[classification]

1. **Introduction**
2. **Background**
3. **Experiment**
   1. **Participants**

245 participants were run in the study. Of these 245, 98 were recruited from a subject pool of introductory Linguistics and Cognitive Science students, and were compensated with extra credit in their course. 147 were recruited from online on the Prolific platform, and were compensated with a $5 Amazon giftcard. The use of two distinct participant pools was necessary in order to recruit a sufficient number of participants from each of the target linguistic demographic backgrounds. Based on previous experimental work reported elsewhere and conducted in our lab, we had no reason to anticipate any differences in performances between the two populations, other than higher attrition rate for participants recruited online. Regardless of their recruitment pool, all participants completed the same online experiment.

Of the participants run, 51 participants were excluded for the following reasons: did not move emojis from the left side of the screen into groups on the right side (n=10); omitted 8 (>10%) emojis from the task altogether (n=1); moved emojis from left side of screen into grid on right with no distinct grouping whatsoever (n=14); took less than 15 minutes on the task (n=21); did not save or submit file (n=3); attempted to participate and submit answers twice (n=2). With these exclusions, we were left with 89 participants run from the subject pool and 105 recruited via Prolific, for a total of 194 participants (female: 157, male: 33, transgender: 1, non-binary: 1, no response: 1).

These participants represented three major linguistic categories: English monolingual (exposed to diverse languages) (n=61); fluent/native/heritage speakers of English and one of the target Asian languages (n=75); fluent/native/heritage speakers of English and a non-target language (e.g., Spanish, French, German, Greek) (n=58). Within the English-target Asian language group, three further subgroups were represented: South Asian (e.g., Hindi, Punjabi, Bengali, Urdu) (n=23); South East Asian (e.g., Tagalog-Filipino, Indonesian, Thai) (n=27); East Asian (e.g., Korean, Mandarin, Japanese) (n=25).

* 1. **Stimuli**
     1. **Language/speaker selection**

We identified 45 speech samples from the Speech Accent Archive online. This database contains recordings of speakers reading the same script. We targeted the first two sentences of each recording for presentation, since these contained segmental features that helped to distinguish between the targeted languages/ dialects. Each wav recording lasted approximately 8 to 13 seconds. (Previous research (Bent et al. 2016) targeted an even shorter segment of the productions for classification purposes, so we were confidence that this length provided enough auditory data to guide participants’ classifications.)

(#) Please call Stella. Ask her to bring these things with her from the store: six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob.

Three main language families were selected (Asian, American English, and International English). Within the Asian family, there were three regional varieties (South, South East, East)

Within these groups of Asian languages, three languages were selected for each (South Asian: Bengali, Gujarati, Urdu; South East Asian: Indonesian, Tagalog, Thai; East Asian: Japanese, Korean, Mandarin)

Within the American English family, three regional dialects were chosen (New England, Midland, Southern)

Within the International English category, three areas were targeted (Australian, British, Africaans)

We conducted two independent analyses of the target sound files in order to verify independently that the speakers’ productions of the selected passage co-varied with their language backgrounds, and that these differences were perceptible, independent of the classification process.

* + 1. **Impressionistic analysis of morphosyntactic and phonological features**

In the first analysis, we targeted four features that are common features of dialectal and cross-linguistic variation (REFS): one morphosyntactic (omission of plural marking), and three phonological (epenthesis, cluster reduction, and presence or absence of word-final rhoticity). We then performed an impressionistic listening task to determine whether these features co-varied with the languages/dialects.

Each one of these features was coded independently by three to four research assistants using a binary system for every single speaker sound file. Researchers were provided with the sound file, and a guide for each feature assigned to them expressed in terms of the target dependent measure, as indicated in the examples below, so that for each instance in each production, they could report Yes/No.

absence of pluralization

(#) Please call Stella. Ask her to bring these thing**s** with her from the store: six spoon**s** of fresh snow pea**s**, five thick slab**s** of blue cheese, and maybe a snack for her brother Bob.

epenthesis

(#) Please call **[**ə**]** Stella. Ask her to bring these things with **[**ə**]** her from the store: six **[**ə**]** spoons of **[**ə**]** fresh **[**ə**]** snow peas, five **[**ə**]** thick slabs of blue cheese, and maybe a snack for her brother Bob.

cluster reduction

(#) Please call Stella. Ask her to **br**ing these things with her from the store: six **sp**oons of fresh snow peas, five thick slabs of **bl**ue cheese, and maybe a snack for her **br**other Bob.

presence of rhoticity

(#) Please call Stella. Ask he**r** to bring these things with he**r** from the sto**re**: six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack fo**r** he**r** broth**er** Bob.

