

Memorization-Based Training and Testing Paradigm for Robust Vocal Identity Recognition in Expressive Speech Using Event-Related Potentials Analysis

Wenjun Chen¹, Xiaoming Jiang^{1,2}

¹Institute of Linguistics, Shanghai International Studies University ²Key Laboratory of Language Science and Multilingual Artificial Intelligence, Shanghai International Studies University

Corresponding Author

Xiaoming Jiang

xiaoming.jiang@shisu.edu.cn

Citation

Chen, W., Jiang, X. Memorization-Based Training and Testing Paradigm for Robust Vocal Identity Recognition in Expressive Speech Using Event-Related Potentials Analysis. *J. Vis. Exp.* (210), e66913, doi:10.3791/66913 (2024).

Date Published

August 9, 2024

DOI

10.3791/66913

URL

jove.com/video/66913

Abstract

Recognizing familiar speakers from vocal streams is a fundamental aspect of human verbal communication. However, it remains unclear how listeners can still discern the speaker's identity in expressive speech. This study develops a memorization-based individual speaker identity recognition approach and an accompanying electroencephalogram (EEG) data analysis pipeline, which monitors how listeners recognize familiar speakers and tell unfamiliar ones apart. EEG data captures online cognitive processes during new versus old speaker distinction based on voice, offering a real-time measure of brain activity, overcoming limits of reaction times and accuracy measurements. The paradigm comprises three steps: listeners establish associations between three voices and their names (training); listeners indicate the name corresponding to a voice from three candidates (checking); listeners distinguish between three old and three new speaker voices in a two-alternative forced-choice task (testing). The speech prosody in testing was either confident or doubtful. EEG data were collected using a 64-channel EEG system, followed by preprocessing and imported into RStudio for ERP and statistical analysis and MATLAB for brain topography. Results showed an enlarged late positive component (LPC) was elicited in the old-talker compared to the new-talker condition in the 400-850 ms window in the Pz and other wider range of electrodes in both prosodies. Yet, the old/new effect was robust in central and posterior electrodes for doubtful prosody perception, whereas the anterior, central, and posterior electrodes are for confident prosody condition. This study proposes that this experiment design can serve as a reference for investigating speaker-specific cue-binding effects in various scenarios (e.g., anaphoric expression) and pathologies in patients like phonagnosia.

Introduction

Human vocal streams are rich in information, such as emotion^{1,2}, health status^{3,4}, biological sex⁵, age⁶, and, more importantly, the individual vocal identity^{7,8}. Studies have suggested that human listeners have a robust capacity to recognize and differentiate their peers' identities through voices, overcoming within-speaker variations around speaker identity's average-based representation in the acoustic space⁹. Such variations are brought about by acoustic manipulation (fundamental frequency and vocal tract length, i.e., F0 and VTL) that corresponds to no clear pragmatic intentions⁹, emotion prosodies¹⁰, and vocal confidence that conveys speakers' feeling of knowing¹¹. Behavioral experiments have focused on many factors that influence listeners' performance in recognizing the talkers, including language-related manipulations^{8,12,13}, participants-related characteristics such as music experience or reading ability^{14,15}, and stimuli-related adaptations like backward speech or nonwords^{16,17}; more can be found in literature reviews^{18,19}. A few recent experiments have investigated how individual variation of the speaker identity representation might undermine the recognition accuracy, considering aspects including high versus low emotional expressiveness¹⁶ and neutral versus fearful prosodies⁵; more possible scenarios open for further investigation, as suggested by a review²⁰.

For the first research gap, the study proposes that the neurological underpinnings of speaker identification have yet to fully explore how within-speaker variation challenges listeners' brain activities. For example, in an fMRI-based speaker recognition task by Zäske et al., participants' right posterior superior temporal gyrus (pSTG), right inferior/middle frontal gyrus (IFG/MFG), right medial frontal gyrus,

and left caudate showed reduced activation when correctly identified as old versus new talkers, regardless of the linguistic content being the same or different²¹. However, an earlier electroencephalography (EEG) study by Zäske et al. did not observe this old/new effect when speaker identity variation was introduced through different texts²². Specifically, a larger, late positive component (LPC) ranging from 300 to 700 ms, detected at the Pz electrode when listeners encountered their familiar trained talker expressing the same text (i.e., hearing a replay with non-varied linguistic content), was absent when the talkers delivered new texts.

In support of the assertion made by Zäske et al.²¹, this study suspects that an old/new effect can still be observed despite differences in linguistic content between training and testing sessions in event-related potential (ERP) analyses. This rationale stems from the notion that the absence of the old/new effect in Zäske et al.²², under conditions where different texts were used, may be attributed to the lack of an additional check session during the training task to ensure thorough and effective identity learning, as suggested by Lavan et al.²³. Consequently, the first objective of the study is to examine and validate this hypothesis. This study aims to test this by adding a checking session to the training-testing paradigm²².

Another key question this study aims to address is the robustness of speaker identification in the presence of speech prosody. Previous behavioral studies have suggested that listeners particularly struggle to recognize talkers across different prosodies, which indicates a modulatory role of prosodic context - listeners underperformed in the different training-testing prosody conditions. This study aims to test

this by exposing listeners to recognize familiar talkers in either confident or doubtful prosodies²⁴. This study expects that the observed ERP differences will help explain how speech prosody influences identity recognition.

The core objective of the current study is to investigate the robustness of the old/new effect in speaker recognition, specifically examining whether there are differences in recognizing talkers in confident versus doubtful prosodies. Xu and Armony¹⁰ performed a behavioral study using a training-testing paradigm, and their findings suggest that listeners cannot overcome prosodic differences (e.g., trained to recognize a talker in neutral prosody and tested on fearful prosody) and can only achieve accuracy lower than chance level¹⁰. Acoustic analysis indicates that speakers expressing varied emotive states are associated with VTL/F0 modulation; for example, confident prosody is characterized by lengthened VTL and lower F0, whereas the opposite is true for doubtful prosody^{11,24}. Another piece of evidence comes from the study by Lavan et al.²³, which confirmed that listeners can adapt to VTL and F0 changes of the speaker and form average-based representations of the talkers. This study reconciles that, from a behavioral data perspective, listeners are likely to still recognize the talker's identity across prosodies (e.g., trained to recognize one in confident prosody but tested in doubtful prosody; reported in a separate manuscript in preparation). Yet, the neural correlates of speaker identification, specifically the generalizability of the old/new effect observed by Zäske et al.²², remain unclear. Hence, the current study is committed to validating the robustness of the old/new effect in confident versus doubtful prosodies as contexts for testing.

The study introduces a departure from previous research paradigms in old/new effects studies. While past research

focused on how old/new talker recognition influences perception, this study extends this by incorporating two confidence levels (confident versus doubtful) into the paradigm (thus, a 2+2 study). This allows us to investigate speaker recognition within the contexts of confident and doubtful speech prosodies. The paradigm enables the exploration of the robustness of old/new effects. The analyses of memory effects and regions of interest (ROI) within both confident and doubtful speech contexts serve as evidence for this investigation.

