Calling Stella Long Distance:

Free Classification of Regional United States Dialects, International Dialects and Asian Nonnative Accents by Listeners from a Diverse Demographic

Joy Lu

Dr. Kristen Syrett

Rutgers University

Acknowledgments

I would like to thank my advisor Dr. Kristen Syrett for her guidance, encouragement and support throughout the entire process. I would also like to thank the Rutgers Linguistics Department at Rutgers University, New Brunswick for the opportunity to complete this thesis. Thank you to Jason Geller from the Rutgers Center for Cognitive Science for the statistical analyses. I would like to thank my fellow thesis students Karen and Nate for their never ending support and camaraderie. Last but not least, a big thanks to my fellow research assistants at the Laboratory for Developmental Language Studies for their help and emotional support, and to our lab coordinator Meg Gotowski for helping with organizing and delegating tasks related to this thesis.

Abstract

People need language for communication. Communication takes a variety of forms and characteristics, including verbal communication, body language, facial expression, prosody, pitch, and many more. Figuring out essential aspects of the language learning process can help improve learning systems and thus facilitate a language learner's experience. Indexical properties like cultural norms, social context, tone, and dialect are also important in understanding a speaker. Sociolinguistic knowledge can affect speech intelligibility and speech processing; in any given language or location, there are cultural and dialectal variations that exist. These variations could range from vowel changes, to consonant insertions and deletions, to dialect specific phrases. Previous research from Bent et al. (2016) shows that listeners can differentiate between native and nonnative talkers in a free classification task. The current study uses a free classification task to examine participants with exposure to and language backgrounds of various asian languages's ability to categorize regional and international dialects, as well as nonnative accents. Speech samples that were categorized in the task included three United States regional dialects, three international dialects, and nine nonnative accents. We found evidence that listeners exposed to diverse languages are able to differentiate between native English dialects and nonnative Asian accents and further between South Asian talkers and South East Asian and East Asian talkers.

1. Introduction

Imagine you are watching a soap opera on TV and a new character, who was born and raised in London, is introduced. By the way the character speaks, can you tell that they did not grow up in the United States? Maybe you can even tell that they are from the UK, or that they are from London. How are you able to extract this information just by hearing a couple of scripted lines from a TV show? Listeners are able to obtain many cues about a talker's identity, from acoustic cues like vowel and consonant qualities to indexical cues relating to age, and gender (Bent et al. 2016; Clopper & Bradlow 2009). Not only can listeners register these cues, they are also able categorize talkers at above chance level based on indexical properties like gender and dialect (Clopper & Bradlow 2008).

Learning a language involves the familiarization with and eventually the acquisition of the phonological system, phonological patterns, and prosodic structures (Clopper & Bradlow 2009), along with meaning, sentence structure, and sociocultural cues. The impact of sociocultural cues in the speech stream has long been studied with native and nonnative populations. A question that remains to be answered is what factors affect a listener's ability to perceive acoustic cues in speech: is it language background, language exposure, or is it something else? Previous research has shown that adults acquiring a second language from the same first language background do so in a consistent way, but those with different language backgrounds do so differently from the target language norms (Bent & Bradlow 2003). If listeners, both native and nonnative, are able to consistently differentiate between various languages and dialects, there must be certain cross-linguistic variations and cues within the input that they are tuning in to. In an increasingly global and linguistically diverse world, being able to

identify features of various dialectal and accented differences can vastly improve the second language learning process (Bradlow et al. 2010).

1.1 Vowel and Consonant Variations in Dialects and Accents

One distinguishing indexical feature of a talker is the speaker's dialect or accent. Although there are some discrepancies dependent on geopolitical factors that influence the definition of a language versus a dialect, a dialect is a variety of a language that is typically based on geographic location or region. Consonant and vowel qualities are defining features within speech that help listeners identify dialect and accents, while also helping with speech intelligibility for nonnative listeners. Examples of vowel and consonant variations between dialects are discussed in Clopper & Bradlow (2009) when examining dialectal differences of U.S. regional dialects using audio files from the TIMIT Acoustic-Phonetic Continuous Speech Corpus. In order for learners of English to process and comprehend English, they need to be able to be able to perceive the differences between monophthongs and diphthongs that exist in English, like the difference between [a:] and [aj], and that r-lessness and presence of r in the word "wash" is a dialectal difference and not producing a different word (Clopper & Bradlow 2009).

Results from earlier research indicate that perceptual sensitivity to phonemic and subphonemic differences within a speech stream is strongly influenced by a listener's language background, as well as the relationship between the listener's first and second languages (Best 1995; Clopper & Bradlow 2009). The effects of native language on perceptual sensitivity can provide great insight on the adult language learning process, as well as address difficulties listeners have as nonnative learners when it comes to speech intelligibility. The phonetic distances between languages demonstrate patterns in perception as well as production of cross-

language and second-language communication (Bradlow et al. 2010). The outcome of Clopper & Bradlow's (2009) study suggests that nonnative listeners can reliably use the acoustic properties of consonantal categorical differences to distinguish between dialects, for example between the [s] and [z] in "greasy". However, the clustering analyses in Experiment 2 of Clopper & Bradlow (2009) showed that the nonnative listeners did not pattern similarly to native listeners and did not attend to the diphthongization of /aj/ to classify the talkers. The participants of Clopper & Bradlow (2009) were monolingual English speakers and nonnative English speakers recruited from Northwestern University in Illinois, a location with less language diversity than the current research which took place on the Rutgers New Brunswick camps in central New Jersey. This brings up the question of the effects of language exposure, not just language background, on the ability to perceive phonemic and subphonemic differences and distinguish between dialects and accents.

1.3 Speech Intelligibility

Speech intelligibility is affected by a wide variety of factors that include vowel and consonant quality, individual speaker differences, the language background of the listener, noise levels, and many others (Clopper & Bradlow 2008). Bent & Bradlow (2003), investigated the phenomenon of "interlanguage speech intelligibility benefit": the benefit that comes from a shared interlanguage between a non-native talker and the listener. This phenomenon was tested experimentally by Bent & Bradlow (2003) in which non-native speakers of English with native Chinese, Korean, Bulgarian, Dutch, French/Douala, German, Greek, Hindi, Japanese, Serbian, Spanish, and Tamil language backgrounds produced a list of sentences, which was then played into headphones for participants. The participants were asked to listen to the sentence stimulus and write down what they heard on answer sheets. The results of the study showed that for native

listeners, there was greater intelligibility for native talkers than for nonnative talkers. Meanwhile, the nonnative listeners provided evidence for "matched interlanguage speech intelligibility benefit" namely, that intelligibility of a nonnative talker from the same background as the listener was equal to that of the intelligibility of the native talker. The nonnative listeners also exhibited "mismatched interlanguage speech intelligibility" in which the intelligibility of a nonnative talker from a different language background was greater than or equal to the intelligibility of the native talker. The question arises, then, how will speakers who are not necessarily multilingual, but are routinely exposed to multiple languages, perform in auditory perception tasks in which they are asked to listen to native and nonnative talkers.

