6-1-2013

Development of a New Experiment Demonstrating Categorical Perception

Kelsey Montzka
Boise State University

Development of a New Experiment Demonstrating Categorical Perception

Kelsey Montzka
English Department, College of Arts and Sciences
Faculty Advisor: Dr. Michal Temkin Martinez

Abstract

In this project, we present a new online pedagogical demonstration to illustrate the psycholinguistic concept of categorical perception. Human speech sounds are characterized by a vast amount of variation and diversity, resulting in a lack of one-to-one correspondence between the acoustic signal and discrete sound categories. The human brain resolves this “lack of invariance" problem and still maps the signal that reaches the brain onto distinctive speech sounds through categorical perception (CP; Liberman, et al., 1967). To teach the concept of CP, some linguistics instructors have been using an online demonstration (Eriksson, 1997) wherein students partake in an online experiment-like procedure containing two tasks. The 1997 demonstration utilizes a labeling task and a discrimination task to illustrate CP of place of articulation in voiced stops (the difference between [b], [d], and [g]), as perceived in gradual shifts in F2 formant transitions from stops to following vowels using synthesized speech that is quite non-human-sounding. This becomes problematic in light of Schouten, Gerrits and van Hessen (2003) postulation that human-like speech is a component of CP of speech sounds (Minagawa-Kawai, Mori, & Sato, 2005; van Hessen & Schouten, 1999). The 1997 demonstration also utilizes the AX stimulus presentation method, which Gerrits and Schouten (2004) also demonstrate as yielding inconsistent CP results. The potential of students receiving inconsistent results counteracts the purpose of the demonstration which is to aid student learning.

In this project, we have created an online portal containing a multi-faceted demonstration of CP. Unlike the 1997 demonstration, we utilized the 2I2AFC (two-interval, two alternative forced choice) stimulus presentation method which should elicit more categorical results from students, helping to better demonstrate the phenomenon (Gerrits & Schouten, 2004), and provided different acoustic cues illustrating CP. In addition to illustrations of CP for F2 formant transition, the new online portal is used to illustrate voice onset timing (VOT; the cue that distinguishes between the sounds [b] and [p]), as well as how contextual factors and participants’ linguistic backgrounds affect CP. The contextual factors used are that of phonological context where categorical boundaries are shifted in order to accommodate phonological variation as described in Sumner (2011), and that of lexical context wherein categorical boundaries are shifted to give preference to perception of a word compared to a non-word, as described in Miller (2001). In demonstrating the effect of linguistic background on CP, the portal includes streams that demonstrate variation in native English speakers’ and native Japanese speakers’ CP of the sound, containing a set of previously collected results for each participant group, allowing students to compare their own results to others’ (McCandliss, et al., 2002; Minagawa-Kawai, Mori, & Sato,

2005; McKain, Best, & Strange, 1981). Finally, a stream within the portal demonstrates to what extent English-speaking participants' can categorically perceive the contrast between two non-English sounds. The goal of designing this experiment is to equip instructors with a detailed and effective tool to supplement their instruction and facilitate a broader understanding of the complexities of speech perception.

An additional feature of this portal is the automatic tabulation of student results from each stream, eliminating the need for instructors to require their students to send individual results to them and then manually tabulate the results (currently necessary using 1997 demonstration). These improved structural features and expanded scope demonstrate how this portal serves to more comprehensively demonstrate the concept of CP.

Keywords: Language, Speech Perception, Pedagogy, Psycholinguistics, Language and Mind, Demonstration, Categorical Perception, linguistics, VOT.

Introduction

There is a high amount of diversity in the sounds that are heard and processed by the human brain. This is called the *lack of invariance problem*-- a lack of one-to-one correlation between the acoustic signal and the way sounds are conceptualized by speakers. The abundance of variation can come in several forms, including acoustic variation and allophonic variation. Acoustic variation arises from variations of the source of the sound--in the case of speech sounds, the vocal folds-- and of the resonance of the chamber through which that sound passes--in the case of speech sounds, the resonance of the vocal tract. This difference accounts for differences between two adults whose vocal folds are different thickness (i.e. children vs. adults, males vs. females) and whose vocal tracts are different lengths (i.e. tall vs. short adults).