Altogether, the study aims to update the understanding of the EEG correlates of voice recognition, with the hypotheses that the enlarged LPC of the EEG old/new effect is observable even when 1) the linguistic content is not the same, and 2) with the presence of confident versus doubtful prosody. This study investigated the hypotheses through a three-step paradigm. Firstly, during the training phase, participants established associations between three voices and their corresponding names. Subsequently, in the checking phase, they were tasked with identifying the name corresponding to a voice from a selection of three candidates. This checking, following Lavan et al.²³, aims to overcome insufficient old speaker familiarization, which led to the unobserved old/new effect when the text in the training and testing phases differed⁶, and talkers could not recognize talkers across neutral and fearful prosodies¹⁰. Finally, in the testing phase, participants distinguished between three old and three new speaker voices in a two-alternative forced-choice task, with speech prosody presented as either confident or doubtful. EEG data were collected using a 64-channel EEG system and underwent preprocessing before analysis. Statistical analysis and event-related potential (ERP) analysis were conducted

in RStudio, while MATLAB was utilized for brain topography analysis.

Regarding design details, this study proposes a speaker identity learning experiment that controls for the talker's height, which is related to VTL and influences impressions of who is talking²³. This aspect also influences social impressions, such as perceived dominance²⁵, and such higher-level impression formation might interact with decoding speaker identity²⁶.

Protocol

The Ethics Committee of the Institute of Linguistics, Shanghai International Studies University, has approved the experiment design described below. Informed consent was obtained from all participants for this study.

1. Preparation and validation of the audio library

1. Audio recording and editing

1. Create a Chinese vocal database following the standard procedure of making a previous English version while making adaptations where needed to fit into the context of China¹¹. For the experiment here, 123 sentences containing three types of pragmatic intentions, namely, judgment, intention, and fact, were used. To do this, refer to an existing English statement corpus¹¹ and create a localized Chinese version with additional localized scenarios.
2. Recruit 24 speakers (12 females) to express these sentences in neutral, doubtful, and confident prosodies while referring to and adapting specified instructions of past recording tasks^{11,24}.

1. For the speakers here, enlist 24 standard Mandarin speakers from Shanghai International Studies University, 12 females and 12 males, with demonstrated proficiency in Mandarin through scores of 87 to 91 on the Putonghua Proficiency Test. Male participants averaged 24.55 ± 2.09 years in age, with 18.55 ± 1.79 years of education and an average height of 174.02 ± 20.64 cm. Females averaged 22.30 ± 2.54 years in age, with 18.20 ± 2.59 years of education and an average height of 165.24 ± 11.42 cm. None reported speech-hearing impairments or neurological or psychiatric disorders.

3. Ask the speakers to repeat each text two times. Set the sampling rate at 48,000 Hz in the software Praat²⁷. Ensure that no stream is longer than 10 min, as Praat can break down, causing recording loss.
4. Edit the long audio stream into clips per sentence with Praat. Since there are two repeats of the same text, select the version that best represents the intended prosody as the target sentence.

2. Audio selection

1. Normalize the audio library at 70 dB and the sampling rate at 41,000 Hz with Praat script²⁸. To do so, open Praat, load the sound files, and select them in the Objects window. Go to the Modify menu, choose **Scale intensity...**, set the New average intensity (dB SPL) to 70 in the settings window, and click **OK** to apply the normalization.
2. Recruit 48 independent listeners to rate each audio on one 7-Likert scale about confidence level: 1 for

not at all and 7 for very confident¹¹. Ensure each sentence has been rated by 12 raters.

3. Select the audio that suits designated thresholds with one major principle: ensure that the average rating for confident-intending is higher than doubtful-intending audio. Ensure these thresholds are consistent across 12 talkers of the same biological sex. For instance, if these talkers expressed two sentences, each with confident and doubtful prosodies, significant differences in ratings must be observed.
4. For the purpose of the current experimental design, use four blocks of audio, totaling 480 audio clips, with each block containing 120 audio.
 1. Divide 24 talkers into four groups of six, with two groups of males and two groups of females, each group consisting of talkers of the same biological sex.
 2. For each group, select audio clips based on perceptual ratings (on the same text), ensuring that the average confidence ratings were higher than the doubtful ratings for each sentence. These four blocks differ in the following ways: 1) the combined six talkers - their identities are different; 2) half of the blocks are expressed by males and the other half by females; and 3) the text expressed in each block is different.
5. Before the selection process begins, document the height data for each speaker. Use this information to divide the speakers into four independent groups based on gender and height.

1. There are 24 speakers in total, divided equally between males and females. Within each gender group, sort the 12 individuals by height.
6. Split these 12 individuals into two groups in an alternating fashion; for example, from a sorted list from 1 to 12, individuals 1, 3, 5, 7, 9, and 11 would form one group, and the other half would form the second group. Within these groups, perform the selection of speakers for the audio clips at regular intervals based on their height.

NOTE: The inclusion of height as a control factor is based on findings suggesting that speaker height-related acoustic measures (VTL and F0) influence talker and speaker identity recognition²³.

2. Programming for EEG data collection

1. Design the experiment matrix
 1. The study employs a within-subject design. Prepare a testing session that presents as per for each subject while adapting the training session. Prepare four blocks, with male and female speakers taking each half of two blocks. Assign two blocks for being trained in confident prosody and tested on both confident and doubtful, as well as trained in doubtful prosody and tested on both confident and doubtful, as suggested in **Figure 1**.
 2. Decide the duration of the functioning screens by referring to existing EEG studies on speaker identification and vocal confidence perception^{22,29}. Organize the sequence of the four blocks with a Latin square matrix between participants^{30,31}. Customized Python coding is recommended to prepare such a list. See the Code Snippet for the

Latin Square matrix and trial list for the PsychoPy program on OSF³².

3. Select talkers upon each interval from a height sequence of the same biological sex. For each block, select six speakers from the original 24 talkers, who group into four lists according to the talkers' reported height.
4. Select the first 24 names in China's Hundred Family Surnames. Randomly assign the surnames to the 24 talkers who expressed the audio by addressing them like Xiao (Junior in Chinese) ZHAO.
5. Put together all relevant information in a spreadsheet with columns for Speaker (1 to 24), Biological Sex (male or female), People Name (from the 24 surnames), Confident Level (confident or doubtful), Item (text index), Rated Confidence Level (averaged score from the perceptual study), Sound (e.g., sound/1_h_c_f_56.wav),
6. Correctly recognize one out of three (1, 2, or 3), and correctly recognize old and new (old or new). Additionally, make sure columns named training_a, training_b, training_c, check, and test have been added.
7. Add the columns training_a_marker, training_b_marker, check_marker, and testing_marker to the spreadsheets to send EEG markers. Format these markers with three digits, meaning even the number 1 is written as 001.

2. Building up the three sessions

NOTE: PsychoPy is recommended to build up the program, mainly by utilizing builder mode. The Code Component in the builder is additionally used to connect the program with the EEG data collection system,

counterbalancing the F and J buttons and calculating the accuracy to be reported on the screen.

1. Before all else, click on the **Edit Experiment Settings** icon and adjust the Experiment Info cell into two fields, namely, Participant and Block. Leave the default for both of them as blank. In this study, among the 40 participants, each having four blocks, 4/40 participants went through certain blocks again (if the accuracy in the Check session is lower than 10/12), with a rate of redoing of 19 redo-counts/4 blocks x 40 participants = 11.875%.