1.4 Explicit Categorization Judgments

Other methodologies have been used to investigate dialect differentiation and social attitudes related to language, such as the explicit categorization judgment task used by Wagner, Clopper, & Pate (2014). Child participants, with the average age of 6, were presented with two dialect samples and instructed to categorize the samples into groups. The auditory stimuli used were a home dialect, or U.S. dialect, a regional dialect, or international dialect, and a second-language dialect, an international population with another first language alongside English. The majority of participants were born and raised in Ohio, and the rest were from neighboring states; all of the participants had little to no exposure to other languages and dialects. Children were able to distinguish between the home dialect and second language sample, but they did not perform above chance for home dialect vs. regional dialect or regional dialect vs. second language dialect. Although this methodology is effective in measuring perception and ability to categorize based on that perception, it lacks flexibility for larger groups of stimuli.

1.5 Free Classification Task

A free classification task asks participants to listen to auditory stimuli and then group the stimuli into groups according to their perception of the stimuli in accordance to the constraints of the study. An advantage to using a free classification task is that this task can help eliminate bias of an a priori geographic framework or structure of constraints set by the experiment (Atagi & Bent 2013; Bent et al. 2016; Clopper & Bradlow 2009; Clopper & Pisoni 2007). There have been some variations in previous research that involved free classification tasks with U.S. regional dialects only (Clopper & Bradlow 2009; Clopper & Pisoni 2007), nonnative accents only, and U.S. regional dialects, international dialects, and nonnative accents.

Clopper & Bradlow (2009) tested participants with stimuli from the TIMIT Acoustic-Phonetic Continuous Speech Corpus. The stimuli included speech samples from four U.S. regional dialects: New England, North, Midland, and South. The participants were given the speech samples and told to organize them into groups based on how similar they sounded with no limitations on number of groups or number of samples per group. The participants themselves consisted of three groups: native speakers of English, nonnative speakers of English with various first languages, and nonnative speakers of English with Mandarin as their first language. The overarching results showed that nonnative participants performed with a decent amount of accuracy with consonant variations, but for a speech sample with individual talker differences, only the native listeners categorized correctly. This study thus provides evidence that native listeners were able to use indexical properties of the speech sample that nonnatives weren't able to use to aid in categorization. With vowel variations, nonnative listeners had a more difficult time categorizing, especially the heterogeneous group of nonnative listeners.

While Clopper & Bradlow (2009) tested exclusively for U.S. regional dialects, Atagi & Bent (2013) focused on nonnative accents. The stimuli used were from the Hoosier Database of Native and Non-native Speech for Children. The study included 28 talkers from the native backgrounds of American Midland dialect English, French from France, German from Germany, Spain from Colombia, Japanese, Korean from Seoul, and Mandarin from Beijing. The participants of this study were adult monolingual English speakers recruited from the Indiana University campus. The results of the task showed that participants were able to separate the audio stimuli into broad categories belonging to different regions. It could be that the talkers could have been judged by the participants based on the number of phonological variants perceived as differences from English that resulted in the categorizations. As this study was conducted with monolingual English speakers with relatively limited language exposures, the results may vary if the participants of a free classification study were exposed to spoken English that contains a wider variety of phonological variation.

Bent et al. (2016) tested for six U.S. regional dialects, six international English dialects, and nonnative twelve accents with 72 talkers from the Speech Accent Archive. The dialects and accents included Mid-Atlantic, Midland, New England, North, South, and West U.S. regional dialects; Australia, England (London), Ireland, New Zealand, Scotland, and South Africa international English dialects; Arabic, French, German, Gujarati, Japanese, Korean, Mandarin, Russian, Spanish, Somali, Swahili, and Thai nonnative accents. Three talkers were selected for each dialect or accent and only male talkers were selected. The target sentence chosen from the complete utterance was, “Please call Stella. Ask her to bring these things with her from the store.” This study examined whether or not a larger set of auditory stimuli would make a big difference on native and nonnative listener performance using a free classification task and a

ladder task following the classification task with the same participants. For the ladder task, participants were instructed to use standard American English as a baseline and arrange the speech samples vertically according to closeness to standard American English. The free classification showed that listeners are able to distinguish between native and nonnative speakers, as the results appeared in separate clusters in the cluster analysis. The native English talkers were split into two clusters, with the first cluster containing the U.S. regional dialects, New Zealand talkers and Ireland talkers, and the second cluster containing international dialects of Scotland, Australia, England, and South Africa. The nonnative accents were grouped in a large cluster with three branches: one with French and German, and Asian branch with Korean, Japanese, Mandarin and Thai, and a third branch containing the rest of the nonnative accents. This shows that while the native and nonnative talkers were perceptually distinct from one another, there were not many distinctions within the Asian nonnative accents or within the large branch of "other" nonnative accents.

The flexibility that exists for this methodology provides a space for the participant to sort without constraints, and for patterns generated across participants to be more representative of the organization of perceived sounds in listeners. There are limitations, however, to this methodology. The task is a taxing task for participants, and if participants are not focused for the entire task, sorting carefully according to their perception, it is very easy for the data to not show any clustering patterns at all. It is also difficult to see if there are any trends within the data with the raw data before the clustering analysis is performed.

1.7 Remaining Questions and Current Research

Previous research looked at free classification tasks with native and nonnative listeners categorizing U.S. regional dialects in which the nonnatives were separated into a heterogeneous

language background group and a homogeneous Mandarin native language group (Clopper & Bradlow 2009). Studies were also conducted to investigate the effect of a larger auditory stimulus sample, including twelve dialects and twelve nonnative accents (Bent et al. 2016). The current research aims to explore the effects of language exposure on participants living in a diverse language community and of language background on the ability to differentiate and categorize accents from East Asia, South East Asia, and South Asia. According to the 2010 United States Census Bureau, the population in Middlesex County, New Jersey, the county that Rutgers New Brunswick is in, is 60.1% white, 12% Black or African American, 0.7% American indian and Alaska Native, 24.9% Asian, 0.1% Native Hawaiian and other Pacific Islander, and 22.1% Hispanic or Latino. The United States Census Bureau also has "Detailed Languages Spoken at Home and Ability to Speak English for the Population 5 Years and Over: 2009-2013" information available, which indicates that about 70% of the population speaks only English at home, and 30% of the population speak a language other than English at home. The Rutgers New Brunswick population is 37.5% white, 24.9% Asian, 12.3% Hispanic or Latino, 7.27% Black or African American, 3.34% two or more races, 0.169 % Native Hawaiian or Other Pacific Islander, and 0.0518% American Indian or Alaska Native[1]. Along with the demographic information, the Research Advisory Council, as a part of the School of Arts and Sciences Language Engagement Project administered a university wide multilingualism survey at Rutgers. The survey results showed that of the students and faculty surveyed, 59% understand and/or speak a language other than English, 63% considers themselves to be a heritage speaker of a language other than English, and 84% studied a language other than English in the past. With such a diverse demographic on top of high percentages of multilingualism at Rutgers New Brunswick, there is

[1] Retrieved from: https://datausa.io/profile/university/rutgers-university-new-brunswick

a higher likelihood that participants from the area will have had exposure to the target Asian accents.

In the modern world, it's almost uncommon, especially in metropolitan areas, to find a monolingual speaker who has not had any type of second language or third language exposure. With a large majority of the population self-identifying as heritage speakers, bilingual speakers, or multilingual speakers, for those acquiring languages other than their native language, understanding the cues and components of speech that people are attuning to can improve the way we learn.

2. Classification Task

The study was initially proposed to be administered in person, and in both adult and children populations, but due to the restrictions of the COVID-19 pandemic, the experiment was adapted to become an online study for adult participants only.

