UBC ISCI 422 “Models in Science”

Project 3: Model Construction – Report

Making Sense of the Unknown:

A Model for Increased Cognitive Load for Recognizing Ambiguous Words

by

Vivian Pan

Abstract

People can distinguish and categorize a wide range of phonological variations in speech into distinctive categories of words. However, as the speech sound becomes more degraded (with noise or with differential attention and speech production), it becomes harder to hear and takes longer for us to recognize (Andruski, Blumstein, & Burton, 1994, Hoff, 2001, Aydelotte & Bates, 2004, & Blumstein, 2004). This paper will propose a mathematical model of processes that contribute to the increased processing time for ambiguous sounds. Using Microsoft Excel and the construction of an artificial language, the model will use a probabilistic approach to describe sounds. This study will explore the frequency effect and the contextual (lexical) effects of sounds on word recognition. The number of steps taken to achieve word recognition is compared to quality of sound input. Results and predictions made by the model are found to be comparable to observations of reaction times over acoustic continua constructed with varying voice onset time. Suggestions of future improvements involving the Bayesian Probability Theory and Neural Networks are also suggested.

1 Introduction

One of the unique features and mysteries of the human mind is its ability to obtain and use language to communicate. People speak in different tones, with different accents, and pronounce words differently, yet we learn to generalize and ignore these differences in order to distinguish the wide range of speech sounds produced. On the other hand, we can easily distinguish between two similar sounding words such as "coat" and "goat", which only differ from each other in milliseconds of the time of air release in speech production. It seems that people have the ability to be extremely tolerant to variations and degradations in speech signal while at the same time are extremely sensitive to the differences in these sounds. The phenomenon where we perceive sounds as distinctive categories is also known as categorical perception (Hoff, 2001). How do we differentiate such a wide range of speech sounds into limited categories of sounds? Do we process all sounds the same way? Would ambiguous sounds take more cognitive power to process?

The question of language perception is a complex one but to oversimplify it, the problem can essentially be broken down into three parts: the input, the black box mechanism, and the output. As expected, researches tackling the problem can be categorized into through these three approaches: the characteristics of sounds we hear in language (the input), the mechanism that our brain employs (the black box), or our perception (the output).

Extensive research has been done in the past to identify the different phonetic features of language such as voice onset time (VOT), glottal excitation, burst amplitude, and vowel length (e.g. Pisoni & Luce, 1987, Blumstein, 2004). Blumstein (2004) suggests that phonetic categories are characterized by a set of acoustic properties that are time varying, relative, and graded. More importantly, acoustic properties have perceptual consequences – graded acoustic stimuli demonstrate the classic categorical perception pattern, but vary in reaction time within each category depending on the quality of the sounds (Blumstein, 2004). Further more, the structure and perception of speech seems to affect higher language processing such as lexical selection and integration. This suggests that differential input does affect output in particular ways. We known that graded acoustic input results in categorical outputs and that the more ambiguous the input is, the longer it takes the "black box" mechanism to process the input to produce an output. So how does our brain do this?

One proposal is the possibility of activating multiple semantic neural networks in speech recognition. Using semantic priming techniques, Blumstein (2004) concluded that acoustic variability in phonetic category structure affects both the perception of phonetic categories of speech as well as higher language processing such as lexical semantic networks. Andruski, Blumstein, & Burton (1994) also suggest that low-level acoustic differences influences speech processing even when listeners judge the phonetic identities to be the same.

In addition, phonetic manipulation seems to have an affect on only words and not nonwords, indicating that acoustic information is important in lexical processing and that it may be used to anticipate the patterns in the mental lexicon that will match incoming speech signal. Ganong (1980) also found a lexical effect where reaction times for phoneme recognition in words differ from nonwords. Another robust feature of language that influences speech production and perception is the frequency of the words. Word frequency effect is the tendency for common words to be perceived correctly at much lower speech to-noise ratios than uncommon words (Savin, 1963).

