

Sounding San Diegan: Speech Perception

Rukmini A. Ravi

University of California, San Diego

COGS 119

Professor Ayse P. Saygin

December 13, 2018

Abstract

Speech production and perception are two phenomena that provide the foundation for this study. The idea that language perception “relies on a pre-analytic store of linguistic items in memory” was utilized to develop the following expectations for this study: a) native English speakers, i.e., American English speakers, will tend to emulate the local accent of San Diego with more perceived accuracy by listeners in San Diego (having limited to high amounts of pre-analytic storage in memory regarding San Diego accents), as opposed to non-native English speakers (from different English-speaking nations other than America, e.g., U.K. and Australia), regardless of how many years, more or less, the native speakers have lived in San Diego than the non-native English speakers, and b) listeners, who have inhabited San Diego for various periods of time, will utilize prior experiences being exposed to certain linguistic stimuli at a national and local “San Diegan” level to perceive each speaker’s accents (Johnson, 1997). For the purpose of this study, two different sets of speakers’ voices were recorded, the first set being native English, i.e., American English speakers, and non-native English speakers, i.e., speakers from non-American English-speaking countries, e.g., U.K., Australia. Four listeners’ ratings of each speaker’s accent, with the main question being how “San Diegan” each accent sounded according to each listener (since San Diego was the local region at hand for this study) were recorded and analyzed in relation to the initial expectations for the study.

Keywords: speech production, speech perception, linguistic memory

Sounding San Diegan: Speech Perception

In the 21st century, individuals of all ages in America are becoming acquainted with the reality of living in a globally connected society and the implications of this phenomenon. With a global, inter-connected framework and culture comes a greater responsibility to become informed about not only different languages and cultures, but also, the ways in which a variety of languages and cultures are perceived from different avenues and mindsets. Living in San Diego, a city among others in California that is revered for its diverse ethno-linguistic population, it is important to begin research at a local level and take note of how the local San Diego accent is perceived and characterized. There are many variables that contribute to the formation of how one perceives the San Diego accent, two such variables being the bilingual culture fostered by many of San Diego's residents and the overall role the city plays in the global "innovation economy" ("Population"). According to the San Diego city website, "40.8% of the population, ages 5 or older, speaks a language other than English at home. (That's nearly double the national average of 21.1 percent)" ("Population").

Andrew Shibata, a graduate student who was a part of our group for this study, and Betina Karshaleva, an undergraduate student in our group, both currently work in a lab at UC San Diego that focuses on research concerning linguistics and cognition. Andrew created this study for us all to be a part of and informed us about relevant texts for our study at hand. Due to Andrew's prior research, our group became acquainted with the idea of pre-analytic storage of linguistic patterns in memory, after familiarizing ourselves with an Ohio State University linguistics paper titled "The auditory/perceptual basis for speech segmentation" by researcher Keith Johnson, that presents the idea of internal representations of speech and speech patterns being acoustically embedded into one's storage of linguistic memory (Johnson, 1997). Thus, the overarching objective of this study is to provide insight into how the San Diego accent is perceived and characterized based on the results from our study's participants listening to audio samples from three native, American English speakers and three non-native English speakers from English-speaking countries and providing their ratings regarding how "San Diegan" the

accents in each of the audio samples sounded. We believe our preliminary results can be used to provide scope for further research regarding how listeners internally represent and characterize speech.

Although it is the case that the number of bilingual individuals in San Diego significantly outweighs the national average, our group, coming into this study, still came in with the expectation that an American English speaker, even if not specifically from San Diego, could easily emulate a so-called “San Diegan” accent with more perceived accuracy by the listener of the audio samples, as opposed to a non-American English speaker, based on the notion that living in America and being familiar with American English provides the speaker with more personal insight than a non-American English speaker into what types of local accents there are nationwide and how they can be emulated, based on proximity to regions such as San Diego. However, the accuracy of the listener’s ratings for each of the audio samples depends on how long they have lived in San Diego and whether ample time has been allotted to providing them with a reasonable internal representation of the “San Diegan” accent.