2.1 Participants

95 undergraduate linguistics students of varying language backgrounds were recruited from the Rutgers University Campus in New Brunswick, NJ. 3 participants were excluded for taking less than 20 minutes total on the task, and 3 participants were excluded for being research assistants in the same lab. All of the participants were speakers of English: there were 44 monolingual English speakers, 7 East Asian language native speakers, 7 East Asian language heritage speakers, 3 South Asian language native speakers, 5 South Asian heritage speakers, and 23 speakers of nontarget languages.

The language backgrounds of the participants were collected through a Google Form completed prior to the start of testing. The participants were asked: "What languages do you speak with native or heritage fluency? Please list them out", "Of the languages you listed above,

at what age did you start learning them, and where did you learn it from?", and "How long have you lived in the U.S.? Have you lived in other countries, and for how long?". These questions were chosen to determine participants' language background as well as the level of fluency they possess. The demographic information was then cross-checked with the prescreening data received through the Rutgers Linguistics SONA testing system. The undergraduate students who participated received extra credit for their undergraduate linguistics classes.

2.2 Experimental Design

The experiment was administered through Google Slides and distributed online through a link. The platform was chosen and adapted to accommodate for a completely virtual study in the midst of the global pandemic. Previous free classification tasks were administered in the respective labs that ran the experiments, and this cut out any online trace of the participant to their data. For a virtual study to accommodate the audio files and editing arrangement capabilities for participants, the researchers devised many iterations of the design and procedure of the study before deciding on the present one. Each participant received an individualized Google Form linked with an individualized Google Slides. It is essential for the researchers to create a procedure that would ensure the anonymity of the participants while allowing them to edit the task virtually with a way to reliably save and code the results.

2.3 Stimuli

2.3.1 Auditory Stimuli

45 speech samples were selected from The Speech Accent Archive. The talkers included three American English regional dialects, three international English dialects, and nine nonnative accents. The nonnative accents were split into three accents from East Asia, three accents from South Asia, and three accents from Southeast Asia. The American English dialects included the

New England dialect, the Southern dialect, and the Midland dialect. The international English dialects included British English, Australian English, and Africaans. The native languages of the nonnative-accented talkers were Mandarin, Korean, and Japanese from East Asia, Bengali, Gujarati, and Urdu from South Asia, and Indonesian, Tagalog, and Thai from Southeast Asia.

The speech samples were edited to select a shorter utterance from the full utterance provided by the Speech Accent Archive. The purpose of using a shorter sample was to lower the cognitive load for the participants, although the chosen segment for this study is longer than in previous research that used the same source (Bent et al. 2016). The segment selected was, "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." There were three talkers selected that represented each dialect or accent, and the genders of the talkers were balanced across all of the talkers. A full list of the talkers used can be found in **Table 2** in **Appendix B**.

2.3.2 Visual Stimuli

The visual stimulus materials included a Google slide that included the speech samples as moveable and clickable icons represented by emojis to the left of an open grid layout. The icons played the speech sample when double clicked, and the icons could be dragged across the screen with a mouse. The target sentence was included at the bottom left corner of the slide as reference. A sample grid is shown in **Figure 1**. The emojis were arbitrarily assigned by the researchers to the selected audio files and a key was for the researchers to keep track of which emoji corresponded to which talkers (**Figure 4**).