In summary, researchers have identified the different acoustic features (the input) that may influence word recognition. The perceptual phenomenon as a result of the varying acoustic features (categorical perception) has also been studied extensively. However, there is no consensus between the existing models for the mechanisms of language perception. As language and human perception are complicated, I will focus on a small aspect of language perception and attempt to model the phenomenon of subphonetic sensitivity in language (it takes us longer to process ambiguous sounds even though we can't consciously tell the difference between the two sounds).

2 Methods

Understanding the mechanism of how we recognize and perceive speech sounds will be an important part of understanding both learning and language but we have yet to achieve that goal. Despite the fact that some "language centers" have been localized in the brain, we fail to find a pathway for speech recognition. One obstacle in developing a testable mechanism is the fact that we don't fully understand the functions of all the brain structures and how it is integrated. Most of what we've learned about language has come from patients with language deficiency, clearly not a good representation of the overall population. Compound this obstacle with ethical considerations to do experiments on the human brain; a computer model might be useful and more suitable to help us understand the mechanisms of how speech may be recognized.

Mathematical modeling is used here to describe and integrate the simple sound matching and lexical facilitation mechanisms. A mathematical model provides many advantages to study the complex phenomenon of language perception. Mathematical models offer a great deal of manipulating power using little to no resources. Microsoft Excel is used in the current model because it is able to change the parameters, conduct the experiments, and compile data all on one Excel data sheet. Excel worksheets provide a way to produce large amounts of data while being very transparent about the process in which it is done. The mathematical expressions used here are also intuitive and solvable, which help us understand the mechanisms and allow a more realistic model of greater complexity to be investigated by non-mathematicians. In addition, simulations can be run over and over again with minimal cost since the calculations can be done in a short amount of time. The system can also be manipulated on a wide range of scales, allowing the possibility to expand the model to adopt a more complicated and extensive language. The model can also be extended to include more processes as we discover more underlying mechanisms.

2.1 Model Description

Language researchers agree that language perception can be characterized into three distinctive processes: first, feature extraction or information gathering from acoustic or visual input of language; second, the selection or matching of this information to our mental lexicon; and third, the integration of these information in a semantic context. (Aydelott & Bates, 2004). Traditional language models propose a simple matching mechanism for word recognition where the sound input from the environment is matched to a memory of the sound representing words in our mind; word recognition is achieved when the sounds match (Connine et al., 1997). However, more recent studies have found that subphonemic variations and other cues such as contextual information (e.g.Ganong, 1980, McClelland & Elman, 1986) or word frequencies (e.g. Savin, 1963) affect our perception of word recognition. This suggests that information in addition to just the acoustic stimuli influences our perception in some way. On the basis of this assumption, I propose a model where additional information helps us identify words with ambiguous sounds at the cost of increased cognitive load (Figure 1).