These raw judgments were compiled in a spreadsheet, and the average of the responses was calculated. The findings are presented in Table ## in aggregate.

The summary in the table indicates that even at an impressionistic level, these sound files differ perceptibly from one another in the manner of production. For example, of the American English dialects, rhoticity is noticeably lacking in the New England utterances.[[1]](#footnote-1) The International English dialects displayed some rhoticity, but not at the level of the Midland and Southern American English dialects. Among the Asian languages, the Thai speakers were most likely to omit plural morphology and reduce consonant clusters. And while English speakers did not produce epenthesis, this was a small but noticeable feature of the Asian speakers.

Table ##

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **absence of plurals** | **epenthesis** | **cluster reduction** | **rhoticity\*** |
| **Asian** |  |  |  |  |  |
| South | Bengali | 0.0% | 16.7% | 0.0% | 100.0% |
|  | Gujarati | 0.0% | 0.0% | 0.0% | 100.0% |
|  | Urdu | 0.0% | 5.6% | 0.0% | 100.0% |
| South East | Indonesian | 0.0% | 5.6% | 0.0% | 100.0% |
|  | Tagalog | 0.0% | 11.1% | 0.0% | 100.0% |
|  | Thai | 33.3% | 0.0% | 41.7% | 50.0% |
| East | Japanese | 8.3% | 16.7% | 0.0% | 100.0% |
|  | Korean | 8.3% | 27.8% | 0.0% | 100.0% |
|  | Mandarin | 0.0% | 0.0% | 0.0% | 83.3% |
| **English** |  |  |  |  |  |
| American | New England | 0.0% | 0.0% | 0.0% | 0.0% |
|  | Midland | 0.0% | 0.0% | 0.0% | 100.0% |
|  | Southern | 0.0% | 0.0% | 0.0% | 83.3% |
| International | Australian | 0.0% | 0.0% | 0.0% | 50.0% |
|  | British | 0.0% | 0.0% | 0.0% | 50.0% |
|  | Africaans | 0.0% | 0.0% | 0.0% | 33.3% |

* + 1. **Acoustic Analysis of Vowel Formants**

Next, we loaded each wav recording into Praat, created an annotation tier, and a researcher annotated the sound files by target segments. A random subset of these boundaries were spot-checked by a second researcher to ensure accuracy. We then wrote and ran a Praat script to conduct an acoustic analysis of these segments, focusing on vowel formants.