Figure 1. Sample Classification Task Template

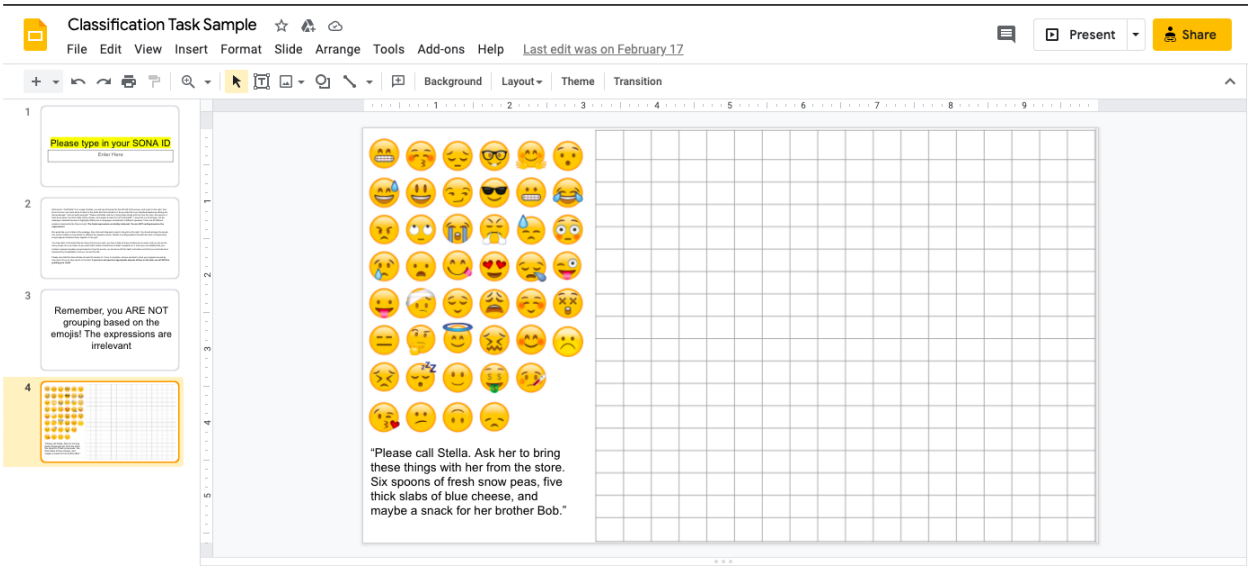


Figure 2. Classification Task Audio

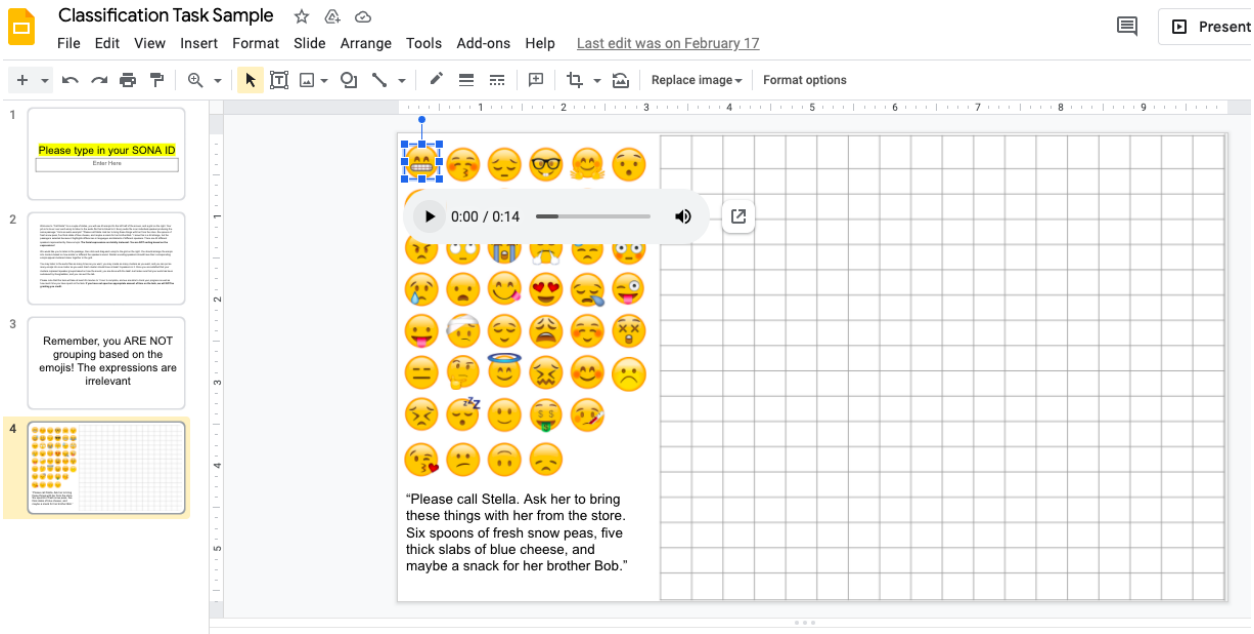


Figure 3. Classification Task Sample Groupings

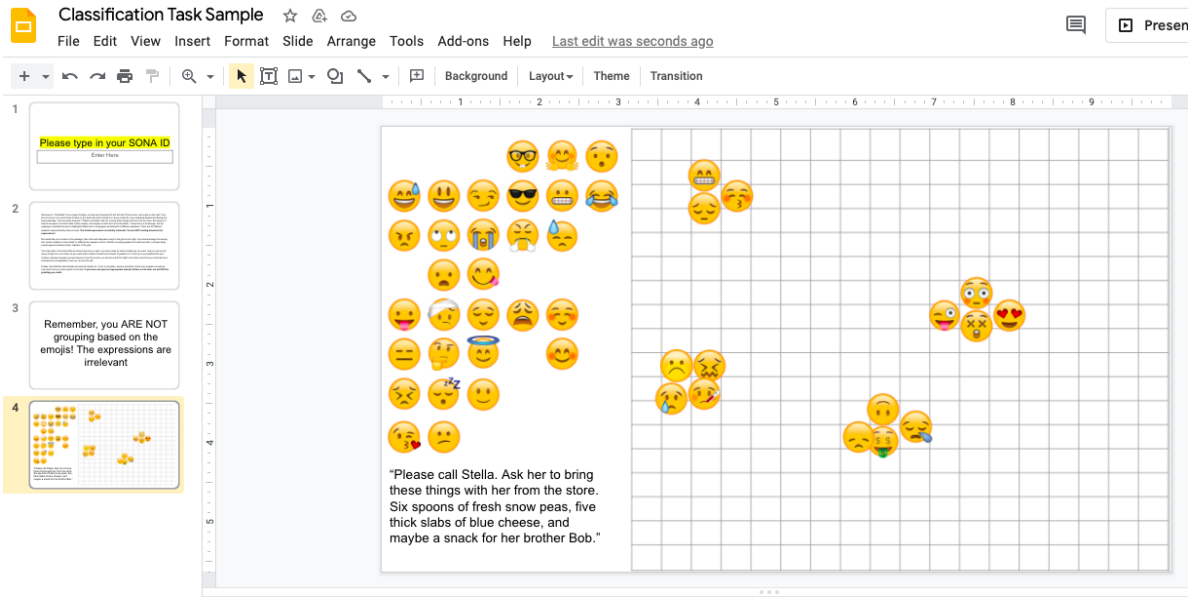
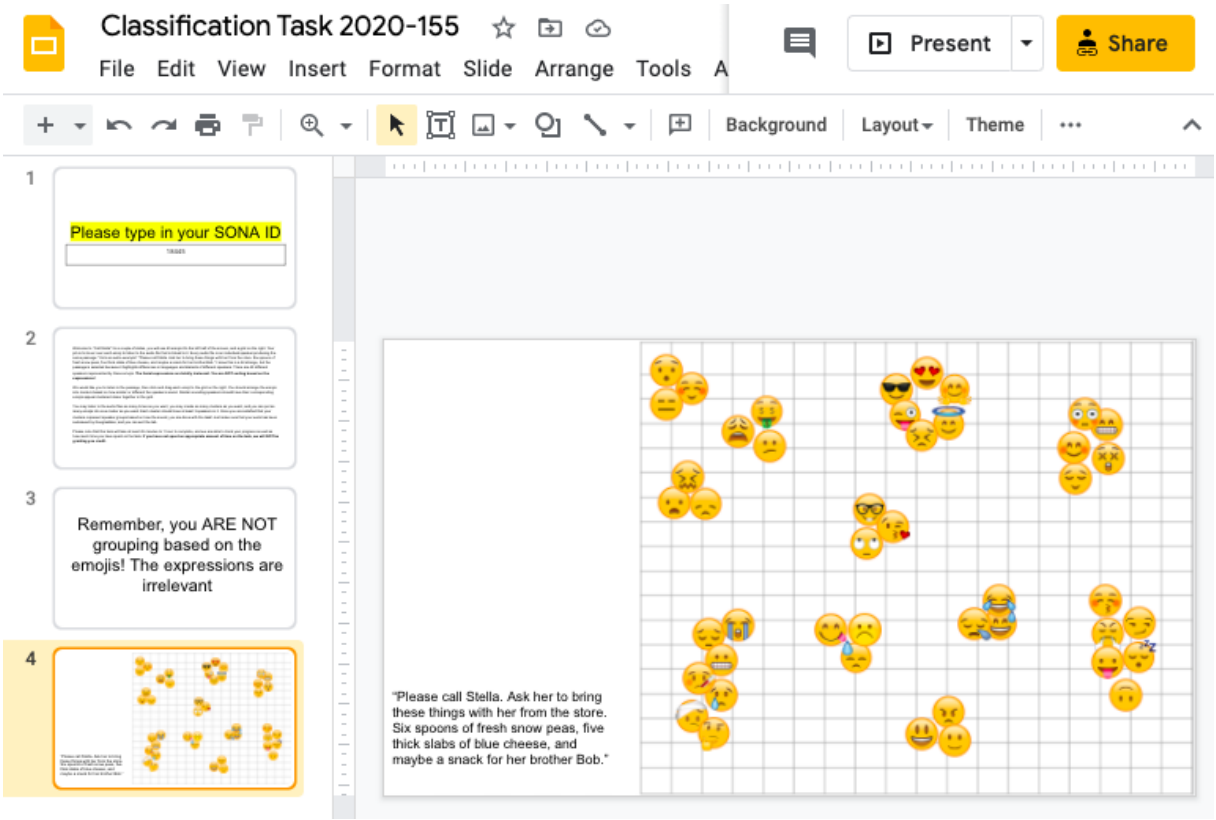


Figure 4. Completed Classification Task



2.4 Procedure

Prior to the start of the experiment, the participants completed a Google Form that contained a consent form, language background questions, and a video tutorial on how to complete the task. The script for the instructions video on the Google Form presented to the participants is in **Appendix A**. At the bottom of each Google Form was a link to an individualized Google Slides copy of the classification task. The experiment was distributed online to the participants through individualized Google Form links. The first slide contained a text box for participants to input their SONA ID number so that the researchers could grant them credit through the Sona System. The participants were instructed through auditory instructions in the video and written instructions on the second slide of the Google Slides to listen to the icons and drag them into groups based on the talkers's geographic location and proximity. The third slide is a reminder that says, "REMEMBER, you ARE NOT grouping based on the emojis! The expressions are irrelevant". Participants could listen to the samples as many times as they needed to, and could organize the samples into as many groups as they wanted to.

