Data Article

Functional magnetic resonance imaging data for the neural dynamics underlying the acquisition of distinct auditory categories

Zhenzhong Gan [a,b,c], Suiping Wang [a,*], Gangyi Feng [d,e,*]

[a] *Philosophy and Social Science Laboratory of Reading and Development in Children and Adolescents (South China Normal University), Ministry of Education, Guangzhou, Guangdong, China*
[b] *Guangdong Provincial Key Laboratory of Mental Health and Cognitive Science, South China Normal University, Guangzhou, Guangdong, China*
[c] *School of Psychology, South China Normal University, Guangzhou, Guangdong, China*
[d] *Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China*
[e] *Brain and Mind Institute, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China*

ARTICLE INFO

ABSTRACT

How people learn and represent auditory categories in the brain is a fundamental question in auditory neuroscience. Answering this question could provide insights into our understanding of the neurobiology of speech learning and perception. However, the neural mechanisms underlying auditory category learning are far from understood. We have revealed that the neural representations of auditory categories emerge during category training, and the type of category structures drives the emerging dynamics of the representations [1]. The dataset introduced here was derived from [1], where we collected to examine the neural dynamics underlying the acquisition of two distinct category structures: rule-based (RB) and information-integration (II) categories. Participants were trained to categorize these auditory categories with trial-by-trial corrective feedback. The functional magnetic resonance imaging (fMRI) technique was used to assess the neural dynamics related to the category learning process. Sixty adult Mandarin native speakers were recruited for the fMRI experiment. They were assigned to either the RB

($n = 30$, 19 females) or II ($n = 30$, 22 females) learning task. Each task consisted of six training blocks where each consisting of 40 trials. Spatiotemporal multivariate representational similarity analysis has been used to examine the emerging patterns of neural representations during learning [1]. This open-access dataset could potentially be reused to investigate a range of neural mechanisms (e.g., functional network organizations underlying learning of different structures of categories and neuromarkers associated with individual behavioral learning success) involved in auditory category learning.

Specifications Table

Subject	Biological Sciences, Neuroscience: Cognitive
Specific subject area	Cognitive Neuroscience, Neurobiology of Language, Behavioral Neuroimaging, Functional Magnetic Resonance Imaging, Auditory Category Learning
Type of data	Table and MRI images
How the data were acquired	MRI data were acquired using a Siemens 3T Tim Trio MRI system with a 12-channel head coil in the Brain Imaging Center at South China Normal University. The experiment was programmed, controlled, and recorded by E-Prime software (Psychology Software Tools, Inc., Sharpsburg, PA, USA).
Data format	Raw
Description of data collection	The fMRI images were acquired with a customized spare-sampling T2*weighted gradient echo-planar imaging (EPI) pulse sequence (TR = 2500 ms with 800-ms silent gap). High-resolution T1-weighted structural images were acquired using a magnetization-prepared rapid acquisition gradient-echo sequence (MP-RAGE) sequence.
Data source location	Institution: Brain and Mind Institute, The Chinese University of Hong Kong City/Town/Region: Hong Kong SAR Country: China
Data accessibility	Repository name: OpenNeuro Data identification number: 10.18112/openneuro.ds003764.v1.0.5 Direct URL to data: https://openneuro.org/datasets/ds003764/versions/1.0.3
Related research article	G. Feng, Z. Gan, H.G. Yi, S.W. Ell, C.L. Roark, S. Wang, P.C.M. Wong, B. Chandrasekaran, Neural dynamics underlying the acquisition of distinct auditory category structures, NeuroImage. 244 (2021) 118,565. https://doi.org/10.1016/j.neuroimage.2021.118565.

Value of the Data

- This MRI dataset can be used to examine theoretical hypotheses related to auditory category learning, especially testing the single vs. dual-learning systems hypothesis.
- The data revealed distinctive local neural encoding patterns of rule-based (RB) and information-integration (II) auditory categories during training. Large-scale network-level encoding mechanisms can be explored with this dataset.
- The data can be used to investigate the spatiotemporal neural dynamics underlying successful learning within and across sessions of categorization training.
- The data can be used to reveal functional network reorganizations underlying the two distinctive types of auditory category learning. Functional network organization can be assessed by measuring the changes in simple or effective inter-regional connectivity across training sessions.