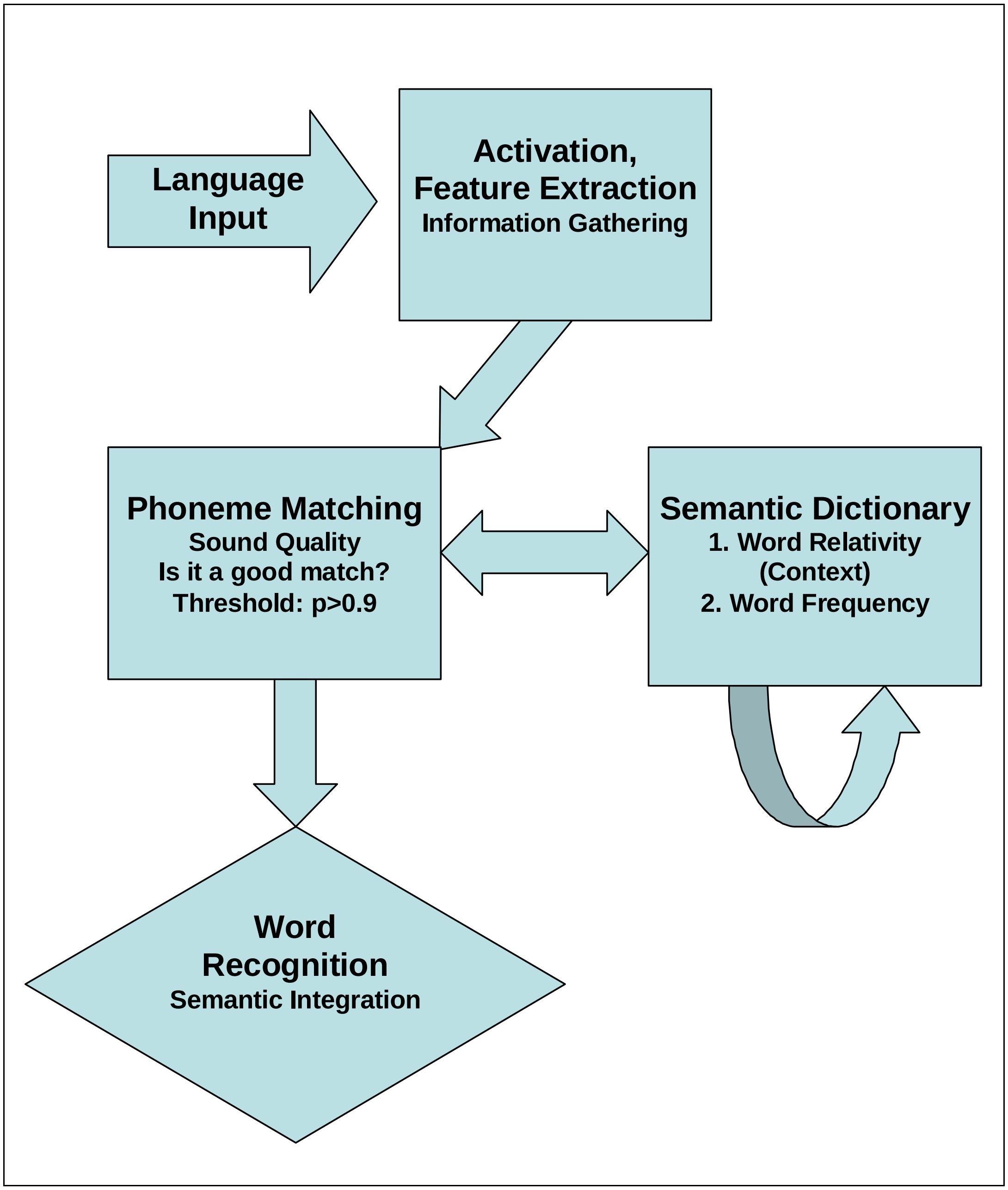

Figure 2.1 Conceptual representation of the model. The Language Input contains a wide range of acoustic information (described as a matrix, see example in Table 2.1). The ears serve as an information gathering "funnel" which relays this information to different parts of the brain where features relevant for processing are extracted. The extracted features are then used to match an existing prototype in our memory. When the input matches the prototype, the sound is recognized as a word. When the input does not match the prototype, additional information as provided in an internal semantic dictionary is used to help the recognition process.

Table 2.1 Matrix describing each sound input. Each sound is described as a potential word consisting of three "syllables" or phonemes. Each phoneme is described as the probability of being sound "A", sound "B", and sound "C".

	First Phoneme	**Second Phoneme**	**Third Phoneme**
Probability of A	0.4	0.8	0.7
Probability of B	0.5	0.1	0.2
Probability of C	0.1	0.1	0.1

First, an artificial language is constructed (see Appendix A) because the English language (or any language) is so immensely complex; it would be extremely difficult to understand the way we perceive speech no matter how simple the actual mechanism is. The artificial language in the current model consists of only three phonemes (sounds) represented by "A", "B", and "C". Ten of the possible 27 combinations are words. Phonemes "A" and "B" for the first ”syllable” mimic minimal pairs (pairs of words that differ in only one phoneme and have distinct meanings) in the English language. For example, in the English language, /b/ and /p/ are minimal pairs that sound very similar and differ minimally, yet at the same time the perception can only be one or the other but never both. Each input of sound is described as a matrix of probabilities of each phoneme representing the actual phoneme (Table 2.1). Second, this model will focus on three variables that may affect the process of word recognition: sound quality, context, and word frequency. Each variable will be described in probabilistic expressions and the threshold for word recognition is set as $p>0.3$.