After the participants were satisfied with the arrangement they came up with, they were instructed to check that their edits have been auto-saved by the Google Slides, and then to exit the window. From there, the researchers screenshotted the Google Slides screens after the time slots for the study closed and compiled the screenshots into folders on Box to which all of the research assistants had access to. Each Google Slide was labeled with a lab ID number so that the researchers could keep track of which screenshot corresponded with which participant.

Two rounds of data collection were conducted with one round in the Fall semester, and one in the Spring semester. There were no changes made to the study between these two rounds of data collection, and there were no repeated participants for the two rounds. The participants in

the free classification task were also not involved in the Qualtrics survey task that was administered as a post hoc.

2.5 Results

2.5.1 Data Treatment and Analysis

Data coding was performed using Microsoft Excel. The data collected was in the form of screenshots; one screenshot per participant. The response entry form used to record the participant data contained a column with the emoji, a column with the emoji description, a column for the speaker the emoji corresponds with, and a column for the assigned category (**Figure 4**). The procedure for organizing the classification task was made available to research assistants on Box. The exact instructions presented to the research assistants are available in **Appendix B**. To log the data into the response entry documents, each cluster of emojis was arbitrarily assigned a number starting from 1. After coding the clusters into the response entry document, the cluster numbers were copied and pasted into an Excel spreadsheet that contained the cluster results of all of the participants.

Many of the screenshots exhibited clear distinguishable clusters of emojis, but there were 3 that were difficult to sort. Each screenshot has been categorized with the response entry document and coded again for reliability purposes by a second research assistant. The clusters were determined and judged by relative distances of the emojis from each other on the grid. Screenshots with ambiguous groupings were discussed amongst the researchers and a consensus was reached on how to best code the groupings. There were three participants with no distinct emoji clusters whose screenshots were discarded and not included in the final analysis. One more participant's data was discarded when running the hierarchical clustering analysis due to a coding issue.

Figure 4: Response Entry Template for Classification Task

emoji	emoji_description	speaker	categories
	shy_blush	bengali_9	
	peaceful_smile	bengali_13	
	big_frown	bengali_16	
	blush_kissy	gujarati_5	
	tongue_out	gujarati_13	
	snoozing_sleeping	gujarati_14	
	smirky_face	urdu_2	
	huff_puff	urdu_15	
	upside_down_smiley	urdu_27	
	angry_face	indonesian_1	
	thinking_face	indonesian_8	
	ironic_smiley	indonesian_10	

Figure 5: Screenshot of Sample Response Entry

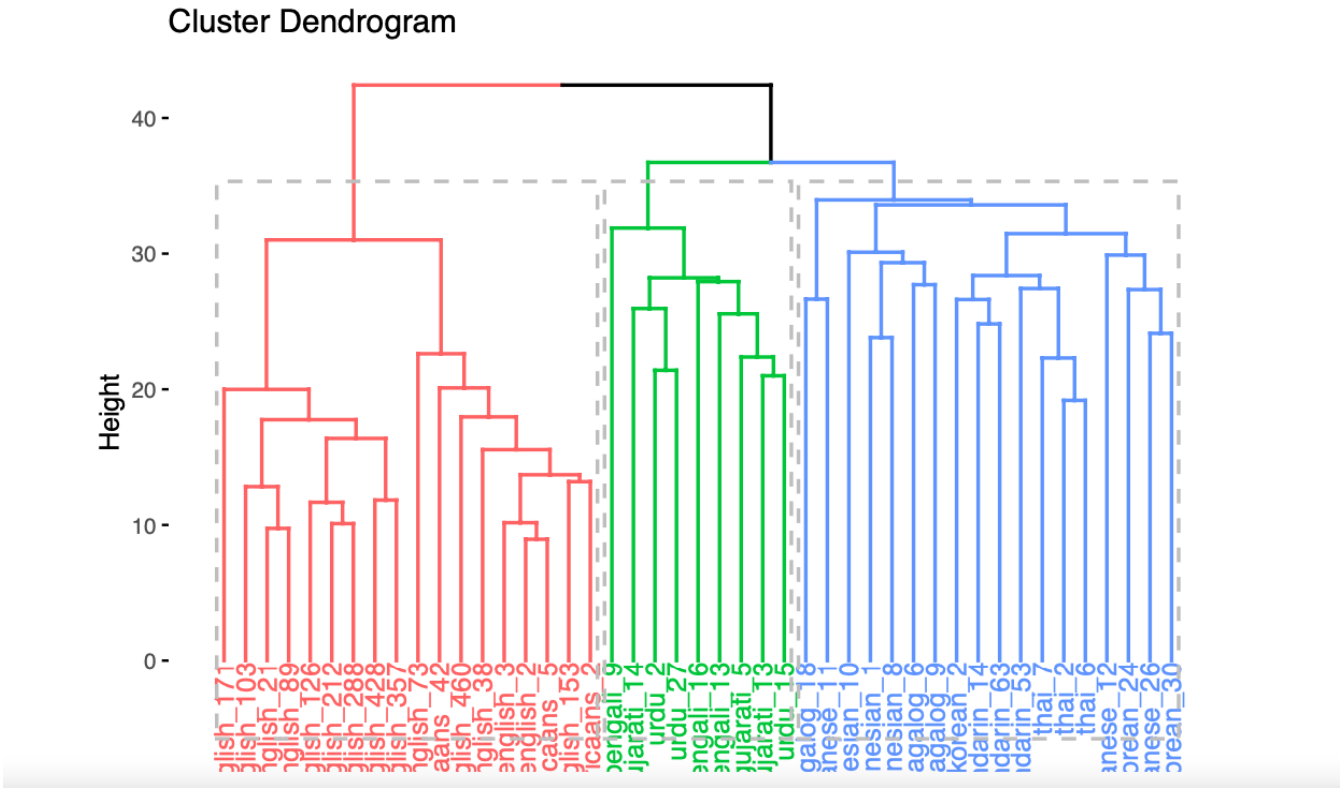
emoji	emoji_description	speaker	categories
	shy_blush	bengali_9	2
	peaceful_smile	bengali_13	2
	big_frown	bengali_16	2
	blush_kissy	gujarati_5	2
	tongue_out	gujarati_13	2
	snoozing_sleeping	gujarati_14	2
	smirky_face	urdu_2	2
	huff_puff	urdu_15	2
	upside_down_smiley	urdu_27	2
	angry_face	indonesian_1	1
	thinking_face	indonesian_8	6
	ironic_smiley	indonesian_10	1

2.5.2 Classification Task Results

Hierarchical clustering was performed on the organized data to examine the natural clusters through the free classification task. The data was clustered using average link clustering: it computes all of the pairwise dissimilarities between the elements and uses the averages of the

dissimilarities to represent distance between the formed clusters[2]. The dissimilarity matrix was calculated using euclidean distance, and the cluster solution was then plotted. In the resulting dendrogram, the individual labels are names of the audio files, and based on similarity, they form branching as the tree moves up. The height provided on the vertical axis demonstrates dissimilarity between the audio files: the higher the height of the connecting branching, the more different they are from each other. A k-means analysis, another clustering algorithm, was then performed to determine the best number of clusters. The dendrogram cut into 3 clusters is shown below in **Figure 6.**

Figure 6: Cluster Dendrogram



2.5.3 Discussion

[2] The data analyses and resulting diagrams were performed, created, and compiled by Jason Geller from the Rutgers Center for Cognitive Science.