In addition to acoustic variation, there is allophonic variation, which can be described as the way the pronunciation of sounds is affected by its position in a given word. There are two types of allophonic variation– one-to-many and many-to-one. In one-to-many variation, one target sound is pronounced in different ways when spoken, as is the case with the target form "t" and its realized forms of [t]-- as in "trumpet"-- and [d]-- as in "butter." In many-to-one variation, multiple target forms end up being realized as the same sound, as is the case with the target forms [d] and [t] in "madder" and "matter," respectively, which are both realized as the same "d"-like form. Liberman (1967) describes further that acoustically, even the same sound perceived as [d] will have widely varying acoustic properties depending on the vowel following it.

Despite the *lack of invariance* problem, the human brain still derives meaning from the acoustic signal and this process is known as categorical perception (CP). The brain has a series of language-specific categories for different acoustic cues that will accept acoustic variation that falls within each one and designate the sound as being a representation of that category's target form. If a speech sound falls between the categorical boundaries, it is considered to be a representative member of that category, but once a sound-- within that cue continuum's variation-- crosses a categorical boundary, it will be perceived as a different sound.

Psycholinguists have explored the phenomenon of CP, analyzing the brain's ability to discriminate different acoustic features and form categories that allow for variation. Research over the years has supported this interpretation of speech perception and has examined the CP of various levels of speech, ranging from the phonological (speech sound) level to the discourse (conversational) level, and how a listener's linguistic background affects their speech perception (Ladd & Morton, 1997; McKain, Best, & Strange, 1981; Sumner, 2011; Wilson & Iacoboni, 2006; Grichkovtsova, Morel, & Lacheret, 2011; Minagawa-Kawai, Mori, & Sato, 2005; Sumner, 2011).

This process has been discovered to be more encompassing than originally perceived as the brain simultaneously uses two different modes of hearing when categorizing sounds. One of these modes is acoustic, which focuses on the acoustic signal and the other phonetic, which analyses the sounds utilizing a conceptual framework of how similar the acoustic signal received by the brain is to a predetermined phonetic form (Pisoni & Lazarus, 1974; Pisoni & Tash, 1973) Gerrits and Schouten (2004) found that the more the phonetic mode was utilized, automatically, when discriminating between two sounds, the more categorical the results were. Furthermore, CP is highly contextual, with phonetic mode and phonetic boundaries being somewhat flexible depending on word-level and dialectal context (Miller, 2001; Sumner, 2011; Repp & Liberman, 1987)

1.1 Eriksson's 1997 demonstration

Anders Eriksson's (1997) online pedagogical demonstration of categorical perception is used to teach students how the brain processes speech sounds using CP. The demonstration includes two tasks -- a labeling task and a discrimination task. During the categorization task, participants are subjected to a series of randomized synthesized speech stimuli that come from various points along the continuum for second formant (F2) transitions from consonant to the following vowel. The formants of the three sounds [ba], [da], and [ga] can be seen in the spectrogram (a visual representations of the acoustic properties of sound) in Figure 1.