Sound quality is defined here as the probability of a potential word based on the input alone. The overall sound quality of a word is built in the way the sounds are described in the artificial language (Table 2.1) because the matrix describes each phoneme as the probability of sounds. The sound quality is therefore defined as:

p(ABC) = p(first phoneme as A) * p(second phoneme as B) * p(third phoneme as C).

Equation 2.1 Probability of a potential word based on the input alone.

Word frequency is defined as how often a word is used in a language and this information has an effect on our perception (see Appendix B). Research has shown that words with higher frequencies seem to be more recognized, this is called the frequency effect. The frequency effect helps the recognition of the word by:

p(ABC) = p(sound as ABC) * p(word frequency)

Equation 2.2 Probability of a potential word based on word frequency

Context is defined as the influence of the meaning of other words on the recognition of the sound of interest. Priming is often used in language research to study the effect of context on perception, which refers to the idea that when a related word is presented before a sound, people use that information to help them recognized the word. The idea is that when someone hears a word preceding another word (a prime), they use that information to help them identify the word following it. A table of prime relativity (see Appendix C) that describes the "relatedness" of two words (the probability of a word occurring when preceded by a prime) is constructed. The effect of context is calculated by:

p(context) = p(sound)*p(prime relativity given prime).

**Equation 2.3 Probability of a potential word based on the context.
Probabilities are normalized at the end of this stage.**

The probabilities for each possible 27 combinations of the sounds are calculated at each stage. The sound with the highest probability is chosen and checked against the threshold value. If the probability is higher than the threshold, the model would report the sound as either the word or as a nonword. If the probability is lower than the threshold, the word is not recognized and the model would respond “Don’t Know” and it would go on to use the next step(s) where more information is processed to increase the probability of word recognition. There are a total of 3 steps: the initial step of sound quality evaluation, and the additional steps of frequency and context. The number of steps it takes for the model to recognize a word is compared to the initial sound quality. The number of steps is expected to increase as the sound quality decreases.

This model assumes that the processes occur in serial rather than parallel. Since the purpose of the current model is to emulate whether ambiguous stimuli increases processing time and not the amount of time, the order in which the processes occur should not matter. When the processes are put in different orders, it still displays the same pattern in result. In this sense, the model may or may not represent the exact interactions between all the factors, however we do not know enough about our brain processes to test or even propose a more detailed and comprehensive interaction. The current model is sufficient to show that additional processes may be needed to manage information that we get from the environment. The current model only includes three processes and they are given equal weights. In actual word recognition, more processes may be involved in different ways and combinations however simplifications to the model is needed and it would be impossible to account for all the factors involved in word recognition. Sound quality, context, and word frequencies are chosen as criteria in the current model because previous researches have shown significant effects of these factors on our perception (e.g. Ganong, 1980, Savin, 1963, Aydelott & Bates, 2004 & Blumstein, 2004).

2.2 Model Verification

Depending on the dataset, the slightest changes can affect outcome and predictions of a mathematical model, therefore it is extremely important to verify that the model is consistent with the parameters of the actual phenomenon under investigation. The data input in this model is an artificial language, thus we have to be very careful about the predictions that we make from the model. Because we do not yet know the interactions between specific features of language to produce our perception, the artificial language used in this model doesn't assume any interactions between features of the language. The artificial language describes in an inclusive way that encompasses all the interactions of the features and express it as individual probability values. This model also tries to incorporate a continuum of probabilities to mimic the minimal pairs in the English language. While the artificial language may not represent the English language in every aspect, it does mirror certain aspects of language.

The variables of a mathematical model are limited to those that can be described in mathematical terms and may not be a realistic reflection of phenomenon. In this particular model, phenomenon is described in terms of probability and may encompass different types of information in just one number. It is near impossible to tease the information apart because we simply do not have enough understanding of language. However, the probabilistic approach allows us to study the phenomenon without really understanding it fully. Thus, this model cannot predict behaviours of individual characteristics of language.