- The data can be used to identify neuromarkers related to individual learning success and build brain-based learning prediction models to predict future learning behaviors. Significant individual differences in learning speed and outcome have been revealed by the behavioral analyses for each learning task. Potential neuromarkers linked to learners' behavioral performance could be identified with predictive modeling approaches [2]. These putative neuromarkers have the potential to be used to build generalizable predictive models to predict unseen learners' future learning behaviors [2–5].

1. Data Description

The data collected are the raw behavioral and neuroimaging data for our previous study focusing on examining the neural mechanisms underlying auditory category learning [1]. This data is open access in OpenNeuro (https://openneuro.org/datasets/ds003764/versions/1.0.5).

A folder named "source data" in the main folder contains the behavioral data in .xlsx format. This Excel file stored the experimental groups, types of stimuli, participants' trial-by-trial responses, and the timing information of each stimulus for each participant and fMRI run. Specifically, "Subject code" refers to the subject ID; "block" refers to the training block number; "structure" refers to the category structure ("ii" = information-integration structure; "rb" = rule-based structure); "trial" refers to the trial number; "sound" refers to the sound filename in each trial; "item" refers to the sound ID (each sound has a unique ID); "category" denotes the category labels; "stim_onset" denotes the onset time of each sound in a training block; "feedback_onset" denotes the onset time of each visual feedback in a training block; "stim.RESP" denotes subjects' categorization responses ("1, 2, 3, 4" represents category 1 to 4 respectively; "0" represents no response); "stim.ACC" represents the correctness of each categorization response ("1" = correct; "0" = incorrect); "stim.RT" represents the categorization response time in each trial; "handresp" represents the responding hand ("1" = right hand, "2" = left hand, "0" = no response); "tempMod" denotes the temporal modulation of each ripple sound; "specMod" denotes the spectral modulation frequency of each ripple sound. Sixty folders named after each participant's number (e.g., "sub-01") contain the neuroimaging data, including two types of MRI files: (1) a high-resolution T1-weighted image in subfolder "anat" and (2) six T2-weighted BOLD fMRI images for the six runs in subfolder "func." The neuroimaging data are in NifTi format.

## 2. Experimental Design, Materials and Methods

2.1. Participants

Sixty native Mandarin speakers (mean age: 21.2 years; SD = 2.1; age range: 18–28 years) were recruited from communities around South China Normal University for the MRI experiment. They were all right-handed, had normal or corrected-to-normal vision, and had minimal formal music training experience (< 1 year). Some participants can speak different dialects to certain degrees other than standard Mandarin. No participants reported any major psychiatric conditions, neurological or hearing disorders, head trauma, or use of psychoactive drugs or psychotropic medication. They were assigned to either RB or II group (n = 30 each). The two groups of participants were matched in age ($t_{(58)}$ = 0.249, P = 0.805), gender (II: 22 females; RB: 19 females; Chi-square = 0.693, P = 0.405), and years in musical training ($t_{(58)}$ = 1.644, P = 0.106). Participants received monetary compensation for their participation. All materials and protocols were approved by the ethics review board of the School of Psychology at the South China Normal University and the Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee. We obtained written informed consent for each participant before the experiment.

2.2. Stimuli

Ripple sounds were generated by modulating a white noise stimulus, varying along spectral and temporal modulation dimensions (duration = 0.5 s; digital sampling rate = 44.1 kHz; low-pass filtered at 4.8 kHz). Previous studies revealed that the spectral- and temporal-modulation dimensions are independent of each other at multiple levels of processing [6,7,8]. The spectral-modulation dimension ranged from 0.25 to 1.33 cyc/oct. The temporal-modulation dimension ranged from 4 to 10 Hz (Fig. 1A). These modulation frequencies were selected because they are strongly represented in the human auditory cortex [9] and reflect the natural spectro-temporal acoustic complexity of speech signals [9]. The amplitude of modulation depth was 30 dB. All sounds were synthesized in Matlab (Mathworks, Natick, MA, USA), and the RMS amplitude was normalized to 80 dB.

