8206L-Predicting Spatial Preposition Naming Accuracy in Mandarin-Speaking Children from Acoustic Features Using Machine Learning

The University of Texas at Austin
Moody College of Communication

The University of Texas at Austin
Dell Medical School

Mengxuan Wu*, Beiming Cao*, Alyssa Regner*, Daniela Rodriguez-Orozco*, Jun Wang*^, Rajinder Koul*
*Department of Speech, Language and Hearing Science, University of Texas at Austin, ^ Department of Neurology, University of Texas at Austin

Introduction

Studies in the Augmentative and Alternative Communication (AAC) field have indicated that the majority of graphic symbols are categorized as nouns, making the design of symbols for spatial prepositions challenging (Koul et al., 2001). Recently, we developed a new set of graphic symbols for children and assessed their effectiveness using a common naming task (Schlosser et al., 2014, 2012, 2011).

In Mandarin-speaking contexts, spatial relationships are typically indicated using a "zai" structure (e.g., "on"-"zai shangmian").

However, during the critical developmental phase of transitioning from an egocentric to an objective viewpoint, children aged 3 to 5 may inaccurately name objects in spatial tasks with confidence. Some verbal responses to Chinese spatial prepositions (e.g., Shang (on) vs. Xia (below)) by preschoolers can be difficult to perceptually discriminate during the transcription process. Previous studies have shown that these metrics can reveal emotional states and cognitive stress, extending beyond communicative efficiency (Sukumaran & Kousalya, 2021; Jones et al., 2011).

Research Goal

We propose a novel approach using machine learning and acoustic features to predict correctness. We hypothesize that voice features within acoustic features may reflect the emotional state of the child during the task, providing additional insights into accuracy.

Methods

Participants: Mandarin-speaking children (n =145)

Age range: 3 to 5

Inclusion criteria: (a) typical Mandarin speaking children with a chronological age ranging from 3 to 5yeras, as determined from daycare records,; (b) Mandarin being the primary language spoken at home, as determined from day care records ; (c) absence of uncorrected visual or hearing difficulties, as indicated by day care records; (d) age-appropriate receptive vocabulary as determined by the Peabody Picture Vocabulary Test -Revised (PPVT-R) Chinese version and (e) receptive or expressive knowledge of spatial prepositions used as stimuli in the experiment.

Settings
The study was conducted in a quiet room where a Lenovo touchscreen laptop was employed to display graphic symbols.

Materials
Stimuli and verbal prompts were presented via text-to-speech female voice using E-prime 3.0 software. The target spatial prepositions used in the tasks were behind, in, in front of, next to, below, on, out, and between.

Experiment procedure

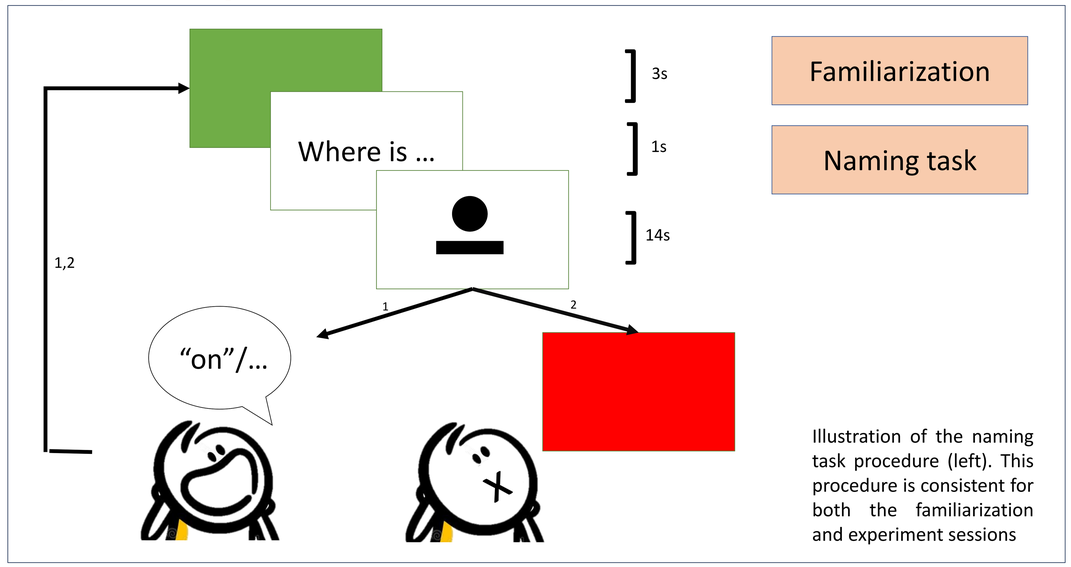


Illustration of the naming task procedure (left). This procedure is consistent for both the familiarization and experiment sessions

Data

- ASR transcribed
- reviewed by two qualified native Chinese speakers (interrater reliability ratio =99.95%)
- The dataset in this study includes a total of the sound files for both types (*n*=1135).