Design

The design of the process we were planning on implementing for our study and collection of results was split into two parts or plans. The first plan entailed figuring out how and when to collect the external stimuli, i.e., the audio samples, and the second plan entailed planning the different parts of the script we were going to code through the computing environment MATLAB. We made sure that we had a plan for the code to be written in MATLAB by focusing on the main, central parts of our program first, as opposed to thinking about the small details such as how many times the program needed to be looped through depending on how many participants wanted to take part in the preliminary trials of our study.

We first determined that it would be easiest for Andrew to collect the audio samples for the non-native English speakers, from English-speaking countries, since his lab mates are from countries such as the U.K., and I volunteered to collect the audio samples for the native, American English speakers. We initially decided on recording five native, American English speakers and five non-native English speakers, but had to reduce the number of recordings so that we keep three native, American English speakers’ recordings and three non-native English speakers’ recordings instead (this will be discussed in

further detail in the “Implementation” section of this paper). These audio samples were collected as .wav files and were shared with each member of the group via Google Drive. Subsequently, we decided as a group how we were going to split off the tasks for writing the actual code that was going to be utilized for the script in MATLAB. We split off the tasks based on our respective interests, experience with using MATLAB, and favorite parts of the study in terms of the code that was going to be implemented. I volunteered to read each of the audio files into MATLAB, and write the associated code required for the audio files, e.g., creation of cell arrays and structs, and the writing of loops (which will be discussed in further detail in the “Implementation” section of this paper). Betina volunteered to write in Psychophysics toolbox extensions into our script that would display a message before each audio file played (Brainard, 1997; Pelli, 1997; Kleiner et. al, 2007). The messages she planned to write were the question “Does this person sound like they are from San Diego?”, and two response texts, “Yes” and “No,” below the question. She also planned to write in Psychophysics toolbox extensions that would be able to determine whether the user entered in a keyboard input corresponding to each of the response texts, with the “q” keyboard input indicating that the participant chose the “Yes” response text and the “p” keyboard input indicating that the participant chose the “No” response text (Brainard, 1997; Pelli, 1997; Kleiner et. al, 2007). Our other member of the group, Yuwei Zhang, volunteered to design structs that would save the keyboard inputs of the participants, along with information regarding the speakers for the audio files. She planned to save information regarding the speaker such as whether they are native, American English speakers or non-native English speakers and how many years each of the speakers has lived in San Diego. Andrew planned to help Yuwei with the structs and create an additional analysis script that would convert the “yes” keyboard input, i.e., “q” and the “no” keyboard input, i.e., “p” from the participants into integers. These values would then be used to calculate the average rating for each speaker sounding “San Diegan,” with 1 being the highest and 0 being the lowest.

Implementation

After coming up with an effective way to design our code, our group was ready to go in terms of the implementation of the design. We had created such a detailed plan that when it came time to implement

this plan in terms of collecting the audio files and construction of the MATLAB code itself, we were able to work together to implement and enhance our preexisting foundation.

For the audio file collection, Andrew and I decided to use the Voice Memos application on our respective iPhones. However, when I was recording my native, American English speakers for the audio files, I noticed that the sound from the background kept creeping into the audio files, so I made all my speakers wear headphones and record their files again. This worked, and the voices came out very clear and audible via the recording. For Andrew's audio files, one speaker could not be found, so Betina volunteered to find a non-native English speaker but still could not find one, and another speaker spoke very fast to the point that when his audio recording was read into MATLAB, it sounded sped up and high-pitched in comparison to the other audio files. We decided to then take out two speakers from the audio file collection for the non-native English speaker category. In order to balance out both categories, two speakers were taken out of the native, American English speaker category as well.