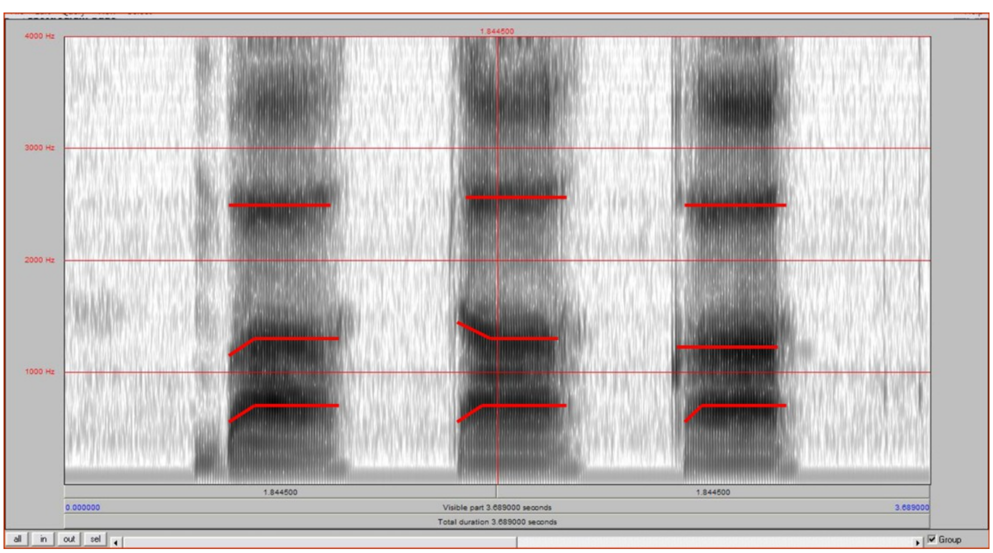

Fig. 1
Image Credit Dr. Temkin Martinez

In Figure 1, the far left column indicates the spectrogram for the sound [ba], the middle column indicates the sound [ga], and the far right column indicates the sound [da]. The dark bars along the spectrogram indicate the formants of the vowel following [b], [d], and [g], and formants are highlighted by the bold red line. The middle formants for each column are all the second formant (F2) of each sound. The difference in shapes of the F2 bar indicate the F2 transition from the different consonant sounds into the vowel [a].

Figure 2 indicates how the F2 transition varies on the continuum of sound from [ba] to [ga]. Sounds with the F2 formant transition that is illustrated by beginning at a higher frequency and decreasing to the F2 bar are the most [ba]-like. Sounds with a constant F2 formant transition frequency will be the most [da]-like. Finally, sounds with a F2 formant transition that begins at a lower frequency and increases will be the most [ga]-like. The stimulus tokens in Eriksson's demonstration are drawn from various points along the continuum with the most [ba]-like sounding stimulus on one end of the continuum, and the most [ga]-like sounding stimulus on the other end of the continuum.

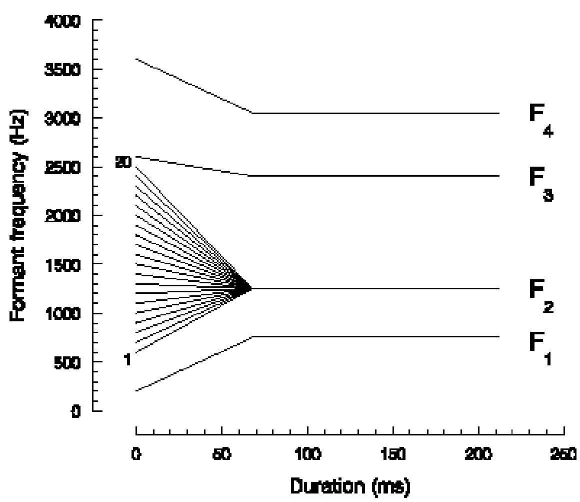

Fig. 2
Image credit: Dr. Anders Eriksson

The variations within the F2 transition are the primary acoustic cues that distinguish between the sounds [ba], [da], and [ga], so participants listen to synthesized speech stimuli that come from some point along the continuum between these three sounds. Their first task is to decide which category each sound belongs to and click on the corresponding button on the screen, designating the sound as being one of those three sounds-- [ba], [da], or [ga]--. Immediately after the selection is made, the next stimulus is presented.

The discrimination task tests whether the participants are thinking of two acoustically different sounds as belonging to one or two different sound categories. The participants are always presented with two speech stimuli that are one step apart along the F2 transition continuum. Participants determine if the speech stimuli are the same or different. When two sounds drawn from the same category are presented, they should be designated as being the same sound. Likewise, when two sounds straddling a categorical boundary are presented, they should be designated by participants as being different sounds. The stimuli are presented in AX presentation method, which is one of several stimulus presentation methods for sensory stimuli. The AX presentation method is when two stimuli, which may be the same or different, are played in close succession. "A" indicates the first stimulus, and "X" indicates the second stimulus, which may be the same as "A" or different. The participants are then asked to indicate whether the stimuli are the same or different. In the case of Eriksson's demonstration, the pairs all differ from one another acoustically by one step along the F2 transition continuum. The pairs are presented in a randomized order along the continuum with the hypothesis that only two sounds that straddle a categorical boundary will be perceived as different. There are some shortcomings in the experiment that lead to results and do not serve the interest of clarifying the concept to students.