In another effort to simplify language perception, only two processes in addition to the "sound matching" are included in this model. In reality, there are many more processes and factors that we do not understand and cannot yet describe. An advantage of this model is that future knowledge about the weights of context and frequency and their interactions can be incorporated. Additional processes can also be added to expand the model to make it more complete and realistic.

Despite every precaution, there are still potential errors in simplifications of parameters. There is a danger in believing the results of the model even if it is completely incorrect. It is essential to check the numerical model against observations. This paper will also compare the results from the mathematical model with the reaction time in a human experiment.

2.3 Experiment Description

A randomly generated list of probabilities that describes the sounds is used to form the artificial language used to test the model. 100 sounds were used to test the model. p(A) for the first phoneme ranges from 0.01 to 0.99, p(B) ranges from 0.99 to 0.01, and p(C) is always zero. p(A), p(B), and p(C) for the second and third phonemes are completely random and averages to be around 0.33. To evaluate whether the order of difference processes matter, the model was simulated twice: the first time the Context process is used before the Frequency process, and vice versa for the second time. The numbers of steps to achieve recognition as well as the number of words recognized at each stage were recorded.

3 Results

First, initial sound quality is compared with the number of steps it takes to achieve word recognition. Both simulations of the model shows that with increased probability based on input alone, the number of steps required to achieve decisions decreases (See Figure 3.1).

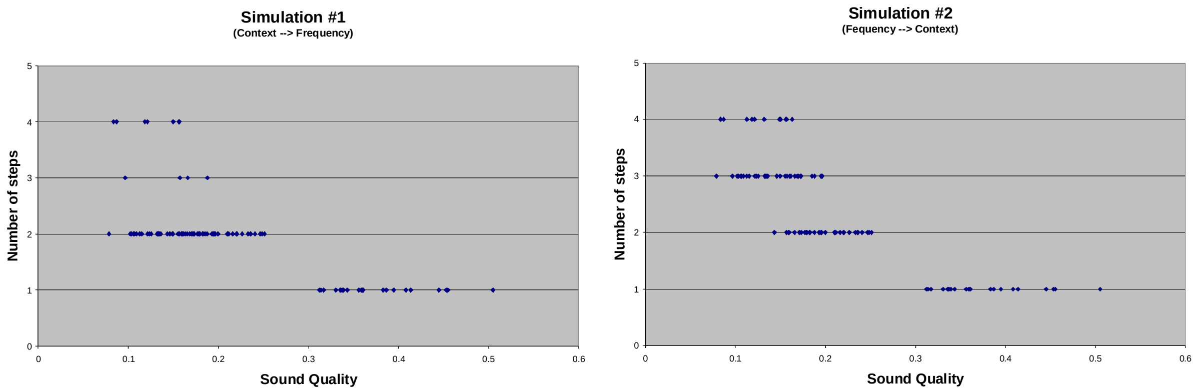

Figure 3.1. Number of Steps vs. Sound Quality. Simulation #1 and #2 show similar patterns. Number of steps required to achieve word recognition increases as sound quality decreases. Step 4 means that the word is never recognized.

When comparing the number of words recognized in each step of the processes (Table 3.1), a difference in actual number of words recognized at each step was found. More words are recognized by the Context process than by the Frequency process in general. When words are recognized by context already, frequency seems to have little effect on word recognition. There is always an upward trend where the more processes are involved, the more words are recognized. Only approximately 15 % of the words are recognized with information from the input alone. However, there is a dramatic increase to approximately 80 – 85% of all words recognized when the additional processes are used.

	Sound Match Alone	**Process 1**	**Process 2**
Simulation #1	14	81 (Context)	85 (Frequency)
Simulation #2	14	47 (Frequency)	81 (Context)

Table 3.1 Number of words recognized at each stage of the process. Context seems to be more effective at increasing the probability of word recognition than Frequency. The order of the processes also seems to have an effect on the number of words recognized – Frequency is only useful if there is no contextual information. However, the order of the processes does not effect the overall number of words recognized by the model.