Now, I will discuss our MATLAB code. The first part of our code is a loop, that loops through four subjects, or participants. We had four participants take our experiment, so we decided to loop through each of their numbers in the MATLAB code, i.e., 1, 2, 3, 4, so that each participant's responses could be recorded into the structs in the rest of the code. A message displayed for each participant one through four, stating, "Your subject number is:" followed by their respective participant number (depending on where they fall in the order of who takes the experiment). This part of the code was written jointly, by all of us.

After this for loop, Betina wrote some code (`[w, wRect] = Screen('OpenWindow', 0, [255 255 255])`) to open up a new full screen window with a white background, using the Psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et. al, 2007). On this white window, instructions for the user were displayed with text, telling the user that they will participate in a study where they will need to pick a "yes" or "no" option for whether the speaker from the audio file sounds San Diegan or not, using corresponding keyboard keys "q" and "p," respectively. Betina then wrote a separate function all the way at the bottom of the script called "dispText" that took in the "w" and "wRect" variables that were

established with the Screen function from above (that was used to open the full, white screen with black text). This function was designed in order to display the text “Yes” and “No” underneath the main question “Does this person sound like they are from San Diego?” before each audio file played for the user to listen to. This kept going until the last audio file was played.

After Betina’s initial call to the Screen function, I wrote the code related to the audio files. I first read in each of the audio files for each of the speakers using the function “audioread” in MATLAB, that converts any .wav file into a matrix. I then saved the label names for each of the speakers using the following format: “speakerNumber_audioFileLetter_nativeEnglishORnon-NativeEnglish.” The speaker number was determined by a chart we all used on Google docs, where the first three speakers listed were non-native English speakers (listed in a random order on the chart), and the last three speakers listed were native English speakers (listed in a random order on the chart). Below is a sentence list and a chart that reflects our updated audio stimuli:

List 1

*strikethroughs reflect that the sentence was deleted

1. The San Diego Chargers have never won a Super Bowl. (non-native English speaker)
2. ~~San Diego produces the most avocados in the United States. (non-native English speaker)*~~
3. ~~Hypnosis is banned in San Diego public schools. (non-native English speaker)~~
4. San Diego has more fleas than any other city in the US. (native English speaker)
5. May 29th is Tony Hawk Day in San Diego (native English speaker)
6. ~~San Diego imports 80 to 90 percent of its water (native English speaker)~~
7. The Hotel Del Coronado is the country’s largest wooden structure. (non-native English speaker)
8. San Diego was the world Tuna capital from 1930 to 1970 (native English speaker)
9. All San Diego lakes are manmade. (non-native English speaker)
10. ~~There are over 125 breweries in San Diego (native English speaker)~~
11. The San Diego Zoo opened in 1916 (native English speaker)
12. La Jolla Cove has a history of shark attacks (non-native English speaker)

- ~~13. Water visibility in La Jolla Cove can exceed 30 feet~~
- ~~14. UCSD contains over 200,000 eucalyptus trees~~
15. San Diego hosts more than 35 million visitors each year. (native English speaker)
- ~~16. The San Diego Zoo has over 800 animal species (native English speaker)~~
17. Christmas lights still up after February 2 lead to fines (non-native English speaker)
18. The median age in San Diego is 34.9 (native English speaker)
- ~~19. Drivers spend an average of 256 hours per year in their car (native English speaker)~~
20. UCSD was founded in 1960 (non-native English speaker)

Table 1

File name	Recorded by	American?	Years in SD	Sentences Read
1. 2a_na.wav, 2b_na.wav	Andrew	No	0	1, 12
2. 4a_na.wav, 4b_na.wav	Andrew	No	3	9, 17
3. 5a_na.wav, 5b_na.wav	Yuwei	No	3	7, 20
4. 6a_ae.wav, 6b_ae.wav	Rukmini	Yes	14	5, 11
5. 7a_ae.wav, 7b_ae.wav	Rukmini	Yes	1	4, 8
6. 8a_ae.wav, 8b_ae.wav	Rukmini	Yes	2.5	15, 18