The data shows an overall trend of participants having the ability to distinguish between U.S. regional dialects, international English dialects, South Asian and South East Asian/East Asian accents. The ability to distinguish between native dialects and nonnative accents aligns with previous research (Bent et al. 2016). A concern at the start of the experiment was that participants would not take the task with focus and care, resulting in random clusters without any type of pattern. The results show, however, that the listeners in the experiment showed consistency in the way they organized, and patterns appeared based on talker origin.

3. Sound File Analysis

3.1 Method

To analyze the selected auditory stimuli, we took a two pronged approach. The target segments for the selected speech sample were identified by the researchers, and then analyzed with Praat by running a Praat script[3]. The first, second, and third formants were obtained after running a script and compared across the speech samples to demonstrate distinct acoustic differences between the dialects and accents. We decided to use an acoustic analysis to pinpoint phonemic and subphonemic differences between the dialects and accents to determine which differences the listeners were attuning to while taking the free classification task.

On top of the acoustic analyses run through Praat, the research assistants of the lab also listened to the selected sound files and coded in a Microsoft Excel spreadsheet the presence of cluster reduction, epenthesis, plurals, and rhoticity in the target segments of the file. The coded excel files were then compared and agreed upon by the research assistants into the compiled Excel spreadsheet.

3.1.2 Procedure for Praat Analysis

[3] The Praat script was written by Professor Adam McCollum, an Assistant Professor in the Department of Linguistics at Rutgers University.

Each selected audio file was uploaded back onto Praat and an accompanying textgrid was created. A key was made separating out the targeted segments for reference while the research assistants segmented. We needed to target the same segments across speakers, and International Phonetic Alphabet (IPA) would result in different annotations for the different dialects and accents. We chose representative letters for each respective segment so that the Praat script can consistently analyze the same target segments, and whether or not the letters were uppercase or lowercase was arbitrarily decided. The full instructions for the research assistants in Praat segmentation, as well as the key that was used to base segmentation off of are available in **Appendix E**. Files were randomly selected for reliability coding to check for discrepancies. Following segmentation, the audio files and textgrid files were separated out into folders for male and female speakers in preparation for running the files through Praat.

3.1.3 Procedure for Impressionistic Listening Task

The research assistants were assigned to listen to all of the sound files in the classification task and code “yes” or “no” if they heard epenthesis, plural reduction, cluster reduction, or rhoticity in the files. Each acoustic feature was assigned 3 research assistants to listen for it and plural reduction was assigned four. From the full length audio file, the segments targeted for epenthesis were between “call-Stella”, “ask-her”, “with-her”, “six-spoons”, “of-fresh”, “fresh-snow”, and “five-thick”. The segments targeted for plural reduction included “things”, “spoons”, “peas”, and “slabs”. The segments targeted for cluster reduction included “bring”, “spoons”, “blue”, and “brother”. Lastly, the segments targeted for rhoticity included “store” and “brother”. The results from the research assistants were then consolidated and averaged for a percentage that can be found in **Table 1**.

3.2 Results

3.2.1 Praat Analysis Results

The results are currently in the process of being obtained through script running and consolidation on Excel. The obtained values from each sound file, the midpoints of the first formant, second formant, and third formant will be averaged and a z-score will be obtained for each talker, then compared with normalized z-scores across talkers.

3.2.2 Impressionistic Listening Task Results

The consolidated data is shown in **Table 1**.

Table 1.

		average surface-level characteristics perceived			
		epenthesis	plural reduction	cluster reduction	rhoticity*
English					
American	Midland	0.0%	0.0%	0.0%	100.0%
	New England	0.0%	0.0%	0.0%	0.0%
	Southern	0.0%	0.0%	0.0%	83.3%
International	Africaans	0.0%	0.0%	0.0%	33.3%
	Australian	0.0%	0.0%	0.0%	50.0%
	British	0.0%	0.0%	0.0%	50.0%
Asian					
East	Japanese	16.7%	8.3%	0.0%	100.0%
	Korean	27.8%	8.3%	0.0%	100.0%
	Mandarin	0.0%	0.0%	0.0%	83.3%
South Asian	Bengali	16.7%	0.0%	0.0%	100.0%
	Gujarati	0.0%	0.0%	0.0%	100.0%
	Urdu	5.6%	0.0%	0.0%	100.0%
South East Asian	Indonesian	5.6%	0.0%	0.0%	100.0%
	Tagalog	11.1%	0.0%	0.0%	100.0%
	Thai	0.0%	33.3%	41.7%	50.0%

*in two places: *(from the sto)re* and *(her broth)er (Bob)*

3.2.2.1 Discussion

The percentages show that there are definitely certain distinguishing acoustic cues that listeners are focusing in on. From the table, we can see that the listeners picked up on perceptible cues that differentiate native talkers from nonnative talkers, and these cues also help with some distinctions within the nonnative Asian accents. The Thai talkers are the most distinct out of the

nonnative accents, demonstrating plural reduction, cluster reduction, as well as rhoticity, while the rest of the Asian talkers patterned without plural and cluster reduction, but included epenthesis which Thai talkers did not exhibit. The rhoticity cue is the most helpful cue in distinguishing between native dialects, while the other cues provide a basis to differentiate between native and nonnative talkers. There are certain phonological processes that are useful to listeners in differentiating native form nonnative talkers, and with the results to come from the Praat script analysis, the rhoticity results may be highlighted more for distinguishing between dialects.

4. Post Hoc Qualtrics Survey

4.1 Method

To account for any discrepancies found in the classification task data, a set of Qualtrics surveys with 22 questions each were distributed to native and heritage speakers of the nonnative accent target languages to obtain judgments on the quality of the sound files used by the experiment. There was an individual qualtrics survey for each of the nonnative accents in the classification task. With nine surveys total, there was a survey for Mandarin, Korean, Japanese, Urdu, Gujarati, Bengali, Tagalog, Thai, and Indonesian. The questions template to the Qualtrics surveys can be found in **Appendix D**.

4.1.1 Participants

The participants for the Qualtrics survey task were recruited by word of mouth, people who were unfamiliar with the project, as well as undergraduate linguistics students from Rutgers University New Brunswick. The participants who took the survey were native or heritage speakers of the target language of the Qualtrics Survey. There were 6 participants for Japanese, 11 participants for Urdu, 14 participants for Mandairn, 1 participant for Tagalog, 4 participants

for Korean, 5 participants for Thai, 3 participants for Indonesian, 7 participants for Bengali, and 7 participants for Gujarati.

4.1.2 Experimental Design

The full set of questions in the survey can be found in **Appendix D**. The task consisted of a Pick, Group, and Rank question style from Qualtrics in which the participants were presented with 8 sound files from the target accent and asked to sort them into categories of "Include these speakers" and "Exclude these speakers" from the final set of representative speakers of the target accent. The survey contained two sets of training trials that contained audio files from U.S. regional dialects and international English dialects. The participants were instructed to listen for poor sound quality, disfluency, long pauses, and production errors, and to include the talkers that sound representative of the language or dialect, and sound relatively similar to the rest of the set. At the end of the survey, the participants were asked in an open ended answer box the rationale to which they excluded speakers in the actual task.

4.2 Results

The results are in the process of being consolidated from Qualtrics.