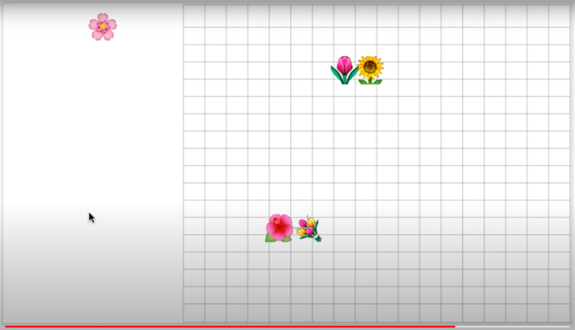
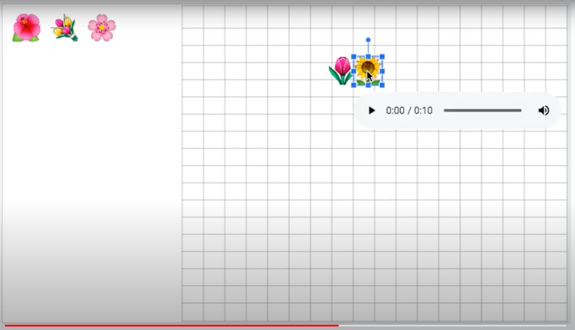
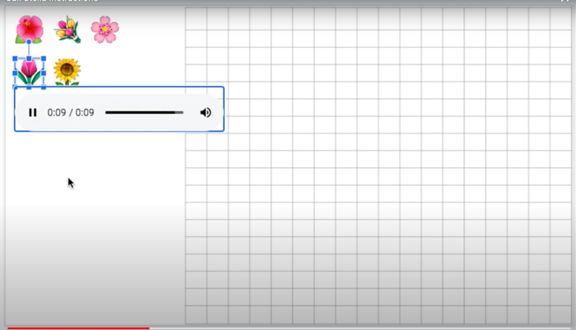
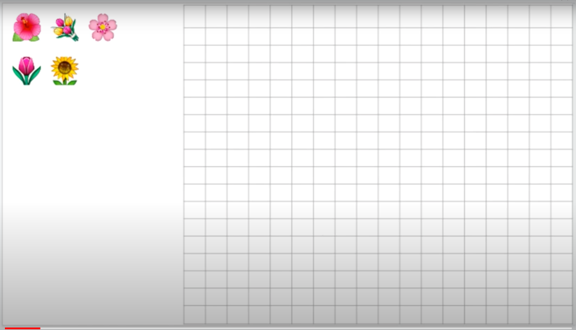
ACOUSTIC ANALYSIS

* 1. **Procedure**

Individuals were invited to participate in a task called “Call Stella” (named after a key phrase from the productions in the Speech Accent Archive “Please call Stella…”). They were brought to an online form where they completed an IRB-approved consent form, and indicate their language background and proficiency, and self-identified gender (male, female, transgender, binary, other). They were asked to indicate the languages they speak with native or heritage fluency, and for each of these, to indicate at what age they began learning it and where they learned it from. For speakers recruited in the subject pool, we also asked them how long they have lived in the US, and if they have lived in any other countries, and if so, for how long.

Once the participants provided consent and demographic background information, we asked them to view a brief 2-minute instructional video to acclimate them to the task before proceeding onto the task proper. The video is available at <https://youtu.be/4rzMBc-zz6Y> [removed for anonymization and blind review, since associated with the lab Youtube channel]. In the video, the narrator (a researcher and one of the authors welcomed the participant and explained that the experiment would be accessed via Google slides available via a link after the video. The narrator then proceeded to explain the setup of the experiment, using an example with five flower emojis on the left side of the screen instead of the 45 actual emojis used in the experiment. See Figure ##.

Figure ##. Images from the instructional training video preceding the experimental task



On the left half of the screen, were 45 emojis, each representing a different speaker from one of the target languages/dialects. The script read by all 45 speakers was printed below the emojis. See Appendix B for the key. Care was taken so as not to have any visual features associated with the languages or dialects, or invited the possibility that these superficial perceptible features could influence the classification in any way. To the right of the emojis was a grid to be used as a background for classification. See Figure ##.

Figure ##. Beginning pre-classification scene with 45 emojis and script

Chart

Description automatically generated with low confidence

Participants were instructed to hover or click on each emoji in order to listen to the audio file linked to it. The participants were told that while the passage may seem strange, it was chosen because it highlights differences in language and dialects of different speakers. They were also told that the facial expressions of the emojis were irrelevant, and that the participants should listen to the audio files, and drag the emojis to the grid, arranging them into clusters based on how similar or different they sound with similar ones close together. (In the video, the narrator demonstrated each of these steps.) Participants were told that they could listen to the audio files as many times they wanted, or create as many clusters as they wanted or put as many emojis in a cluster as they wanted, but that each cluster should have at least three speakers in it, but otherwise. The task was done when the participants were satisfied that they had successfully grouped the emojis into clusters based on how they sounded.

Figure ##. Examples of eight actual participant classifications/cluster creations of emojis/speakers

|  |  |
| --- | --- |
| Chart, scatter chart  Description automatically generated | Timeline, scatter chart  Description automatically generated |
| Background pattern, timeline, scatter chart  Description automatically generated | Timeline, scatter chart  Description automatically generated |
| Timeline, scatter chart  Description automatically generated | Chart, timeline, scatter chart  Description automatically generated |
| Timeline, scatter chart  Description automatically generated | Timeline  Description automatically generated |

The work was autosaved on Google slides. Participants were not granted credit or given compensation if they spent less than 10 minutes on the task. Typically, participants took between 15 to 25 minutes to complete the classification clustering.