The shortcomings of this particular demonstration primarily revolve around two factors. The first is that the experiment uses computerized speech synthesized using old technology since it was created nearly a decade and a half ago. The synthesized speech sounds quite automated and non-human. Schouten, Gerrits and Van Hessen (2003) postulate that human-like speech is a component of achieving CP of speech sounds. In fact, there is a disclaimer on the introductory page to the experiment explaining that, "acoustically, [the] structures [of the sounds] are somewhat stylized. They may, therefore, seem a bit unnatural, but you should disregard that and just determine which of the syllables [are which]" (Eriksson, 1997). This aspect of the experiment ought to be revisited as it has the opportunity to significantly affect results and the efficacy of the demonstration

The second shortcoming of the experiment is that the presentation method of the stimuli for the discrimination task is uniformly AX, which Gerrits and Schouten (2004) have found to significantly bias results as it leads participants to inconsistently favor phonemic labeling. Phonemic labeling is a situation in which the listener compares and categorizes sounds heard based on idealized sound forms, utilizing generalized concepts of sound tokens. In the context of the experiment, "[a] subject's response could be completely dominated by a subjective phoneme-based criterion, very close to one end of a scale between same and different" (p. 364). There is a variety of methods for stimulus presentation which elicit either phonetic or acoustic labeling. Although this tendency for different stimulus presentation methods leads Gerrits, Schouten, and van Hessen (2003) to dismiss any discrimination task as helpful in demonstrating or testing CP, this is not a universal opinion. Furthermore in the interest of promoting the demonstration of the phenomenon, it might be advantageous to utilize a method that elicits more categorical results. Figures 3 through 6 indicate the results of the shortcomings of Eriksson's demonstration.

Figure 3 demonstrates the ideal results of a labeling task. In Figure 3, results should indicate that participants should consistently respond "ba" until the F2 transition varies across a categorical boundary, at which point the participants will consistently respond "da" and then, similarly, "ga."

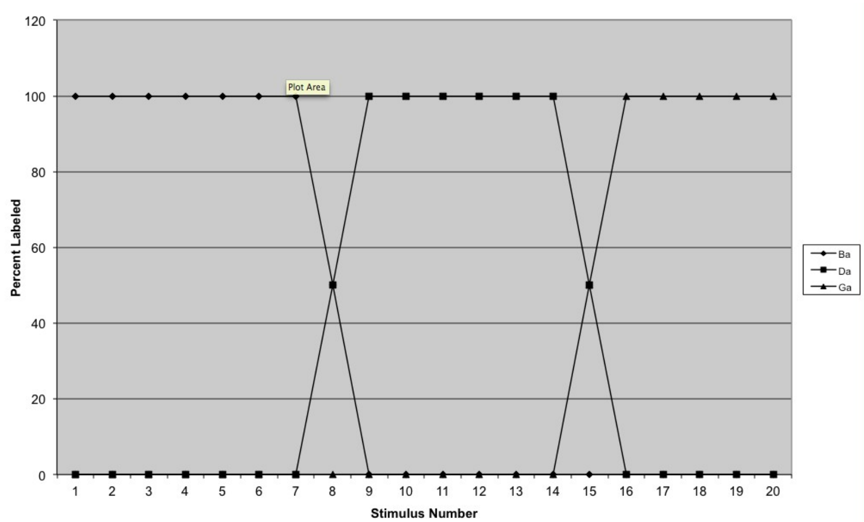

Figure 3
Image credit: Dr. Temkin Martinez

Figure 4 illustrates actual results from the labeling task in Eriksson's demonstration.

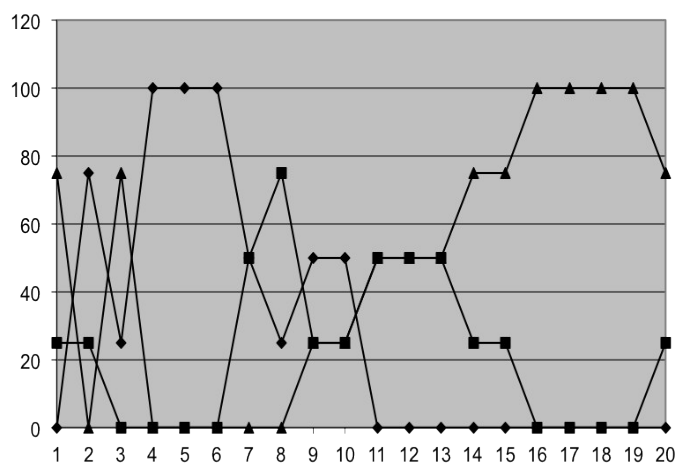

Figure 4
Image Credit: Dr. Temkin Martinez

Figure 5 indicates ideal results for Eriksson's discrimination task. For all stimulus pairs that do not straddle categorical boundaries, participants would indicate them as the same sound. Stimulus pairs that straddle categorical boundaries would be marked as different sounds.