4.2.1 Discussion

5. General Discussion

Talkers from around the world with native regional dialects and nonnative accents display a wide range of cross-linguistic variation in both vowel and consonant production and perception differences. With all of these variations from individual talker differences, to cultural and dialectal differences, to phonemic inventory differences: how do nonnative learners perceive these cues in comparison to native listeners? What components of the speech stream are native and nonnative listeners tuning in to in order to communicate efficiently? Previous research

tackles these questions using an assortment of methodologies like explicit categorization judgment tasks, free classification tasks, and ladder tasks (Atagi & Bent 2013; Bent et al. 2016; Clopper & Bradlow 2008; Clopper & Bradlow 2009; Wagner, Clopper, Pate 2014). The outcomes of this research can better help us understand perception, speech intelligibility, and the effects of language background and language exposure.

The results of previous work left various open questions, and the current research aimed to examine a few of them. Similar to Bent et al. (2016), the current study used a free classification task with auditory stimuli including U.S. regional dialects, international English dialects, and nonnative accents using the Speech Accent Archive. However, in the current study, the nonnative accents were centered around Asian nonnative accents in three areas: East Asia, South East Asia, and South Asia. The results of Bent et al. (2016) showed that native and nonnative talkers are perceptually distinct, but the nonnative clusters did not show very fine-grained branching. This opens up the question of how listeners of varying language exposures and language backgrounds would perform in categorizing nonnative accents.

In the current work, we focused on looking at the effect of language exposure and language background testing in a diverse location, Rutgers New Brunswick campus in New Jersey. Taking advantage of the diverse language communities around Rutgers, as well as the large Asian population, we took a narrower look at nonnative accents from East Asia including Mandarin, Korean, and Japanese, from South East Asia including Indonesia, Tagalog, and Thai, and South Asia including Gujarati, Urdu, and Bengali. The participants were a mix of monolingual English speakers with exposure to diverse language communities, heritage and bilingual speakers of languages from the target regions, and heritage and bilingual speakers of languages that were not targets.

The clustering analysis results showed that listeners were able to distinguish between the U.S. regional dialects and the international English dialects for the native talkers, and there were two clusters for the nonnative accents, an East Asian and South East Asian cluster, and a South Asian cluster. The participants were a mix of monolingual English speakers with exposure to diverse language communities, heritage and bilingual speakers of languages from the target regions, and heritage and bilingual speakers of languages that were not targets. The results demonstrate that participants were able to distinguish between the nonnative accents in a more precise way than participants in previous studies. This could be explained by a variety of factors. Unlike previous studies, the nonnative accents chosen for the current research was based only in Asia, so there is a possibility that the participants were able to focus more on smaller perceptual differences between the talkers because there weren't any non-Asian nonnative accents for the participants to categorize. There could also be a difference in results if the monolingual listener population data and the heritage or bilingual population data from this study were separated. It is currently unclear whether the performance shown in the clustering analysis had to do with the multilingual participants, or if language exposure really had an impact on the monolingual listeners who took the study.

A possible future direction is to use the same task and target populations with East Asian, South East Asian, and South Asian language backgrounds specifically to get more significant participant numbers. It would be that having a language background from target regions could increase the listener's ability to categorize within the region of their language background. For example if a listener spoke Mandarin, would that give them an advantage in distinguishing between the Mandarin, Japanese, and Korean samples over a listener who spoke Gujarati? Or would a listener who spoke Marathi have an advantage in categorizing Urdu, Bengali, and

Gujarati over a Thai speaker because speakers of Marathi are from South Asia? It would be interesting to see if the language backgrounds of the listeners would give them an advantage in their region, or if even knowing more languages would give an advantage overall even if the language wasn't from the target region. Another possible extension of the current research is testing monolingual English speakers from a less linguistically diverse state or region from the U.S. This way language exposure can be more controlled, and the effects will be more apparent.[4]

There are many possible alterations to the task to attempt to answer the open questions remaining, and the current results as well as the remaining analyses that are ongoing in this current research can help provide insight to guide and shape future studies.

[4] A big thank you to Professor Tessa Bent, a professor at Indiana University for discussing my results with me and for helping me with brainstorming possible future research directions.

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.

Best, C.T. (1995). A direct realist view on cross-language speech perception. In: Strange W, editor. Speech Perception and Linguistic Experience. Timonium, MD: York Press; 1995. pp. 171–204.

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.

Clopper, C. G., & Pisoni, D. B. (2007). Free classification of regional dialects of American English. *Journal of Phonetics, 35*(3), 421–438.

Creel, S. (2012). Preschoolers' use of talker information in on-line comprehension. *Child Development, 83*(6), 2042-2056.

Mccullough, E. A., Clopper, C. G., & Wagner, L. (2017). Regional dialect perception across the lifespan: Identification and discrimination. *Language and Speech*, *62*(1), 115–136.

Wagner, L., Clopper, C. G., & Pate, J.K. (2014). Children's perception of dialect variation. *Journal of Child Language, 41*, pp 1062-1084.

Weinberger, Steven. (2015). Speech Accent Archive. George Mason University. Retrieved from http://accent.gmu.edu

Appendix A: Pre-Classification Task Materials

Questions on the Google Form:

There was an individualized Google Form for every sign up and linked to it was a link to the individualized Google Slides. The answer choices in this appendix are represented by bullet points.

Section 1: Consent Form *if the participant chose not to consent, then the form closed immediately*

Please indicate below if you agree to participate in this research study:

- I DO agree to participate in this research study and will CONTINUE to the next page.
- I DO NOT agree to participate in this research study and will EXIT the study now.

Section 2: Please type in your 5-digit SONA ID *this was a mandatory question*

Section 3: Language Background

1. What languages do you speak with native or heritage fluency? Please list them out.
2. Of the languages you listed above, at what age did you start learning them, and where did you learn it from?
3. How long have you lived in the U.S.? Have you lived in other countries, and for how long?

Section 4: Please watch this instructional video before continuing to take the task

Instructional video recorded by the researcher using sample icons and self-recorded auditory stimuli

The link to the video: https://youtu.be/4rzMBc-zz6Y

Script for instructional video:

Welcome to “Call Stella”! In a couple of slides, you will see 45 emojis On the left half of the screen, and a grid on the right. Your job is to hover over each emoji to listen to the audio file that

is linked to it. Every audio file is an individual speaker producing the same passage: *click an audio example* “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.” I know this is a bit strange, but the passage is selected because it highlights differences in languages and dialects of different speakers. There are 45 different speakers represented by these emojis. The facial expressions are totally irrelevant. You are NOT sorting based on the expressions!

We would like you to listen to the passage, then click and drag each emoji to the grid on the right. You should arrange the emojis into clusters based on how similar or different the speakers sound. Similar sounding speakers should have their corresponding emojis appear clustered closer together in the grid.

You may listen to the audio files as many times as you want, you may create as many clusters as you want, and you can put as many emojis into one cluster as you want. Each cluster should have at least 3 speakers in it. Once you are satisfied that your clusters represent speaker groups based on how the sound, you are done with the task! Just make sure that your work has been autosaved by Googleslides, and you can exit the tab.

Please note that this task will take at least 45 minutes to 1 hour to complete, and we are able to track your progress as well as how much time you have spent on the task. If you have not spent an appropriate amount of time on the task, we will NOT be granting you credit.

Appendix B: Materials for the Classification Task

Full transcript for auditory stimuli available from the Speech Accent Archive:

> “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. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.”