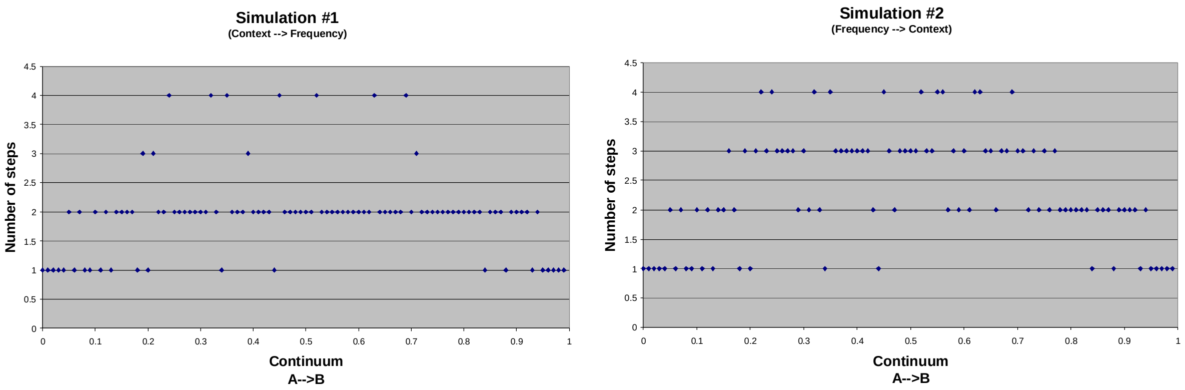

Figure 3.2. Number of Steps across a minimal pair continuum. Simulation #1 and #2 show similar patterns. The number of steps required to recognize the word decreases around the ends of the continuum where sound is more likely to represent the words. More steps are needed to recognize words the sounds in between the two words. If the words are too ambiguous, it may not be recognized at all.

When comparing the number of steps it takes to recognize words across a minimal pair continuum, words representing the ends of the continuum require less processing that words that are ambiguous. The patterns of very similar in both simulations, however simulation one obtained more words recognized in step 2 because of the order in which the model was tested. This result is similar to formats used in current research and can thus be compared.

In conclusion, the current model predicts that as input quality decreases, the processing time required to reach perception would increase. In addition, Frequency seems to have an effect only when contextual information is not available, establishing an order in which actual processes may follow. However, the order of the actual processes in reality can not be concluded from the current model, the model simply suggests that one might make more sense than the other.

4 Discussion

Current literature and research often uses the VOT continuum to study speech perception (e.g. Andruski, Blumestein, & Burton, 1994) because it is unfeasible to evaluate the “quality” of an input as a whole. To get around this problem, researchers degrade only one feature of the language, such as the VOT, to create a continuum of minimal pairs. For example, Blumstein (2004) looked at the reaction times (time it takes participants to respond to a speech sound) compared to the type of stimulus given. VOTs between 0 – 10ms describes one word in English, while a VOT of 40ms describes another. Thus, VOT of 20ms in English speech is an ambiguous sounding word (i.e.. Poor sound quality). Blumstein created a continuum between 0 and 40 ms and found increased reaction times across the phonetic boundary of 20ms where sound quality is poor (Figure 4.1). This finding is comparable to the one found in this paper (Figure 3.2).

As mentioned earlier in this paper, the number of steps taken to achieve word recognition is intended to represent the processing time. We assumed that the more steps are taken, the more processing time will be need, or in other words, more cognitive power is needed. Psychologists agree that increased reaction time is a result of heavier cognitive load, which is directly related to the amount or processes used in our brain. Although specific interactions and pathways of these processes are hard to map, we can make the general conclusion that more processes equals larger reaction times.