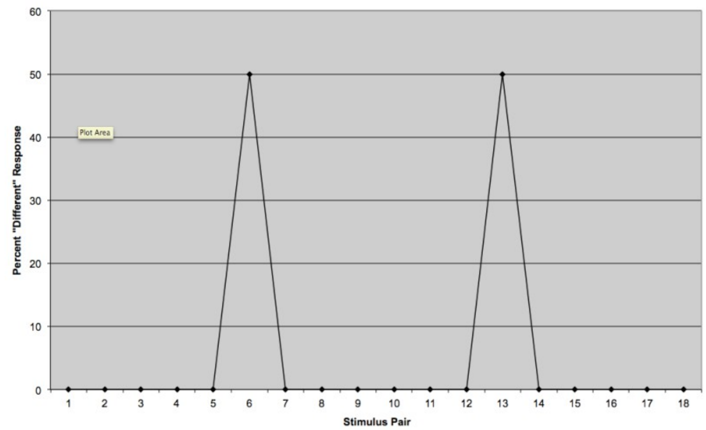

Figure 5
Image Credit: Dr. Temkin Martinez

Figure 6 indicates some actual results from the discrimination task in Eriksson's demonstration.

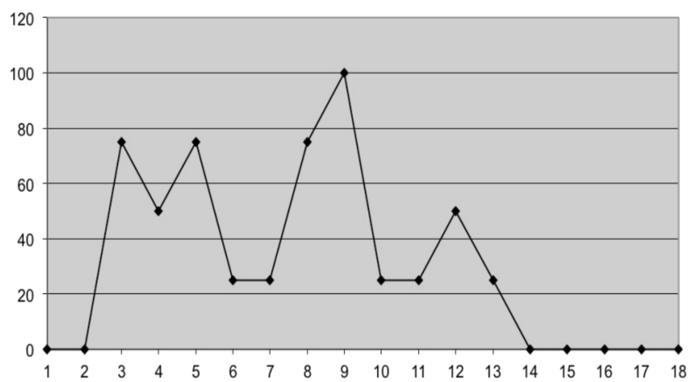

Figure 6
Image Credit: Dr. Temkin Martinez

There are no other readily available online demonstrations besides the one created by Eriksson in 1997. The aims of this project are to create a model that will not only utilize more refined and speech-like stimuli and feature the most effective stimulus presentation method, but also be customizable. As CP is a phenomenon that is utilized for all variety of speech cues, it is possible to create demonstrations that present how CP functions in native versus non-native speakers of a language (Wilson & Iacoboni, 2006; McKain, Best & Strange, 1981), the level of contextual dependence for different sounds (Miller, 2001), how accented speech affects CP (Sumner, 2011), and give basic demonstrations of categorically perceived sounds. This portal will enable teachers to more effectively illustrate the complexities of this language phenomenon.

Methodology

2.1 Background

The initial steps had been focused on researching various shortcomings and information about previous experiments in CP in order to more effectively demonstrate its complexity in a clear and understandable way. This required extensive background research on the phenomenon itself. The literature review not only gave rise to improvements to be made on the current design, but also provided ideas to expand the scope of the portal to demonstrate CP and its effects on multiple speech cues. Finally, it also involved researching what stimuli would have to be synthesized as a part of the work and what stimuli was created recently enough to be eligible for use.

As discussed in the Introduction, it is possible to create demonstrations that compare CP of native versus non-native phonemes (Wilson & Iacoboni, 2006; McKain, Best & Strange, 1981), the level of contextual influence on CP for different sounds (Miller, 2001), how accented speech affects CP (Sumner, 2011), and also simply illustrate CP of sounds. This resulted in the formation of six separate streams. Each stream consists of a two-part "experiment." Each demonstration illustrates how CP works for different speech cues and different contexts.