2. Training session: repeated identity learning for three times

1. Define a loop named Training_A, which contains three screens: Fixation, Presentation, and a Blank. Tick the **Is Trials** option. Keep the nReps 1, leaving Selected rows and Random Seed blank. Write the Condition as below:


```
"$"trials/
{:}_training_a.xlsx".format(explnfor["Participant"],
explnfor["Block"])
```

Where the trials/ is the name of the folder; Participant is the participant's index; Block is the sequence of blocks of the current block.
2. In the Fixation screen, add a Text Component, with Start Time set as 0, Duration Time set as 2 (s), and a + sign put into the Text inputting window that selects **Set Every Repeat**. Likewise, include a similar Text component in the Blank screen with no information in the Text cell, and it lasts 0.5 seconds.
3. In the Presentation screen, perform the following actions:

1. Add a Sound component, with Start Time set as 0, Stop Duration Time left blank, and the Sound cell input with \$Sound and select **Set Every Repeat**. Tick the **Sync Start With Screen**.
 2. Add another Text component, with the Start Condition cell inputted with `Cross_for_Training_A.status == FINISHED`. Leave the Stop Duration cell blank. The text cell shows \$Name. Select **Set Every Repeat**.
 3. Add a Key_Response_Training_A, in which the Start Condition is `Training_A.status == FINISHED`. Leave the Stop Duration cell blank. Tick the **Force End of Routine**. For Allowed keys cell, add space; for setting, select **Constant**.
 4. Add a Cross_for_Training_A. Its Start Time is set as 0; the Stop Condition cell is set as `Training_A.status == FINISHED`. Put a + sign into the Text inputting window and select **Set Every Repeat**.
4. Prepare Training_B by following a similar procedure as Training_A.
3. Checking session: Select the names of the three participants who are talking.
 1. Define a loop named Check, with the same Fixation and Blank screen as the training session.
 2. Use a different presentation than the training by adding a function to collect the reaction from the keyboard. In the Presentation screen, perform the following action.
 1. Add a Sound component and name it Checking_audio, with Start Time set as 0 and leave the Stop Duration cell blank. Set the Sound cell as \$Sound, with Set Every Repeat on.
 2. Add a Text component named Show_names, with Start Condition written with a command:
`Checking_audio.status == FINISHED` and leave Stop Duration blank. Set the text cell to \$ People_Name, with Set Every Repeat on.
 3. Add a Keyboard component and title it Key_Response_Check, with the Start Condition being `Checking_audio.status == FINISHED` and leave Stop Duration blank. Select **Force End of Routine** with the Allowed keys num_1, num_2, and num_3 remaining Constant so that participants could use the number pad to index their choice.
 4. Add a fixation named Cross_Check, with Start Time being 0 and Stop Condition input with `Checking_audio.status == FINISHED`. Add a + to the Text cell, which will select Set Every Repeat.
 3. Insert a Code Component. In the Begin Experiment section, initialize total_trials, current_correct, current_incorrect, and current_accuracy as 0. In the Begin Routine, define user_input as None. In the

Each Frame section, collect the user's input from the keyboard and check against the correct response stored in the spreadsheet file, with a key code of `user_key = Key_Response_Check.keys` to extract 1, 2 or 3. Then, use it to gauge against the stored 1,2 or 3 in a column named `Correctly_recognize_one_out_of_three`.

4. Once out of the loop, ensure a feedback screen appears with the following message: `check_feedbacks.text = f"` The second step is complete.
You have identified the speaker in a total of `{total_trials}` sentences,
Correctly recognized `{current_correct}` speakers,
Incorrectly judged `{current_incorrect}` speakers.
Your overall accuracy rate is `{current_accuracy}%`.
If it is below 83.33%, please signal to the experimenter,
You become reacquainted with the three speakers mentioned above.
If you meet the requirements, please press the space bar to continue.
4. Testing session: classifying the old and new talker
 1. Define a loop titled Testing. It includes Fixation and Blank (the same as in the training session) and a Presentation screen.
 2. Prepare the Presentation section as below.
 1. Add a sound-playing component, `Testing_sound`, with settings identical to those in the training session. Add a `Key_response_old_new` component, which has a Start Condition of `Testing_sound.status == FINISHED`, leave

Stop Duration blank, and tick **Force End of Routine**. In the Allowed keys, include f and j, and select **Constant**.

3. Add a Text component named `Testing_old_new`, with Start Condition being `Testing_sound.status == FINISHED`, leave Stop Duration blank, and leave the Text cell blank with Set Every Repeat - the text will be defined by a later code component.
4. Add a `Cross_Testing`, with Start Time being 0, Stop Condition being `Testing_sound.status == FINISHED`, and a + in the Text cell while Set Every Repeat is on.
5. Add a Code component as described below.
 1. In the Begin Experiment section, initialize the total number of trials (`total_trials_t`), the number of correct trials (`correct_trials_t`), and the number of incorrect trials (`incorrect_trials_t`).
 2. In the Begin Routine section, begin with a conditional check to determine the presentation format based on the participant's ID number (`expInfo["Participant"]`). If the ID number is odd, ensure the instructions for identifying old versus new stimuli are presented in one format, either ("`Old(F) New(J)`") or ("`New (F) 'Old (J)`").
 3. Outside this loop, there is a feedback screen with a code component. Ensure that each frame section reads: `testing_feedbacks.text = f"`You have identified the speaker in a total

of {total_trials_t} sentences,\nCorrectly recognized {correct_trials_t} speakers,\n\nIncorrectly judged {incorrect_trials_t} speakers.\nYour overall accuracy rate is {accuracy_t:.2f}%.\nPlease press the space bar to end this current part.

5. Connect the program with the Brain Products system as described below.

1. Synchronize the marker by setting a marker as the onset of each audio. Before the very beginning of the loop Training_A, define an EEG marker sending protocol in the code component Begin Experiment, as described below.

1. Import essential PsychoPy components, including the parallel module, and configure the parallel port's address using 0x3EFC.
2. Establish a sendTrigger function to transmit EEG markers. This function sends a specified triggerCode through the parallel port with parallel.setData(triggerCode) after verifying if it's a NumPy integer and converting it as needed.
3. Add a short wait of 16 ms to ensure marker capture before resetting the trigger channel to 0 with parallel.setData(0).

2. Sending the marker to the EEG recorder uses sendTrigger(). Include the exact name of the corresponding column in brackets. In this study, there are training_a_marker, training_b_marker,

check_marker, and testing_marker - refer to the column previously defined in the spreadsheet.

3. Collecting EEG data

1. Preparing the venue

NOTE: There are at least two computers available to perform the data collection. One is to connect to the EEG system, and the other is to collect behavioral data. It is recommended that another screen be built to mirror the behavioral-data-related computer. The system consists of an amplifier and passive EEG caps.

1. For this study, recruit participants without any reported speech-hearing impairment. Ensure participants do not have any psychiatric or neurological disorders. A total of 43 participants were selected, with three excluded due to alignment issues with the EEG markers. Of the remaining 40, there were 20 female and 20 male participants. Females were aged 20.70 ± 0.37 years, while males were 22.20 ± 0.37 years old. Their years of education were 17.55 ± 0.43 for females and 18.75 ± 0.38 for males.
2. Assign participant IDs and invite participants to wash and dry their hair within one hour before participating in the experiment.
3. Mix the electrolyte gel and abrasive electrolyte gel in a 1:3 ratio, adding a small amount of water. Stir the mixture evenly in a container with a spoon.
4. Prepare fine-tipped cotton swabs and a dry EEG cap.
5. Have the participant sit comfortably in a chair and inform them that the experimenter will apply the EEG cap. Explain that conductive paste, which is

harmless to humans and enhances brain signal reception, is applied to the cap's holes using cotton swabs.

6. Provide the participant with instructions about the experimental tasks and an informed consent form for the experiment. Proceed with the preparation phase after obtaining the participant's signature.
7. Connect the EEG cap to the amplifier, which in turn connects to the EEG data acquisition computer. This study uses a passive cap, so it is necessary to use an additional monitor to check the color indicators on the 64 electrodes.
8. Open BrainVision Recorder³³ and import a customized workspace file that has defined the recording parameters. Click on **Monitor** to check the impedance. The color bar, from red to green, is influenced by the set resistance levels, with the target impedances ranging from 0 to 10 kΩ.