The audio file letter indicates whether this was the first audio recording of the speaker or the second audio recording of the speaker. Each speaker recorded two audio files, so the first audio recording is denoted by the letter “a” and the second audio recording is denoted by the letter “b.” The last part of the label, indicating whether the speaker is a native English speaker or a non-native English speaker, is denoted by either “na,” standing for “non-American” or “ae,” standing for “American English.” I then saved these label names into the first part of a struct for each speaker’s audio files, and the audio files themselves into the second part of the struct. For example, for audio recording “two_a_na,” I saved the label “two_a_na” to the first part of the struct, i.e., “two_a_na.a” and the audio file that was written into MATLAB with the

audioread function into the second part of the struct, i.e., “two_a_na.b.” After this, I saved all the structs into a cell array titled “audioarray.” For Yuwei’s code, the indices of the original audioarray cell array that were randomized needed to be saved, so I first wrote a code that says “orig_indices = randperm(length(audioarray))” which randomizes the indices of the audioarray based on the length of it and saves the random index numbers to an array. I then converted the cell array values into a loop-able matrix, using the cell2mat function. I then initialized a cell array called file_names that stores all of the label names (in their randomized order using the orig_indices variable that saves the randomized indices from audioarray) from the first part of each of the structs that were saved into audioarray, and then a cell array called audio_files, that stores all of the audio files (in their randomized order using the orig_indices variable that saves the randomized indices from audioarray) from the second part of each of the structs that were saved into audioarray. This ensured that the label names and the audio files corresponding to them stuck together each time, since they had the same index from orig_indices. A for loop was used to complete the file_names and audio_files cell arrays. After this, Yuwei hardcoded the values for the number of years each speaker has lived in San Diego into a struct called speakers.yearsinSD. Subsequently, Yuwei created a for loop that prepares for the keyboard input from the user, initializing a variable called keyIsDown as zero, which was used later in the code with Psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et. al, 2007). Within this for loop, I wrote in the code for the audio file to be updated each time around. I called on the orig_indices array within the audioarray array, with the variable “i” within the for loop calling on each value of the orig_indices (the randomized indices) consecutively, so that the randomized audio array could be called on consecutively within the loop. Each time, the audio file was saved under the name current_audio_file, and this current_audio_file, was played at 44000 Hz using the “sound” function. (This sampling frequency worked well with playing the sound files and was also used during one of our assignments in the COGS 119 course, which was the “Happy Birthday” assignment, where we had to play the song Happy Birthday using the sound function.) Within the loop, I paused the sound each time for the length of each current_audio_file (that was

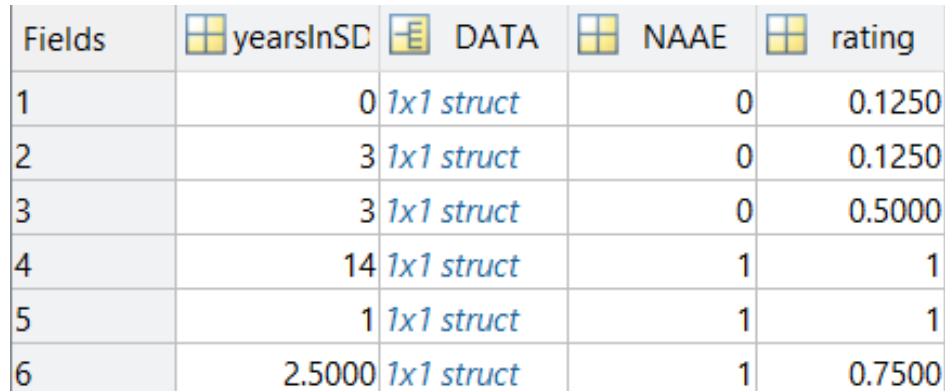
constantly updated) and divided this by 44000, the sampling frequency, so that each audio file could be played completely before the next audio file was played.