To create the RB category structure, we first generated 40 coordinates in an abstract normalized two-dimensional space, with a minimum value of 0 and the maximum value of 1. The coordinates were sampled from four bivariate normal distributions centered on (0.33, 0.33), (0.33, 0.68), (0.68, 0.33), and (0.68, 0.68), with a standard deviation of 0.1 for both dimensions. Values along each dimension were logarithmically mapped onto spectral and temporal modulation frequencies. Thereby, the optimal decision boundaries for the RB category structure reflect the placement of a decision criterion along the spectral modulation dimension at 0.58 cyc/oct and along the temporal modulation dimension at 6.32 Hz (Fig. 1A left panel, dashed lines). We rotated the RB structure by 45° counterclockwise to create the II category structure. Thus, the optimal decision boundaries for the II category structure rely on integrating information from both spectral and temporal modulation dimensions and are hard to describe verbally (Fig. 1A right panel, dashed lines).
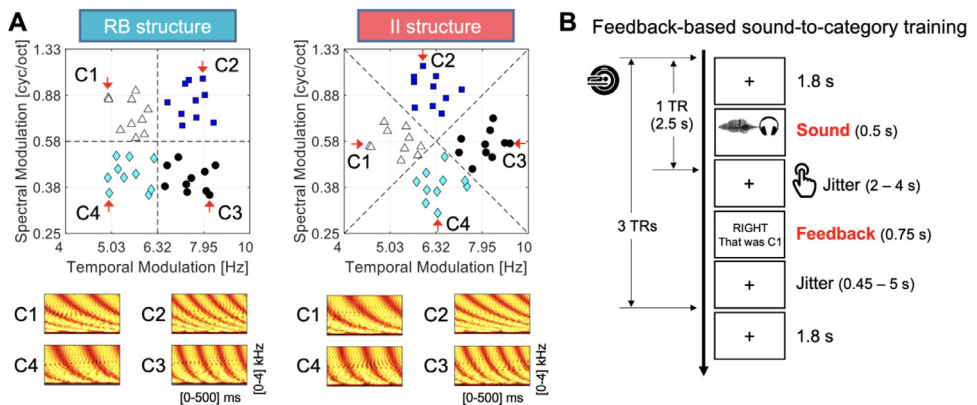
Fig. 1. Illustration of the experimental design. Two types of auditory category structures and the training paradigm adopted in the scanner. A, RB and II category structures and spectrograms of sample sounds. *Left panel*: forty sounds of the RB structure with optimal decision boundaries represented by dashed lines. Four categories (C1 – C4) of sounds are plotted in different shapes and colors in a perceptual space with spectral-by-temporal-modulation dimensions. *Right panel*: forty sounds of the II structure with optimal decision boundaries represented by dashed lines. Four categories (C1 – C4) of sounds are plotted in the same perceptual space. B, the feedback-based sound-to-category training procedure used in the fMRI experiment. TR = repetition time (2.5 s; 1700 ms acquisition time with 800 ms silence gap). Sounds were presented within the silence gap of a TR. Each training trial consisted of three TRs. This graph is modified from Fig. 1 in [1].

2.3. Sound-to-Category Training Procedure

Sounds were presented via MRI-compatible circumaural headphones in the MR scanner. Visual stimuli (e.g., feedback) were presented using an in-scanner projector visible using a mirror attached to the head coil. Participants responded with a two-button response box in each hand. Participants were instructed to attend to the fixation cross on the screen at the beginning (10 s) and the end of each training block (20 s). On each trial, a sound was presented during the 800-ms silence gap following the 1700 ms image acquisition (Fig. 1B; also see Imaging acquisition section for the customized sparse-sampling imaging parameters). This sparse scanning protocol and stimulus presentation procedure minimize the influence of MRI noises on auditory processing. Participants were asked to categorize the sounds into one of four categories. Informational, corrective feedback ("正确, 这是类别 1." ["RIGHT, that was C1."] or "错误, 这是类别 1." ["WRONG, that was C1."]) was displayed for 750 ms after each categorization response. If the participant failed to respond within two seconds, cautionary feedback was presented ("没有反应" ["NO RE-SPONSE"]). A fixation cross was displayed afterward until the onset of the next imaging acquisition. To separate sound- and feedback-evoked BOLD responses, jittered stimulus-to-feedback intervals sampling from a uniform distribution of 2 to 4 s were added. The intertrial intervals after the feedback presentation were also jittered (from 0.45 to 5 s). Each trial lasted for three TRs (7.5 s in total). Ten silence trials with only a fixation cross were randomly inserted between sound trials to jitter the inter-trial intervals in each training block to better estimate single-trial brain activation. The experiment consisted of six training blocks. Each sound was presented once within a block. The order of sound presentation was randomized for each block and participant. Therefore, each sound was repeated six times across six training blocks. Participants were asked to conduct a practice to familiarize the with finger-button mapping before the fMRI experiment.