2. Preparing the participants

1. Ask the participant to sit upright in a chair. Select an appropriately sized (size 54 or 56) gel-based passive electrode system for the participant's head and ensure the electrode system is correctly fitted according to the 10-20 system^{28,34}.
2. Begin by dipping a disposable cotton swab into the conductive paste and applying it into the holes of the cap, making sure to rub against the participant's scalp. An electrode's corresponding indicator turning green on the EEG data collection computer signifies that it is successfully collecting optimal data.
3. After the indicative color for all electrodes on the screen, except for the two independent-sided

electrodes turn green (on the Monitor screen), apply the conductive paste to the side electrodes. Attach the left electrode near the participant's left eye, at the area of the lower eyelid, and the right electrode near the right temple.

4. Once all electrodes are green, place an elastic net over the participant's head to help the EEG cap fit more securely and stably against the participant's head.
 5. Equip the participant with wired headphones (specific air-conduction headphones used in the lab). Close the electromagnetic shielding door and guide the participant's actions through a microphone that allows communication inside and outside. Additionally, monitor the participant's movements through an external monitor, such as reminding them not to move their body significantly; also monitor the participant's progress in behavioral tasks through a behavioral data monitor.
 6. Ask the participant to wear earphones connected to the behavioral collection computer through an audio interface.
- ## 3. Running the experiment block-by-block independently
1. On the EEG data collection computer, open BrainVision Recorder and click on Monitor to double-check the impedance and Stat/Resume Recording and start recording. Create a new EEG recording file and name it accordingly, for example, 14_2, which means the second block for participant number 14.
 2. Open the PsychoPy program's Run experiment (green button) for the behavioral experiment, enter the participant's ID (e.g., 14) and the corresponding

block number (e.g., 2), and click **OK** to start the experiment.

3. Closely monitor the accuracy of data reported on the screen after the participant completes the Check phase on the behavioral data computer. If the accuracy is below 10 out of 12, ask the participant to redo the training session until they achieve the required accuracy before moving on to the testing phase.
4. Pay close attention to the final accuracy of old versus new recognition reported on the screen after the participant completes the testing phase of the block. If the accuracy is exceptionally low (for example, below 50%), inquire about possible reasons from the participant.

4. Post-EEG experiment

1. After the participant has completed all blocks, invite them to wash their hair. Clean the EEG cap by removing residual conductive paste with a toothbrush, taking care not to wet the signal connectors, and wrapping them in plastic bags. Once cleaned, hang the EEG cap in a well-ventilated area to dry.
2. Copy the EEG and behavioral data onto a portable hard drive, ensuring that the EEG data and behavioral data correspond. For example, the EEG data is named with two files, 14_2.eeg and 14_2.vhdr, and the behavioral data as a 14_2.xlsx file.

4. EEG data processing

NOTE: The following descriptions involve EEG data preprocessing, statistical analysis, and visualization using MATLAB and RStudio for batch processing.

1. Preprocessing the EEG data with MATLAB

1. Merging the EEG and behavioral data

1. Given that participants might need to redo the task if they do not reach the required accuracy of 10/12 or above, which affects the naming of EEG and behavioral data, for example, 14_2.vhdr might become 14_2(1).vhdr, standardize the filenames by removing characters other than 14_2. While iterating through each participant's data, name the data files as sub, stripped_filename, .set, resulting in files like sub14_2.set (containing metadata and links to the EEG dataset) and sub10_1.fdt (the actual EEG data) automatically saved. This renames the 14_2.vhdr and 14_2.eeg files to sub14_2.fdt and sub14_2.set.
2. Use the EEG = pop_mergeset() function to merge the data into a single file for each participant, combining different block data in chronological order rather than numerical order of blocks 1,2,3,4.
3. Merge multiple behavioral data files into one spreadsheet per participant based on chronological order, which is essential for later synchronization.
4. Customize code to synchronize trials in the EEG signals with trials in the behavioral signals. For example, testing_list = [37:108, 145:216,

253:324, 361:432] would correspond to the EEG marker points for the four blocks.

5. Convert the behavioral data spreadsheet into a .txt file, resulting in a table with data in both rows and columns. Column names include most of those mentioned in step 2.1.
6. Redefine the content of EEG data by adding information into the EEG data using code similar to the following, for example, `EEG = pop_importepoch(EEG, behav_txt_path, {'Epoch', 'Sound', 'Speaker', 'Gender', 'Confidence_level', 'old_new_speaker', 'same_different_prosody', 'Response'}, 'timeunit', 1, 'headerlines', 1)`. This process merges each participant's corresponding EEG and behavioral data through batch processing.

NOTE: The Response values of 1 and 0 come from behavioral data, where 1 represents a correct judgment, and 0 represents an incorrect one.

2. Preprocessing the EEG data

1. For reference and re-reference^{29,35}, call the `pop_reref` function to re-reference the EEG data to the FCz electrode, ensuring that each signal is calculated relative to the FCz electrode. Use the `pop_reref` function to re-reference the EEG data to channels 28 and 29, representing the bilateral mastoid electrodes located at the posterior scalp, ensuring that each signal is calculated relative to the bilateral mastoids.
2. Set a high-pass filter (for removing linear trends) with `EEG = pop_eegfiltnew(EEG, [], 0.1, 16500, 1, [], 0)`, and perform baseline correction from

-500 to 0 ms with `EEG = pop_rmbase(EEG, [-500 0])`.

3. Manually inspect bad trials: after importing the data with EEGLAB, select **Plot**, then click on **Channel Data (scroll)**, and set the Value's maximum to 50.
4. Delete trials with visible muscular and other types of artefacts and mark bad electrodes: hovering the mouse over the channel's waveform will display its electrode. Record all bad electrodes, return to the EEGLAB main page, select **Interpolate Electrodes** under Tools, choose **Select from Data Channels**, select the electrodes needing interpolation, and confirm with OK. Save the file to a new folder.
5. Conduct principal component analysis (PCA) with `EEG = pop_runica(EEG, 'extended', 1, 'pca', 30, 'interrupt', 'on')`. Manually reject problematic ICAs, removing artifacts from eyes, muscles, and channel noise, and then save the file.
6. Use the `pop_eegthresh` function to set a threshold from -75 to +75Hz to remove extreme values^{34,36,37}.
7. Apply `pop_eegfiltnew` with parameters set (the third input parameter) to 30 to retain frequencies of 30Hz and below³⁸.
8. Customize code to list all conditions of interest, including `old_new_speaker = {'old', 'new'}`; `same_different_prosody = {'same', 'different'}`; `Confidence_level = {'c', 'd'}`; and `Response = {'1', '0'}`. Then, combine these conditions to create data combinations like

sub1_new_different_c_0 and save them as files with a txt extension.