After I finished incorporating this part of the audio file code into the for loop that Yuwei created for initializing the `keyIsDown` variable, Yuwei went on to use more of the Psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner et. al, 2007). She created a while loop, that kept looping until a keypress was input by the participant. She used the `KbCheck` function for this, whose output not only updated the `keyIsDown` variable, to see if a key had been pressed or not, but also saved the keycode entered by the participant. After this, Yuwei wrote in if conditionals that check which speaker the audio file played belongs to and saves the keycodes from the participants using the function “`native2unicode`,” that converts the keycode produced with the `KbName` Psychophysics toolbox extension to an “ASCII” character, in this case, either ‘q’ or ‘p’ (Brainard, 1997; Pelli, 1997; Kleiner et. al, 2007). She then saved the keypresses to the corresponding speakers using a struct, called `speakers.DATA.response`, and just before that, saved the subject number (from the very first for loop that was created to loop through each participant in the study) to the struct `speakers.DATA.info`, to correlate the participant’s responses to the keypresses they input and the speaker associated with the audio files they heard. This “speakers” struct was saved to a .mat file titled “`speakersData.mat`.”

Using this .mat file, Andrew wrote code that loaded the data from the .mat file and first converted the “q” and “p” keypresses to integers 1 and 0, respectively; thus, 1 indicates “yes” and 0 indicates “no” from the participant of the study. This overwrote the preexisting values in the `speakers.DATA.response` struct. Additionally, Andrew wrote a for loop that loops through the speaker struct and finds the average rating for each speaker, based on all the participant’s responses, using the `sum` and `length` functions.

Results.

Using Andrew’s analysis code, the following data was saved for the four participants of our study:



Fields	yearsInSD	DATA	NAAE	rating
1	0	1x1 struct	0	0.1250
2	3	1x1 struct	0	0.1250
3	3	1x1 struct	0	0.5000
4	14	1x1 struct	1	1
5	1	1x1 struct	1	1
6	2.5000	1x1 struct	1	0.7500

Image 1. Screenshot of the .mat file saved from the four preliminary participants taking the experiment.

From this table, the fields denote the speaker number, and the years each speaker has lived in San Diego. The NAAE column lists “0” for those who are non-native English speakers, and “1” for those who are native, American English speakers. The fifth column indicates the mean ratings for how “San Diegan” sounding each speaker was based on each of the participant’s keypresses. We found that the first part of our expectations was true for this preliminary group of trials! The native, American English speakers had a mean rating of 1 or closest to 1 for sounding “San Diegan” as opposed to the non-native English speakers, who had mean ratings of 0.125, 0.125, and 0.500, respectively.

Discussion.

Through this study, we found that with the four participants who participated in our study, the hypothesis/expectation we formulated prior to executing our study, i.e., that native, American English speakers can emulate an “accurate, San Diegan” accent in the eyes (rather, ears) as opposed to non-native English speakers proved to be true based off of the mean ratings from our analysis code. However, we had also formulated another hypothesis before embarking on this study. The other hypothesis of ours that was not tested through this study was that the longer a listener has resided in San Diego, the better their idea is of a “true, San Diegan” accent. This was not examined through this study, and should have been, given that the foundations of our study were built upon this idea of internal speech representation and characterization. However, with more time in the future, this could be tested out as an extension of this study.

Overall, I learnt that how we perceive accents at regional, national, and global levels can affect our internal representation of speech and linguistic associations.

References

- Brainard, D. H. (1997) The Psychophysics Toolbox, *Spatial Vision* 10:433-436.
- Johnson, Keith (1997) "The auditory/perceptual basis for speech segmentation," *OSU Working Papers in Linguistics* 50:101-113
- Kleiner M, Brainard D, Pelli D (2007) "What's new in Psychtoolbox-3?" *Perception* 36 ECVF Abstract Supplement.
- Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, *Spatial Vision* 10:437-442.
- Population*. San Diego: The City of San Diego. Retrieved from <https://www.sandiego.gov/economic-development/sandiego/population>