All of the streams feature a short series of initial steps where students register data about their listening environment and demographic information. The selected options correspond to a series of data labels, and each set of results is "tagged" with the information entered. There is also an option to enter a unique course label, which also generates and attach a new label onto the participants' data. On the results screen for each stream, participants can add constraints based on these labels to limit which comparative data to view. Instructors can enter their unique class tag and view only the tabulated results from their class. Aside from practical applications, this feature also has pedagogical implications whereas, students can view and compare between results from native English speakers and results from native German speakers, etc.

Every stream consists of two parts: a labeling task and a discrimination task. In the labeling task, the participant is asked how, and where, categorical boundaries exist along the continuum for an acoustic cue. This is accomplished by exposing participants to an auditory stimulus drawn from somewhere along the continuum in question and having the participants label the sound as a representative of one of the choices offered. The purpose of the labeling task is to illustrate the aspect of CP that categorical boundaries exist along a continuum of variation. Like in Eriksson's demonstration, the participants' responses will be logged and mapped onto a graph with x-axis being each stimulus number, and the y-axis being the percentage of times the participant label selected a particular labeling response. This will allow participants to view which stimuli along the continuum they labeled as what sound, enabling them to view at what point along the continuum they began to perceive a stimulus as a different sound.

The discrimination task further demonstrates how two sounds which straddle a categorical boundary will be classified as different sounds; whereas two sounds that are located within one category will be perceived as two instances of one sound. Using the Two-Interval Two-Alternative Forced Choice Task (2I2AFC), participants are exposed to a two-part sequence of auditory stimuli that are ordered randomly. The participants are asked to designate the sounds as being ordered in one way or another (i.e. AB or BA?). The principle behind the 2I2AFC task is that participants will be able to more accurately distinguish the order of two sounds that straddle categorical boundaries; whereas, their accuracy for determining the order for sounds drawn from within one category will be lower. Participants' responses will be automatically logged and plotted onto a graph with stimulus pair intervals as x-axis and percentage correct ordering as y-axis, allowing participants to observe how discrimination is easier and more accurate for cross-boundary stimulus pairs.

2.2 Different streams

Although each stream has essentially the same experiment design, they demonstrate different aspects of CP. One stream (S1) is, essentially, an improved version of Eriksson's experiment, demonstrating CP of F2 formant transition. The F2 formant transition is an acoustic cue that distinguishes between the sounds [ba], [da], and [ga]. Sound stimuli from along this continuum, with most [ba]-like sounds on one end and most [ga]-like sounds on the other, is used to demonstrate not only the points along the continuum where the participants distinguish between [ba], [da], and [ga] but also how participants will be able to discriminate better between stimuli from two different categories.

A second stream (S2) illustrates CP of Voice Onset Timing (VOT). VOT is the time elapsed between the release of the stop and the commencing of vibration of the vocal folds and is a feature that distinguishes between voiceless sounds, like [pa], and their voice counterparts, like [ba]. For S2 there is a continuum with most [pa]-like sounds on one end and most [ba]-like sounds on the other end, and gradual variation along it.

A third stream (S3) tests participants' ability to perceive the contrast between two contrasting sounds not featured in English. S3 utilized contrasting sounds in Russian-- the voiced and voiceless retroflex sibilant [ʂ] and [ʐ]. The decision to use these two sounds was based on several criteria. First, the sounds had to be sounds that varied in only one feature, so that they could be placed on a simple continuum. Second, the sounds had to contrast in at least one other language. Third, the sounds had to not be featured in English. The purpose of S3 is to demonstrate how participants will have a less consistent CP of sounds unfamiliar to them.

Similarly, another demonstration (S4) illustrates CP of the sounds [l] and [r]. The sounds on one end of the continuum were the most [l]-like and the sounds on the other end of the continuum were the most [r]-like. These two sounds are sounds that contrast for native English speakers, but do not contrast for native Japanese speakers. Participants will be able to compare their results to those of native Japanese speakers.

Finally, two streams demonstrate the effects contextual information has on CP. One (S5) demonstrates lexical-level effects on CP, with the non-word [peef] being on one end of the continuum and the word [beef] being on the other end. This contrast is essentially the same as the [pa]/ [ba] contrast in VOT, so participants can examine their results of where along the VOT continuum the categorical boundary is and their relative accuracy in discrimination and compare the results between this stream and the results from S2.