2. ERPs analysis with RStudio

1. To organize the data, convert it to a long format. Import all .txt files into RStudio and use the rbind function to append each temporary data frame to alldata, creating a large data frame containing all file data. Rename the Row column in all data to Time for accuracy. Utilize the melt function to convert alldata from wide to long format (Data_Long), where each observation occupies a row and includes all related conditions and channel information.
2. Use the filter function from the dplyr package to select data matching specific conditions: Judgement is 1. Source is h. Memory is either old or new. Prosody is c or d.
3. Define regions based on electrode channels as follows: Left anterior (F3, F7, FC5, F5, FT7, FC3, AF7, AF3). Left central (C3, T7, CP5, C5, TP7, CP3). Left posterior (P3, P7, P5, PO7, PO3). Medial anterior (Fz, AFz, FC1, FC2, F1, F2, FCz). Medial central (CP1, CP2, Cz, C1, C2, CPz). Medial posterior (Pz, O1, Oz, O2, P1, POz, P2). Right anterior (FC6, F4, F8, FC4, F6, AF4, AF8, FT8). Right central (CP6, C4, T8, CP4, C6, TP8). Right posterior (P4, P8, PO4, PO8, P6). Group these regions into anterior, central, and posterior regions.
4. Save the workspace for subsequent data loading. To save, use setwd(); to load, use load().

3. Statistical analysis

1. For EEG data analysis across all the electrodes, filter the dataset to include only relevant data points where Judgement is 1, Source is h, Memory is

either old or new, Subject is not empty, and Time is between 400 and 850 ms.

2. Update the names of the regions of interest (ROI) based on predefined mappings. For example, left anterior, medial anterior, and right anterior are for anterior.
3. Fit a linear mixed-effects model to the data using lmer from the lme4 package³⁹, with Voltage as the response variable and Memory and ROI as fixed effects, including random intercepts for Subject and Channel: `fit_time_window <- lmer(Voltage ~ Memory * ROI + (1|Subject) + (1|channel), data=DATA)`. Replace DATA with combined, confidently only, and doubtful-only data repeatedly. See an example code on OSF³².
 1. Obtain the analysis results from the fitted model: `anova(fit_time_window)`, `eta_squared(fit_time_window)`, and `emmeans(fit_time_window, specs = pairwise ~ Memory * ROI, adjust = "Tukey")`.
4. For EEG data analysis in Pz, when filtering the dataset, follow the same steps as above but also add the condition `Channel == 'ChPz'`. Repeat the above process, but use `lmer(Voltage ~ Memory + (1|Subject))` to analyze Pz data from 400 to 850 ms.
5. To plot ERPs in the Pz (repeat over the combined, confidently only, and doubtful only dataset), filter the dataset to include only relevant data points where Judgement is 1, Source is h, Memory is either old or new, and Subject is not empty.
 1. Define a vector containing multiple electrode points (including Pz), and prefix them with Ch

to match the channel naming convention in the data. Select Pz out.

- Specify the time window for ERP analysis: `time_window <- c(400, 850)`. Define the electrode of interest, in this case, Pz. Loop through the selected electrode and create plots as described below.

- Filter the data for the Pz electrode using `filter(Channel == k)` to isolate the relevant data points.
- Create an interaction factor for line type and color based on the Memory condition using `interaction(current_channel_data $Memory)` and label the conditions as Old and New.
- Compute summary statistics and standard error for the Voltage measurements over time using the `summarySEwithin` function, specifying Voltage as the measure variable and Time as the within variable.
- Generate the ERP plot for the Pz electrode, by adding a background for the specified time window using `geom_rect` with the parameters `xmin`, `xmax`, `ymin`, and `ymax`. Include standard error ribbons with `geom_ribbon`, drawing the mean voltage with `geom_line`. Customize the plot appearance and labels using functions like `scale_x_continuous`, `scale_y_reverse`, `scale_linetype_manual`, `scale_fill_manual`, and `scale_color_manual`.

- Use `theme_minimal` for the base theme and further customize text sizes and legend placement with `theme`.

4. Topography plotting with MATLAB

- Import data and setting conditions, define the list of subjects from 1 to 40 with `subject_list = 1:40`. Define two empty cell arrays to store data for correct classifications of old and new conditions: `human_timelocked_old_correct = {}`; `human_timelocked_new_correct = {}`. Loop through the subject list, import each subject's data, and filter it based on conditions.
- Extract event information from raw EEGLAB data, selecting only events with the Response equals to 1. Select trials with Source equal to h and update the data structure accordingly. Separate data for old and new conditions, limited to correct trials with Source h, and perform time-lock analysis.
 - Calculate the grand average for both old and new conditions: `cfg = []`; `grandavg_old_correct = ft_timelockgrandaverage(cfg, human_timelocked_old_correct{:})`; `grandavg_new_correct = ft_timelockgrandaverage(cfg, human_timelocked_new_correct{:})`.
- Perform the permutation test as described below.
 - Define the neighbor configuration using a specified layout file: `cfg_neigh = []`; `cfg_neigh.method = 'distance'`; `cfg_neigh.layout = 'path_to_layout_file'`; `neighbours = ft_prepare_neighbours(cfg_neigh)`.

2. Configure parameters for the permutation test, including the design matrix and statistical method:


```
cfg = [];
cfg.method = 'montecarlo';
cfg.statistic = 'ft_statfun_indepsamplesT';
cfg.correctm = 'cluster';
cfg.clusteralpha = 0.05;
cfg.clusterstatistic = 'maxsum';
cfg.minnbchan = 2;
cfg.tail = 0;
cfg.clustertail = 0;
cfg.alpha = 0.05;
cfg.numrandomization = 1000;
cfg.neighbours = neighbours;
cfg.design = [2*ones(1, length(human_timelocked_new_correct))
ones(1, length(human_timelocked_old_correct))];
cfg.ivar = 1.
```

 Furthermore, refer to the following link (https://www.fieldtriptoolbox.org/tutorial/cluster_permutation_freq/) for tutorials on using Fieldtrip⁴⁰.
3. Perform the statistical test on the averaged data for old and new conditions:


```
stat = ft_timelockstatistics(cfg, human_timelocked_old_correct{:}, human_timelocked_new_correct{:}).
```
4. Perform custom interval plotting as described below.
 1. Compute the difference between the two conditions:


```
cfg = [];
cfg.operation = 'subtract';
cfg.parameter = 'avg';
grandavg_difference = ft_math(cfg, grandavg_old_correct, grandavg_new_correct).
```
 2. Define time windows:


```
time_windows = {[0.500, 0.800] % LPC}.
```
 3. Create a figure and plot the difference between conditions with


```
ft_topoplotER(cfg_plot, grandavg_difference).
```

Representative Results

The classic old/new effect is characterized by a significant increase in listeners' brain activity on the Pz electrode (between 300 to 700 ms) when the speech content of the testing session matches that of the training session, particularly in the old talker condition compared to the new talker condition²². The protocol unveils an updated version of this effect: Firstly, observing larger positive trends in the Pz electrode and across the entire brain region for the old condition compared to the new talker condition between 400 to 850 ms. Secondly, the speech content in the testing session will differ from that of the training session. Thirdly, both confident and doubtful speech prosody conditions are expected to exhibit these trends. Lastly, the old/new effect is more pronounced in doubtful condition during the testing session (**Figure 2**).

The LMER analysis with the formula

$$\text{lmer}(\text{Voltage} \sim \text{Memory} * \text{ROI} + (1|\text{Subject}) + (1|\text{Channel}))$$

suggests that both memory types (old versus new) and ROI have main effects, as well as an interaction between memory and ROI (**Table 1**). Further post-hoc analysis revealed that, across all brain regions, the old condition exhibits a larger positive voltage than the doubtful condition, including in the anterior, central, and posterior regions (**Table 2**). Comparing the beta values suggests that the old/new effect was more pronounced at central and posterior electrodes than anterior electrodes: for the combined dataset - Anterior $\beta = .40$, Central $\beta = .63$, and Posterior $\beta = .60$; for the confident dataset - Anterior $\beta = .61$, Central $\beta = .63$, and Posterior $\beta = .76$, and for the doubtful dataset - Anterior $\beta = .44$, Central $\beta = .87$, and Posterior $\beta = .69$. The involvement of central

and posterior electrodes was most noticeable in the doubtful prosody condition.