2.4. Imaging Acquisition

MRI data were acquired using a Siemens 3T Tim Trio MRI system with a 12-channel head coil in the Brain Imaging Center at South China Normal University. Functional data were acquired using a customized spare-sampling T2*weighted gradient echo-planar imaging (EPI) pulse sequence (repetition time (TR) = 2500 ms with 800-ms silent gap; TE = 30 ms; Flip angle = 90°, 31 slices, field of view= 224 × 224 mm^2, in-plane resolution = 3.5 × 3.5 mm^2, slice thickness = 3.5 mm with 1.1 mm gap). T1-weighted high-resolution structural images were acquired using a magnetization-prepared rapid acquisition gradient-echo sequence (MP-RAGE) sequence (176 slices, TR = 1900 ms, TE = 2.53 ms, flip angle = 9º, voxel size = 1 × 1 × 1 mm^3).

Ethics Statement

All materials and protocols were approved by the ethics review board of the School of Psychology at the South China Normal University (approval numbers: SCNU-PSY-353 and SCNU-PSY-2021–159) and the Joint Chinese University of Hong Kong – New Territories East Cluster Clinical Research Ethics Committee (approval number: CREC 2018.432). Written informed consent was obtained before the experiment. Data collection was conducted according to the Declaration of Helsinki.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Functional magnetic resonance imaging data for the neural dynamics underlying the acquisition of distinct auditory categories (Original data) (OpenNeuro).

CRediT Author Statement

Zhenzhong Gan: Data curation, Writing – original draft; **Suiping Wang:** Project administration, Funding acquisition; **Gangyi Feng:** Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

References

[1] G. Feng, Z. Gan, H.G. Yi, S.W. Ell, C.L. Roark, S. Wang, P.C.M. Wong, B. Chandrasekaran, Neural dynamics underlying the acquisition of distinct auditory category structures, Neuroimage 244 (2021) 118565, doi:10.1016/j.neuroimage.2021.118565.

[2] D. Meng, S. Wang, P.C.M. Wong, G. Feng, Generalizable predictive modeling of semantic processing ability from functional brain connectivity, Hum. Brain Mapp. (2022) hbm.25953, doi:10.1002/hbm.25953.

[3] G. Feng, E.M. Ingvalson, T.M. Grieco-Calub, M.Y. Roberts, M.E. Ryan, P. Birmingham, D. Burrowes, N.M. Young, P.C.M. Wong, Neural preservation underlies speech improvement from auditory deprivation in young cochlear implant recipients, Proc. Natl. Acad. Sci. U. S. A. 115 (2018), doi:10.1073/pnas.1717603115.

[4] G. Feng, J. Ou, Z. Gan, X. Jia, D. Meng, S. Wang, P.C.M. Wong, Neural fingerprints underlying individual language learning profiles, J. Neurosci. (2021) 7372–7387, doi:10.1523/JNEUROSCI.0415-21.2021.

[5] G. Feng, Y. Li, S.M. Hsu, P.C.M. Wong, T.L. Chou, B. Chandrasekaran, Emerging native-similar neural representations underlie non-native speech category learning success, Neurobiol. Lang. 2 (2021) 280–307, doi:10.1162/nol_a_00035.

[6] D.A. Depireux, J.Z. Simon, D.J. Klein, S.A. Shamma, Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex, J. Neurophysiol. 85 (3) (2001) 1220–1234, doi:10.1152/jn.2001.85.3.1220.

[7] D.R.M. Langers, W.H. Backes, P.V. Dijk, Spectrotemporal features of the auditory cortex: the activation in response to dynamic ripples, Neuroimage 20 (1) (2003) 265–275, doi:10.1016/S1053-8119(03)00258-1.

[8] M. Schönwiesner, R.J. Zatorre, Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 14611–14616, doi:10.1073/pnas.0907682106.

[9] T.M. Elliott, F.E. Theunissen, The modulation transfer function for speech intelligibility, PLoS Comput. Biol. 5 (2009) e1000302, doi:10.1371/journal.pcbi.1000302.