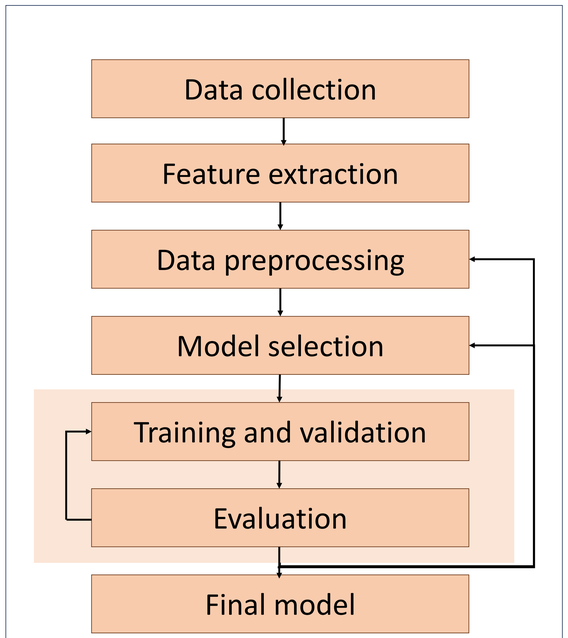
Response Type1 (*n* = 631)
Basic response: no more than 3 characters (zai) shang/xia mian/bian", "dixia" ("below")

Response Type 2 (*n* = 504)
Contextual response: more than 3 characters "xiao qiu/ren zai … de shang/xia mian/bian" or more complicated sentences ("The ball is below the …")

Unclear responses, unintelligible sentences and irrelevant sentences were excluded.

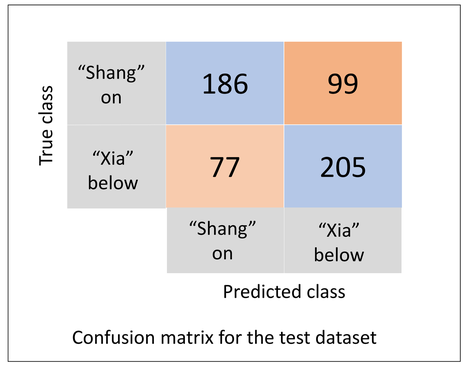
Feature extraction
OpenSMILE toolkit provided a comprehensive set of low- and high-level descriptors, including spectral, prosodic, and voice quality features.

Preprocessing
- Normalization (z-score)
- 5-fold cross validation, 50% training and 50% for testing
- 95% PCA

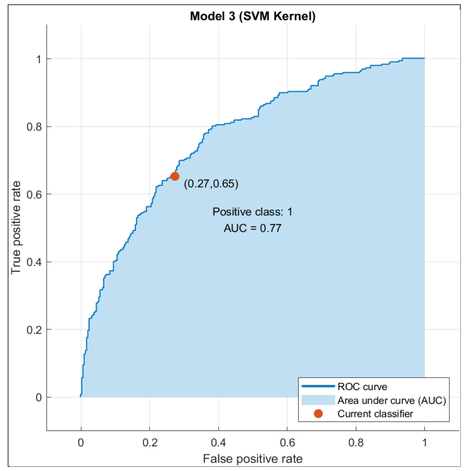


Machine Learning Workflow for Spatial Preposition Naming Accuracy Prediction

Result



Confusion matrix for the test dataset



Based on cross-validation, the SVM with Radial Basis Function Kernel achieved the best balance of accuracy and generalization.

Model performance
Accuracy: 68.9%
Recall:
Shang (on) 65.3%
Xia (below) 72.7%
F1-score:
Shang (on) 67.9%
Xia (below)) 69.9%
Precision:
Shang (on) 70.7%
Xia (below) 67.4%

The Area Under the Curve (AUC) is 0.77, which indicates a moderate to good level of classification performance.
•**False Positive Rate (FPR):** 0.27 (or 27%), meaning 27% of negatives are incorrectly classified as positives at this threshold.
•**True Positive Rate (TPR):** 0.65 (or 65%), indicating that 65% of the positive cases are correctly classified at this threshold.

Conclusion and Future Works

- Longer sentences contain more noise and have an impact on model performance.
- The accuracies of the models are significantly higher than chance level (50%), indicating the feasibility of using machine learning to predict spatial preposition accuracy in a naming task.
- Future work may involve improving the models (i.e., exploring more features and refining the training process) and applying them to children's speech recognition in Mandarin for spatial preposition naming tasks.

References

Jones, M., Anagnostou, F., & Verhoeven, J. (2011, August). The Vocal Expression of Emotion: An Acoustic Analysis of Anxiety. In ICPhS (pp. 982-985). Koul, R. K., Schlosser, R. W., &

Schlosser, R. W., Koul, R., Shane, H., Sorce, J., Brock, K., Harmon, A., … & Hearn, E. (2014). Effects of animation on naming and identification across two graphic symbol sets representing verbs and prepositions. Journal of Speech, Language, and Hearing Research, 57(5), 1779-1791.

Schlosser, R. W., Shane, H., Sorce, J., Koul, R., & Bloomfield, E. (2011). Identifying performing and under performing graphic symbols for verbs and prepositions in animated and static formats: A research note. Augmentative and Alternative Communication, 27(3), 205-214.

Schlosser, R. W., Shane, H., Sorce, J., Koul, R., Bloomfield, E., Debrowski, L., … & Neff, A. (2012). Animation of graphic symbols representing verbs and prepositions: Effects on transparency, name agreement, and identification.

Sukumaran, P., & Kousalya, G. (2021). Towards voice based prediction and analysis of emotions in ASD children. J. Intell. Fuzzy Syst., 41, 5317-5326. https://doi.org/10.3233/JIFS-189854.

Disclosure
The authors declare that they have no relevant or material financial interests that relate to the research described in this poster.