With the formula

$$\text{lmer}(\text{Voltage} \sim \text{Memory} + (1|\text{Subject}))$$

we confirmed the existence of old/new effects in the Pz electrode. At the Pz electrode, a main effect of memory (old versus new) was observed ($F(1, 69341.99) = 120.46, p <$

$.001, \eta^2_p = .002, \beta = .425, SE = .039, z\text{-ratio} = 10.98, p < .001$). In the confident-only condition, a main effect of memory (old versus new) was observed at the Pz electrode ($F(1, 34318.32) = 5.04, p = .025, \eta^2_p = .0001, \beta = .125, SE = .056, z\text{-ratio} = 2.25, p = .025$). In the doubtful-only condition, a main effect of memory (old versus new) was observed at the Pz electrode ($F(1, 34993.20) = 317.02, p < .001, \eta^2_p = .009, \beta = .914, SE = .051, z\text{-ratio} = 17.81, p < .001$).

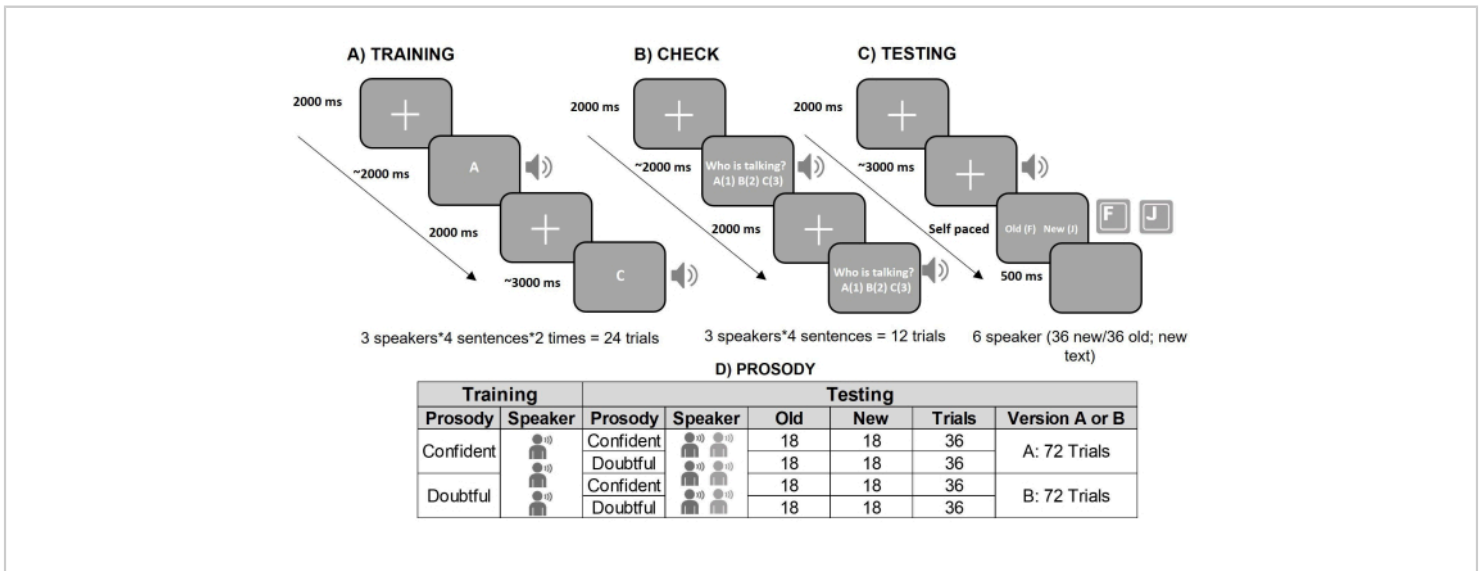


Figure 1: Workflow of the data collection for each block. In (A) Training, listeners hear a voice and associate the name subsequently presented with it. Three old talkers are required to be remembered. The language that appeared in the program was originally Chinese. The A and C represent names such as Xiao (Junior) ZHANG. In (B) Checking, listeners identify the talker's name upon hearing a voice by pressing 1, 2, or 3 on the number pad to associate the voice identity with names like Xiao ZHAO. In (C) Testing, listeners hear a voice and classify it as spoken by the old or new speaker. As illustrated in (D) Prosody Design, listeners learn three talkers express only confidently or doubtfully, but hearing six talkers speak both confidently and doubtfully. The appearance of Version A or B is mutually exclusive. If Version A appears with a male or female speaker, Version B will appear with the corresponding female or male speaker. [Please click here to view a larger version of this figure.](#)

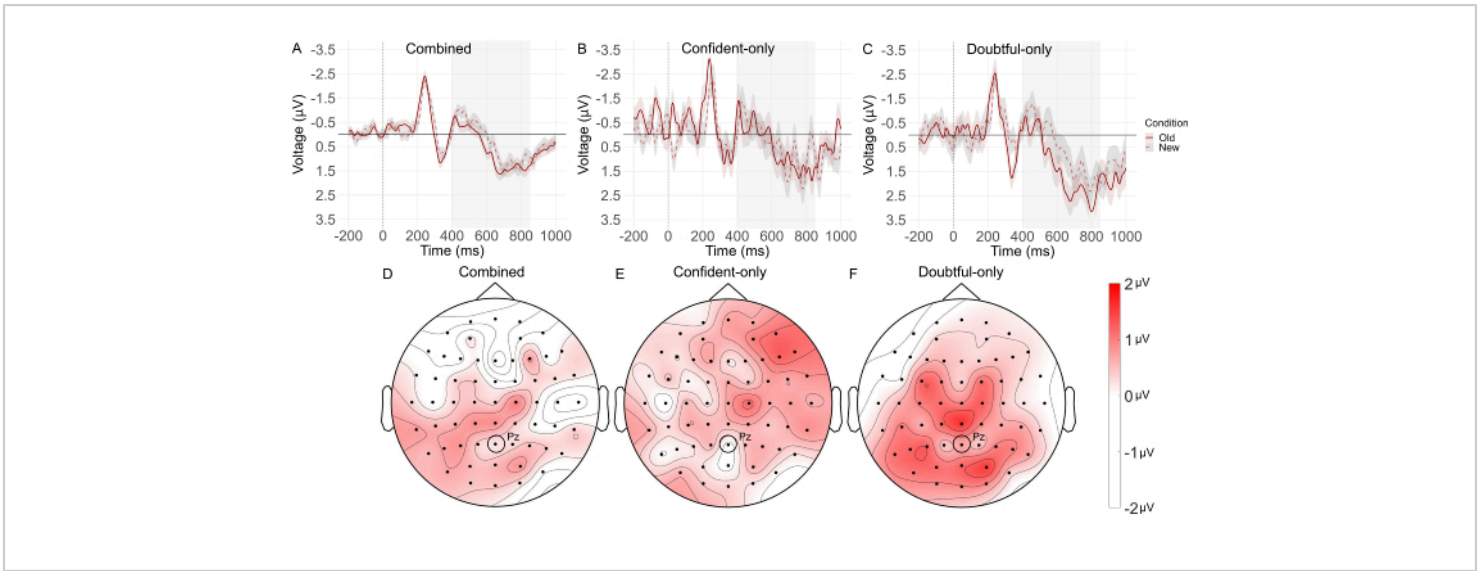


Figure 2: The old/new effect. (A, B, C) Figures display the grey-indicated ERP of Pz electrodes from 400 to 850 ms for the prosody-combined, confident-only, and doubtful-only conditions, respectively. (D, E, F) Figures illustrate the topography of the old minus new condition across all electrodes (depicted as black dots) for the prosody-combined, confident-only, and doubtful-only conditions. [Please click here to view a larger version of this figure.](#)

