auditory letter identity was readily decodable from the MEG signal in both groups of listeners (Fig. 1). Both groups exhibited similar onset trajectories of the emerging letter identity signal, with onsets and bootstrapped 95% confidence intervals at 30 (10-74) ms for the sighted group and 47 (37-60) ms for the blind group. Decoding peaks were at similar accuracies for both groups, with latencies of 100 (96-141) ms for the sighted and 156 (97-259) ms for the blind group. The large bootstrapped confidence interval for the blind group reflects a sustained peak signal in the blind group's decoding curve that is evident in Fig. 1.

Model correlations

The tactile braille model correlated significantly with the MEG response in blind listeners (Fig. 3), reaching significance at ~300 ms and sustained for the rest of the trial. The model did not correlate significantly with the sighted group. The visual DNN RDM did not correlate significantly with either group's time course.

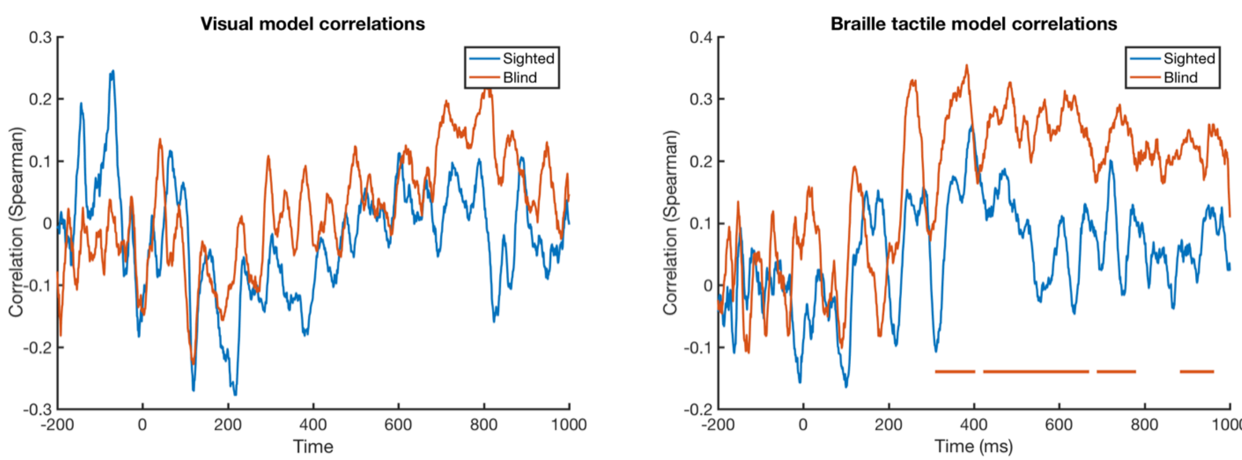

Figure 3. Visual (left) and braille tactile (right) model correlations in sighted (blue) and blind (orange) listeners. The visual DNN RDM comprised feature vectors of visual printed letter images from the first two convolutional layers of a pretrained DNN. The tactile model contained rate vectors from simulated afferent responses to braille letter stimulation on the index finger pad. Onset ~300 ms in the blind braille correlation.

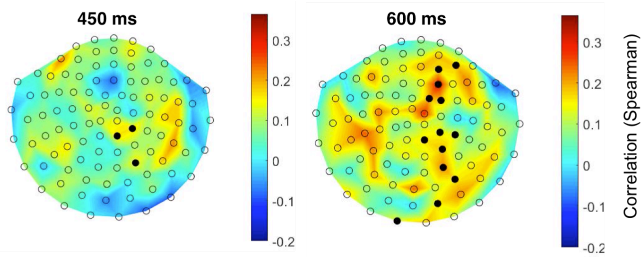

Figure 2. Spatial distribution of MEG-tactile model correlations in blind listeners at t=450ms (top) and 600ms (bottom). Sensor map shows top-down view of head, facing "up" on the page. Black dots indicate sensor positions with significant model fits.

Sensorwise analyses

Correlating the tactile model to sensor-specific MEG data revealed correlations reaching significance at ~450 ms in the blind group, originating at right central-parietal sensors, and largely right-lateralized thereafter (Fig. 4).

Summary and Conclusions

In presenting blind and sighted listeners with spoken alphabetic letter names, we found that the overall decoding time courses of letter perception were similar, but that distinct representations could be identified within the similarity structure of the MEG responses.

Despite the complete absence of a tactile component to the experimental paradigm, listening to spoken letter names elicited a representation in blind listeners consistent with tactile braille signals. An analogous DNN-based model of low-level visual letter features did not correlate significantly with the sighted listeners' MEG signal. Thus, identical auditory stimuli presented to both groups of participants elicited group-specific crossmodal representations. This may result from early-blind listeners' speech-elicited activation of visual cortex, also involved in braille processing (Röder et al., 2002).

The visual DNN correlation time courses did not reach significance in either group. While we expected this null result in blind participants with no visual reading experience, in sighted listeners such a signal would indicate facilitation of multisensory representation. We cannot rule out this possibility, but our result may reflect crossmodal suppression in the sighted group, consistent with prior work showing deactivated visual cortex during tactile tasks (Sadato et al., 1996) and impaired visual processing via auditory interference (Hidaka & Ide, 2015).

Finally, the spatial distribution of the representations is complex and invites further study. Interestingly, however, the tactile model reached significance at about 450 ms in central parietal sensor locations — the same location where one would expect early somatosensory responses when actually presenting tactile stimuli. This result suggests that crossmodal letter representations are referred back to the relevant early sensory areas.

Acknowledgments

This work was supported by the Vannevar Bush Faculty Fellowship program by the ONR to A.O. (N00014-16-1-3116) and grant CI 241-1/1 to R.C., and conducted at the Athinoula A. Martinos Imaging Center, McGovern Institute for Brain Research, MIT.

References

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434-447.

Hidaka, S., & Ide, M. (2015). Sound can suppress visual perception. *Scientific reports*, *5*, 10483.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).

Röder, B., Stock, O., Bien, S., Neville, H., & Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *European Journal of Neuroscience*, *16*(5), 930-936.

Rothlein, D., & Rapp, B. (2014). The similarity structure of distributed neural responses reveals the multiple representations of letters. *NeuroImage*, *89*, 331-344.

Saal, H. P., Delhaye, B. P., Rayhaun, B. C., & Bensmaia, S. J. (2017). Simulating tactile signals from the whole hand with millisecond precision. *Proceedings of the National Academy of Sciences*, *114*(28), E5693-E5702.

Sadato, N., Pascual-Leone, A., Grafman, J., Ibañez, V., Deiber, M.-P., Dold, G., Hallett, M. (1996). Activation of the primary visual cortex by Braille reading in blind subjects. *Nature*, *380*, 11.