Table 2. Full list of speakers used in the classification task

Sound File	Language	Location1	Location2
bengali_9	Bengali	South	Asian
bengali_13	Bengali	South	Asian
bengali_16	Bengali	South	Asian
gujarati_5	Gujarati	South	Asian
gujarati_13	Gujarati	South	Asian
gujarati_14	Gujarati	South	Asian
urdu_2	Urdu	South	Asian
urdu_15	Urdu	South	Asian
urdu_27	Urdu	South	Asian
indonesian_1	Indonesian	South East	Asian
indonesian_8	Indonesian	South East	Asian
indonesian_10	Indonesian	South East	Asian
tagalog_6	Tagalog	South East	Asian
tagalog_9	Tagalog	South East	Asian
tagalog_18	Tagalog	South East	Asian
thai_2	Thai	South East	Asian
thai_6	Thai	South East	Asian
thai_7	Thai	South East	Asian
japanese_11	Japanese	East	Asian
japanese_12	Japanese	East	Asian
japanese_26	Japanese	East	Asian
korean_2	Korean	East	Asian
korean_24	Korean	East	Asian
korean_30	Korean	East	Asian

mandarin_53	Mandarin	East	Asian
mandarin_63	Mandarin	East	Asian
mandarin_14	Mandarin	East	Asian
english_21	New England	American	English
english_89	New England	American	English
english_103	New England	American	English
english_428	Southern	American	English
english_212	Southern	American	English
english_357	Southern	American	English
english_288	Midland American	American	English
english_171	Midland American	American	English
english_126	Midland American	American	English
english_3	Australian	International	English
english_73	Australian	International	English
english_153	Australian	International	English
english_2	British	International	English
english_38	British	International	English
english_460	British	International	English
africaans_2	African/Africaans	International	English
africaans_5	African/Africaans	International	English
english_42	African/Africaans	International	English

Appendix C: Data Treatment for Classification Task

Step by step recording of free classification responses for each Participant

1. Open a blank 'response_entry' doc from the Response Entry folder.
2. Save 'response_entry' doc with participant name at end of results (e.g., response entry_18') in YOUR subfolder (e.g., 'done_Kristen').
3. Open the participant's screenshot of results from the 'Participant Screen Shots (Fall)' folder. Have this open in your browser window as you have the response entry_# file open, too.
4. Locate each of the emojis on the screenshot one by one, and note categories one by one in the response_entry_# doc based on the groupings on screen.
5. Save completed 'response_entry_#' doc with participant's responses in your 'done' subfolder.
6. Proceed to next participant.

emoji	emoji_description	speaker	categories
	shy_blush	bengali_9	
	peaceful_smile	bengali_13	
	big_frown	bengali_16	
	blush_kissy	gujarati_5	
	tongue_out	gujarati_13	
	snoozing_sleeping	gujarati_14	
	smirky_face	urdu_2	
	huff_puff	urdu_15	
	upside_down_smiley	urdu_27	
	angry_face	indonesian_1	
	thinking_face	indonesian_8	
	ironic_smiley	indonesian_10	

Appendix D: Qualtrics Task

1. Consent Form *mandatory question to access rest of survey*
2. Please enter your 5-digit SONA ID. We cannot give you credit without this information.
3. Welcome! In this task, you will be listening to some sound files and sorting them to decide upon a final set. This final set will include files that are of good quality and where the speakers' accents sound similar to each other and representative of that language or dialect.
4. Are you a native or heritage speaker of (nonnative accent target language)?
5. What age did you learn (nonnative accent target language), and from where?
6. Here are some issues you might encounter regarding the quality of the sound file itself:Some sound files may have poor sound quality (like too much background noise or the speaker isn't loud enough or is way too loud)
    a. The speaker may produce a disfluency (like "uh" or "um" or restarting as in "Bo...Bob")
    b. The speaker may produce a long pause between words
    c. The speaker may even produce the wrong word (like saying "from" instead of "for")
7. Here are some questions to ask yourself as you listen to the speakers:
    a. Does the speaker sound representative of that language or dialect?
    b. Do the speakers in the final set of speakers sound relatively similar (or are there any outliers)?
    c. Speakers from a given region may sound different. That's ok and very common in language variation!

d. But in this particular task, we need to arrive at a set of speakers who produce the script with a similar speech style and sound representative of that language or dialect.

8. Training

a. Let's practice listening for issues in the sound file itself.

b. Please listen to this sound file of the speaker reading the following sentences:

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.

i. Did you notice any of the following issues? Select all that apply.

1. Poor sound quality
2. Disfluency
3. Long Pause
4. Production error
5. None of the above

ii. If you clicked on the boxes for any of the above options, then you would probably not want to include this sound file.

iii. Let's try another one.

c. *Repeat of the above question with a different sound file*

d. Now we'll move on to selecting a group of speakers who sound relatively similar and whose productions sound representative of a language or dialect. We'll work on identifying a 'final set' of speakers that match this criterion. Each of these

sound files within the set should be relatively free of the kind of issues we just reviewed.

9. Please listen to each of the following sound files. Then drag and drop their corresponding numbers into the two categories to the right, indicating whether we should include them or exclude them in the final set. Try to put 4 samples in your final set!
    a. Remember to ask yourself:
        i. Does the speaker sound representative of that language or dialect?
        ii. Do the speakers in the final set of speakers sound relatively similar (or are there any outlier?
        iii. You may listen to the sound files in whatever order you like, and as many times as you like.
10. *The following question space contains 7 audio files without labels with a “Pick, Group, Rank” question type below it with 7 arrangeable icons to represent the audio files. Each file and icon are labeled chronologically 1-7; the available categories are labeled as “Include these speakers” and “Exclude these speakers”*
11. Now we're done practicing and we'll move on to the actual sound files we'd like you to listen to.
    a. Remember, we're paying attention to the quality of the individual sound files, and then trying to select a final group of speakers who sound similar to each other and who sound representative of the language or dialect. Let's continue...
12. Please listen to each of the following sound files. Then drag and drop their corresponding numbers into the two categories to the right, indicating whether we should include them or exclude them in the final set. Try to put 4 samples in your final set!

a. Remember to ask yourself:

i. Does the speaker sound representative of that language or dialect?

ii. Do the speakers in the final set of speakers sound relatively similar (or are there any outlier?

iii. You may listen to the sound files in whatever order you like, and as many times as you like.

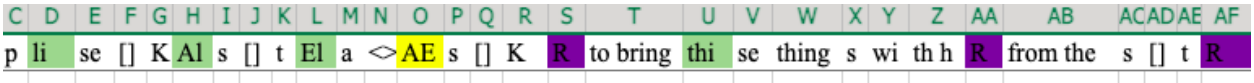
13. *Same format as question 10 with 8 sound files and 8 arrangeable icons instead*
14. Please type in the space below how you sorted the samples, and why you included certain samples and excluded others. Try to give a reason for each sample that you sorted.
    a. For example: I chose to include 5 because of the way the speaker said their "a" in slab; it sounds like an accent from the American South.
15. Thank you for completing this task for us! Your attention to detail and focus is much appreciated. If you have any questions about the task you have just completed, or about the study as a whole, please feel free to contact the lab and we'd be happy to discuss!

Appendix E: Praat Segmentation and Analysis

The Target Segments Segmented with Praat Highlighted in Yellow

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.

Snapshot of Segmentation Protocol Using Praat



Screenshot Example of Praat Segmentation

The F1, F2, and F3 values were obtained for the target segments by the Praat script.