Context	Brain Region	F value	Pr(>F)	Eta2_partial
Combined	Memory	9938.98	.00	.00
	ROI	4.13	.02	.13
	Memory:ROI	182.37	.00	.00
Confident	Memory	7291.22	.00	.00
	ROI	3.60	.03	.12
	Memory:ROI	41.94	.00	.00
Doubtful	Memory	8333.38	.00	.00
	ROI	4.65	.01	.15
	Memory:ROI	290.15	.00	.00

Table 1: Results from LMER analysis for old/new effect across brain regions: Combined, confident, and doubtful datasets. Using post-hoc analysis, * significant at $p < .05$, ** significant at $p < .01$, *** significant at $p < .001$.

Context	Brain Region	Contrast	Estimate	SE	z	p
Combined	Anterior	old-new	.40	.01	43.70	.00***
	Central	old-new	.63	.01	61.74	.00***
	Posterior	old-new	.60	.01	67.51	.00***
Confident	Anterior	old-new	.61	.01	46.63	.00***
	Central	old-new	.63	.01	43.22	.00***
	Posterior	old-new	.76	.01	59.95	.00***
Doubtful	Anterior	old-new	.44	.01	35.95	.00***
	Central	old-new	.87	.01	64.05	.00***
	Posterior	old-new	.69	.01	57.75	.00***

Table 2: Post-hoc test results for old/new effects across brain regions: Combined, confident, and doubtful datasets. Using post-hoc analysis, significant at $p < .001$ (***).

Discussion

The study presents a pipeline for EEG data collection and analysis, focusing on recognizing previously learned speaker identities. This study addresses variations between learning and recognition phases, including differences in speech content²² and prosody¹⁰. The design is adaptable to a range of research fields, including psycholinguistics, such as pronoun and anaphoric processing⁴¹.

The training-testing paradigm is a classic experimental design used to assess participants' learning outcomes on specific topics such as voice learning^{42,43}. This paradigm evaluates how well participants have learned particular information (as reflected in accuracy)¹⁰. It allows researchers to introduce variables incrementally under controlled experimental conditions, such as different prosodies during training and testing phases, to understand their influence on voice recognition accuracy, for example,

VTL/F0 modulated voices²³, fearful versus neutral¹⁰, or doubtful versus confident in this study.

However, the paradigm has limitations. The differences between the learning and testing environments can affect the validity of experimental results, as controlled learning conditions may not reflect the more variable testing conditions. For instance, the training session uses a single prosody rather than a proportionate difference, such as 30% versus 70%⁴⁴. To address this imbalance, ensuring a more diverse learning environment could better replicate real-life scenarios where speakers use varied prosodies while interacting with listeners. Additionally, this study acknowledges that the complexity of the experimental design, involving multiple stages and sophisticated programming (using tools like R Studio, MATLAB, and Python), can be challenging for newcomers.

The primary insight emphasizes the importance of adequate familiarization and a check phase. Xu and Armony's work highlights that listeners struggle to identify old talker identities without sufficient training and checks above chance levels¹⁰. Additionally, Zaske et al. found the LPC old/new effect was only present when the same text was repeated, not with different text²². In this study, the implementation of a check phase revealed the persistence of the old/new ERP effect, even with different text stimuli, supporting the fMRI studies claims²¹. The study suggests that, for training-testing-based paradigms, inserting a check session is critical. It allows listeners to form a robust impression of the speaker's acoustic identity, associating a talker with a specific symbol, such as a name²³. Without sufficient learning of the speaker's representation, listeners may struggle to adapt to within-speaker variations¹⁰.

This study also observed the role of prosody as a binding cue for speaker recognition⁴⁵. Contrary to previous views that prosody may hinder old talker recognition, this study found the old/new effect present across confident and doubtful prosody conditions. This robust effect suggests a modulation role of prosody in speaker recognition. Further analysis revealed differences in anterior region activation across prosody conditions. Confident prosody elicited lower levels of the old/new effect in anterior regions compared to doubtful prosody. This finding suggests that confident speech may make talker identification more challenging due to extended vocal tract length and lowered fundamental frequency, potentially leading to increased attention from listeners^{11,29}.

This study's design can inform future investigations into recognition impairments in patient populations, such as those with prosopagnosia or phonagnosia^{46,47}. Additionally, modifications to accommodate participants with shorter

attention spans, such as individuals with autism spectrum disorders⁴⁸, could enhance study accessibility.

Furthermore, the paradigm extends beyond speaker recognition to investigate pronoun processing and anaphoric comprehension within psycholinguistic research. Coopmans and Nieuwland⁴¹ demonstrate how neural oscillatory synchronization patterns distinguish between antecedent activation and integration in anaphor comprehension, which aligns with this study's exploration of identity-related cues. Cues alike include communicative styles (e.g., literal or ironic statements), word orders (Subject-Object-Verb (SOV), or Object-Subject-Verb (OSV) sentence structure^{44,45,49,50}), and vocal expression types (confident vs. doubtful prosody) in this paper.

Disclosures

There is no information to be disclosed.

Acknowledgments

This work was supported by the Natural Science Foundation of China (Grant No. 31971037); the Shuguang Program supported by the Shanghai Education Development Foundation and Shanghai Municipal Education Committee (Grant No. 20SG31); the Natural Science Foundation of Shanghai (22ZR1460200); the Supervisor Guidance Program of Shanghai International Studies University (2022113001); and the Major Program of the National Social Science Foundation of China (Grant No. 18ZDA293).

References

1. Larrouy-Maestri, P., Poeppel, D., Pell, M. D. The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspect Psychol Sci.* 17456916231217722 (2024).