Once the participants’ clustering was done, a researcher then reviewed the slides and then took a screen shot of the slide with the final sorting (as in Figure ##), saving it with the participant number. The clusters were then reviewed one by one manually,

* 1. **Analysis**
  2. **Results**
  3. **Discussion**

1. **General Discussion**

**References**

Atagi, E., & Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics, 41*(6), 509–519.

Bent, T., Atagi, E., Akbik, A., & Bonifield, E. (2016). Classification of regional dialects, international dialects, and nonnative accents. *Journal of Phonetics*, *58*, 104–117.

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114*(3), 1600–1610.

Bradlow, A.R., Clopper, C.G., Smiljanic, R., & Walter, M.A. (2010). A perceptual phonetic similarity space for languages: Evidence from five native language listener groups. *Speech Communication, 52*, 930-942.

Clopper, C. G., & Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech, 51*(3), 175–198.

Clopper, C. G., & Bradlow, A. R. (2009). Free classification of American English dialects by native and nonnative listeners. *Journal of Phonetics, 37*(4), 436–451.

**Appendix A**

**Appendix B: Key of emojis, speakers, and languages/dialects**

|  |  |  |  |
| --- | --- | --- | --- |
| **emoji** | **Speech Accent Archive speaker** | **Language family** | **Language/English dialect** |
|  | bengali\_9 | South Asian | Bengali |
|  | bengali\_13 | South Asian | Bengali |
|  | bengali\_16 | South Asian | Bengali |
|  | gujarati\_5 | South Asian | Gujarati |
|  | gujarati\_13 | South Asian | Gujarati |
|  | gujarati\_14 | South Asian | Gujarati |
|  | urdu\_2 | South Asian | Urdu |
|  | urdu\_15 | South Asian | Urdu |
|  | urdu\_27 | South Asian | Urdu |
|  | indonesian\_1 | South East Asian | Indonesian |
|  | indonesian\_8 | South East Asian | Indonesian |
|  | indonesian\_10 | South East Asian | Indonesian |
|  | tagalog\_6 | South East Asian | Tagalog |
|  | tagalog\_9 | South East Asian | Tagalog |
|  | tagalog\_18 | South East Asian | Tagalog |
|  | thai\_2 | South East Asian | Thai |
|  | thai\_6 | South East Asian | Thai |
|  | thai\_7 | South East Asian | Thai |
|  | japanese\_11 | East Asian | Japanese |
|  | japanese\_12 | East Asian | Japanese |
|  | japanese\_26 | East Asian | Japanese |
|  | korean\_2 | East Asian | Korean |
|  | korean\_24 | East Asian | Korean |
|  | korean\_30 | East Asian | Korean |
|  | mandarin\_14 | East Asian | Mandarin |
|  | mandarin\_53 | East Asian | Mandarin |
|  | mandarin\_63 | East Asian | Mandarin |
|  | english\_21 | American English | New England |
|  | english\_89 | American English | New England |
|  | english\_103 | American English | New England |
|  | english\_428 | American English | Southern |
|  | english\_212 | American English | Southern |
|  | english\_357 | American English | Southern |
|  | english\_288 | American English | Midland |
|  | english\_171 | American English | Midland |
|  | english\_126 | American English | Midland |
|  | english\_3 | International English | Australian |
|  | english\_73 | International English | Australian |
|  | english\_153 | International English | Australian |
|  | english\_2 | International English | British |
|  | english\_38 | International English | British |
|  | english\_460 | International English | British |
|  | africaans\_2 | International English | Africaans |
|  | africaans \_5 | International English | Africaans |
|  | africaans \_42 | International English | Africaans |

Appendix C

1. Note that none of the /ɹ/ are in an intervocalic position, where an [ɹ] is expected to occur in this dialect. [↑](#footnote-ref-1)