The last stream (S6) would demonstrate sound-level (phonological) effects on CP. Like (S2) and (S5), the context of this stream would be analyzing VOT. The continuum would feature a non-accented token of a word on one end and a heavily accented token of a word on the other end.

Participants will be able to compare their results from this stream to their CP results of S2 and S5.

The author was able to collect a list of already published experiments which utilized the appropriate stimuli for each stream, and contacted the researchers requesting access to their stimulus sets and consent to use the stimuli in this project. One set of stimuli was received. This required the synthesis of five sets of stimuli. These stimuli were synthesized using MATLAB.

The author was able to contract with a coder in order to code the information of the experiment. The coder was required to create the framework into which the information for the experiment was put. The author provided text for instructions, a template of the order of each stream, and a description of the layout of the final project to the coder. The author also provided instructions for how the data collected in the course of the experiments was to be processed and presented. Finally, the author contributed the synthesized sound files for each stimulus used in each stream.

Conclusion

The creation of a new and updated version of this demonstration will enable more instructors to have access to better materials and resources to teach courses on psycholinguistics. Additionally, the platform of the final demonstration is flexible and open-ended, allowing for the possibility of the addition of new streams. Eriksson's demonstration is highly biased towards native English speakers, and although the final portal constructed this spring was created by a native English speaker and, thus, will feature similar biases, the framework in place will allow others to contribute to not only this project, but also the larger body of knowledge on psycholinguistics.

References

Eriksson, Anders. (1997). Categorical Perception. Retrieved from: http://www.ling.gu.se/~anders/KatPer/Applet/index.eng.html

Gerrits, E., & Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Attention, Perception, & Psychophysics*, *66*(3), 363-376.

Grichkovtsova, I., Morel, M., & Lacheret, A. (2011). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*.

Harnad, S. R. (Ed.). (1990). *Categorical perception: The groundwork of cognition*. Cambridge University Press.

Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics*, *38*(2), 185-196.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical?. *Journal of Phonetics*, *25*(3), 313-342.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological review*, *74*(6), 431.

McKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English IT I and/I/by Japanese bilinguals. *Applied Psycholinguistics*, *2*(4), 369-390.

Minagawa-Kawai, Y., Mori, K., & Sato, Y. (2005). Different brain strategies underlie the Categorical perception of foreign and native phonemes. *Journal of cognitive neuroscience*, *17*(9), 1376-1385.

Miller, J. L. (2001). Mapping from acoustic signal to phonetic category: Internal category structure, context effects and speeded categorisation. Language and Cognitive Processes, 16(5-6), 683-690.

Pisoni, D. B., & Lazarus, J. H. (1974). Categorical and noncategorical modes of speech perception along the voicing continuum. *The Journal of the Acoustical Society of America*, *55*(2), 328-333.

Pisoni, D. B., & Tash, J. (1973). Reaction Times to Comparisons Within and Across Phonetic Categories:Evidence for Auditory and Phonetic Levels of Processing. *Cog. Psychol. [Also in Haskins Laboratories Status Report on Speech Research SR-34, 77-88*.

Repp, Bruno H., and Alvin M. Liberman. "Phonetic category boundaries are flexible." *Categorical perception: The groundwork of cognition* (1987): 89-112.

Schouten, B., Gerrits, E., & van Hessen, A. (2003). The end of categorical perception as we know it. *Speech Communication*, *41*(1), 71-80.

Sumner, M. (2011). The role of variation in the perception of accented speech. *Cognition*, *119*(1), 131-136.

van Hessen, A. J. (1997). "Categorical perception as a function of stimulus quality." The Journal of the Acoustical Society of America (0001-4966), 101 (5), p. 3199.

Wilson, S. M., & Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage*, *33*(1), 316-325.

Acknowledgements

The author would like to thank the Student Research Initiative program, and specifically Erin Muggli and Dr. Liljana Babinkostova, for enabling this project to take shape, and for giving undergraduates the opportunity to participate in academic research in a guided, mentored setting. Additionally, the author is extremely grateful to the tireless efforts of Dr. Michal Temkin Martinez whose work in mentorship, inspiration, and guidance played an integral role in not only this project but also in the educational experiences of so many. Thanks to Dr. Julie Fiez for granting access to an invaluable stimulus package. Finally, thanks to Dr. Ed Holsinger for providing his coding expertise to the project.