2. Pell, M. D. Kotz, S. A. Comment: The next frontier: Prosody research gets interpersonal. *Emotion Rev.* **13** (1), 51-56 (2021).
3. Cummins, N. et al. Multilingual markers of depression in remotely collected speech samples: A preliminary analysis. *J Affect Disor.* **341**. 128-136 (2023).
4. Cummins, N., Baird, A., Schuller, B. W. Speech analysis for health: Current state-of-the-art and the increasing impact of deep learning. *Methods.* **151**, 41-54 (2018).
5. Kennedy, E. Thibeault, S. L. Voice-gender incongruence and voice health information-seeking behaviors in the transgender community. *Am J Speech-language Pathol.* **29** (3), 1563-1573 (2020).
6. Zäske, R. et al. Electrophysiological correlates of voice memory for young and old speakers in young and old listeners. *Neuropsychologia.* **116**, 215-227 (2018).
7. Lavan, N., Burton, A. M., Scott, S. K., Mcgettigan, C. Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bullet Rev.* **26**, 90-102 (2019).
8. Perrachione, T. K., Del Tufo, S. N., Gabrieli, J. D. Human voice recognition depends on language ability. *Science.* **333** (6042), 595-595 (2011).
9. Lavan, N., Knight, S., Mcgettigan, C. Listeners form average-based representations of individual voice identities. *Nat Comm.* **10** (1), 2404 (2019).
10. Xu, H. Armony, J. L. Influence of emotional prosody, content, and repetition on memory recognition of speaker identity. *Quart J Exp Psychol.* **74** (7), 1185-1201 (2021).
11. Jiang, X. Pell, M. D. The sound of confidence and doubt. *Speech Comm.* **88** 106-126 (2017).
12. Winters, S. J., Levi, S. V., Pisoni, D. B. Identification and discrimination of bilingual talkers across languages. *J Acoustical Soci Am.* **123** (6), 4524-4538 (2008).
13. Orena, A. J., Polka, L., Theodore, R. M. Identifying bilingual talkers after a language switch: Language experience matters. *J Acoustical Soc Am.* **145** (4), EL303-EL309 (2019).
14. Xie, X. Myers, E. The impact of musical training and tone language experience on talker identification. *J Acoustical Soc Am.* **137** (1), 419-432 (2015).
15. Kadam, M. A., Orena, A. J., Theodore, R. M., Polka, L. Reading ability influences native and non-native voice recognition, even for unimpaired readers. *J Acoustical Soc Am.* **139** (1), EL6-EL12 (2016).
16. Fleming, D., Giordano, B. L., Caldara, R., Belin, P. A language-familiarity effect for speaker discrimination without comprehension. *Proc Natl Acad Sci.* **111** (38), 13795-13798 (2014).
17. White, K. S., Yee, E., Blumstein, S. E., Morgan, J. L. Adults show less sensitivity to phonetic detail in unfamiliar words, too. *J Memory Lang.* **68** (4), 362-378 (2013).
18. Levi, S. Methodological considerations for interpreting the language familiarity effect in talker processing. *Wiley Interdiscip Revi: Cognitive Sci.* **10** (2), e1483 (2019).
19. Perrachione, T. K. Recognizing Speakers Across Languages. *The Oxford Handbook of Voice Perception.* S. Frühholz & P. Belin (Eds.), Oxford University Press, 515-538 (2018).
20. Lavan, N., Burton, A. M., Scott, S. K., Mcgettigan, C. Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bullet Rev.* **26** (1), 90-102 (2019).

21. Zäske, R., Hasan, B. a. S., Belin, P. It doesn't matter what you say: Fmri correlates of voice learning and recognition independent of speech content. *Cortex*. **94**, 100-112 (2017).
22. Zäske, R., Volberg, G., Kovács, G., Schweinberger, S. R. Electrophysiological correlates of voice learning and recognition. *J Neurosci*. **34** (33), 10821-10831 (2014).
23. Lavan, N., Knight, S., Mcgettigan, C. Listeners form average-based representations of individual voice identities. *Nat Comm*. **10** (1), 1-9 (2019).
24. Chen, W. Jiang, X. Voice-Cloning Artificial-Intelligence Speakers Can Also Mimic Human-Specific Vocal Expression. *Preprints*. 10.20944/preprints202312.0807.v1 (2023).
25. Pisanski, K., Anikin, A., Reby, D. Vocal size exaggeration may have contributed to the origins of vocalic complexity. *Philosoph Trans Royal Soc B*. **377** (1841), 20200401 (2022).
26. Belin, P., Fecteau, S., Bedard, C. Thinking the voice: Neural correlates of voice perception. *Trend Cognitive Sci*. **8** (3), 129-135 (2004).
27. Boersma, P. Weenink, D. *Praat: doing phonetics by computer*. <https://www.fon.hum.uva.nl/praat/>. (2022).
28. Jiang, X. Pell, M. D. On how the brain decodes vocal cues about speaker confidence. *Cortex*. **66**, 9-34 (2015).
29. Jiang, X., Gossack-Keenan, K., Pell, M. D. To believe or not to believe? How voice and accent information in speech alter listener impressions of trust. *Quart J Exp Psychol*. **73** (1), 55-79 (2020).
30. Rigoulot, S. Pell, M. D. Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PloS One*. **7** (1), e30740 (2012).
31. Cui, X., Jiang, X., Ding, H. Affective prosody guides facial emotion processing. *Curr Psychol*. **42** (27), 23891-23902 (2023).
32. Chen, W. Jiang, X. *Memorization-based training and testing paradigm for robust vocal identity recognition in expressive speech using event-related potentials analysis.*, <https://osf.io/6zu83/> (2024).
33. Gmbh, B. P. *Brainvision recorder*. <https://www.brainproducts.com/downloads/recorder/> (2024).
34. Jiang, X., Paulmann, S., Robin, J., Pell, M. D. More than accuracy: Nonverbal dialects modulate the time course of vocal emotion recognition across cultures. *J Exp Psychol*. **41** (3), 597 (2015).
35. Jiang, X. Pell, M. D. The feeling of another's knowing: How "mixed messages" in speech are reconciled. *J Exp Psychol*. **42** (9), 1412 (2016).
36. Zhou, X. et al. Semantic integration processes at different levels of syntactic hierarchy during sentence comprehension: An erp study. *Neuropsychologia*. **48** (6), 1551-1562 (2010).
37. Jiang, X., Tan, Y., Zhou, X. Processing the universal quantifier during sentence comprehension: Erp evidence. *Neuropsychologia*. **47** (8-9), 1799-1815 (2009).
38. Acunzo, D. J., Mackenzie, G., Van Rossum, M. C. W. Systematic biases in early erp and erf components as a result of high-pass filtering. *J Neurosci Meth*. **209** (1), 212-218 (2012).
39. Bates, D. Fitting linear mixed models in r. *R news*. **5** (1), 27-30 (2005).
40. Oostenveld, R., Fries, P., Maris, E., Schoffelen, J. M. Fieldtrip: Open source software for advanced analysis

- of meg, eeg, and invasive electrophysiological data. *Computat Intelligence Neurosci.* **2011**, 1-9 (2011).
41. Coopmans, C. W. Nieuwland, M. S. Dissociating activation and integration of discourse referents: Evidence from erps and oscillations. *Cortex.* **126**, 83-106 (2020).
 42. Humble, D. et al. The jena voice learning and memory test (jvlmt): A standardized tool for assessing the ability to learn and recognize voices. *Behavior Res Meth.* **55** (3), 1352-1371 (2023).
 43. Holmes, E., To, G., Johnsrude, I. S. How long does it take for a voice to become familiar? Speech intelligibility and voice recognition are differentially sensitive to voice training. *Psychol Sci.* **32** (6), 903-915 (2021).
 44. Kroczek, L. O. H. Gunter, T. C. Communicative predictions can overrule linguistic priors. *Sci Rep.* **7** (1), 17581 (2017).
 45. Kroczek, L. O. H. Gunter, T. C. The time course of speaker-specific language processing. *Cortex.* **141**, 311-321 (2021).
 46. Schroeger, A. et al. Atypical prosopagnosia following right hemispheric stroke: A 23-year follow-up study with mt. *Cognitive Neuropsychol.* **39** (3-4), 196-207 (2022).
 47. Garrido, L. et al. Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia.* **47** (1), 123-131 (2009).
 48. Schelinski, S., Borowiak, K., Von Kriegstein, K. Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive Affective Neurosci.* **11** (11), 1812-1822 (2016).
 49. Holle, H. Gunter, T. C. The role of iconic gestures in speech disambiguation: Erp evidence. *J Cognitive Neurosci.* **19** (7), 1175-1192 (2007).
 50. Regel, S., Coulson, S., Gunter, T. C. The communicative style of a speaker can affect language comprehension? Erp evidence from the comprehension of irony. *Brain Res.* **1311**, 121-135 (2010).