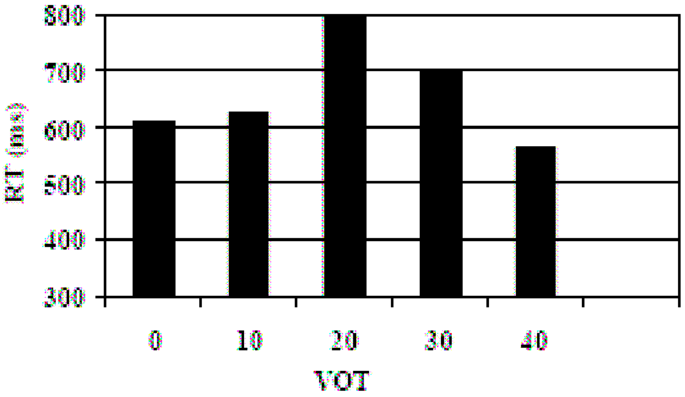

Figure 4.1. Reaction times across a VOT continuum. Graph taken from Blumstein, 2004. The graph shows increased reaction time across the phonetic boundary (20ms) where the sound is ambigurous.

The complexity of language and the difficulties in simplifying the phenomenon will be a task that researchers continue to overcome. In this model, contextual and frequency effects are described in a simple probabilistic approach where odds are simply added and normalized. A better approach may be to use the Bayesian probability theory where probabilities are measures of subjective belief rather than relative frequency of occurrences in an infinite sequence of trials. Many researchers believe that the philosophical tenet of Bayesianism may be more realistic than the existing frequentist statistics.

Recent studies have also tried involving neural networks to model the connections and interactions between different processes and factors (e.g. Stork, Wolff, & Levine, 1992). Neural networks are proposed by many researchers to be close conceptual representation of the human brain because of its ability to learn from large sets of data. The ability of neural networks to learn without explicit “instructions” and its similar performance patterns to the human brain seems to also offer a promising framework to build the language models.

The current model is a rudimentary representation of language processing and perception and further refinements and expansion will be needed to predict more aspects of language and be more representative of the actual phenomenon. Our limited knowledge at the present time as well as the limitations of this project prevents further insight into the problem. However, the model does predict that a more ambiguous sound would take more steps or processing time to match than an unambiguous sound, which is noted in actual observations.

4.1 Summary

A mathematical model is used to represent the process of language perception. Specifically, the model will address the question of how sound quality affects processing time. The model is constructed in Microsoft Excel and is tested with an artificial language consiting of three phonemes and 10 words. The model predicts increased processing time for ambiguous sounds than for non-ambiguous sounds and is representative of observations in reaction time patterns across VOT continuums. The current model is limited to predictions of processing time based on probabilities alone, future improvements can be made with the use of Bayesian Probability Theory and Neural Networks.

References

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*(3), 163-187.

Aydelott, J., & Bates, E. (2004). Effects of acoustic distortion and semantic context on lexical access. *Language and Cognitive Processes, 19*(1), 29-56.

Blumstein, S.E. (2004). Phonetic Category Structure and Its Influence on Lexical Processing. *Proceedings of the 2003 Texas Linguistics Society Conference. Somerville, MA, USA,* 17-25.

Connine, C.M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity Mapping in Spoken Word Recognition. *Journal of Memory and Language, 37, 463 – 480*.

Ganong, W.F. (1980). Phonetic Categorization in Auditory Word Perception. *Journal of Experimental Psychology: Human Perception & Performance, 6, 110 – 125.*

Hoff, E. (2001). Language Development (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.

McClelland, J.L., & Elman, J.L. (19860. The TRACE Model of Speech Perception. *Cognitive Psychology, 18, 1-86*.

Pisoni, D.B., & Luce, P.A.(1987). Acoustic-phonetic Representations in word recognition. *Cognition, 25, 21 – 52.*

Savin, H.B. (1963). Word-Frequency Effect and Errors in the Perception of Speech. *The Journal of the Acoustical Society of America, 35 (2), 200-206*.

Appendix A: Words in Artificial Language

AAA	word
AAB	nonword
AAC	nonword
ABA	word
ABB	word
ABC	nonword
ACA	word
ACB	word
ACC	nonword
BAA	word
BAB	nonword
BAC	word
BBA	nonword
BBB	word
BBC	nonword
BCA	nonword
BCB	word
BCC	word
CAA	nonword
CAB	nonword
CAC	nonword
CBA	nonword
CBB	nonword
CBC	nonword
CCA	nonword
CCB	nonword
CCC	nonword

Appendix B: Word Frequencies

AAA	0.13
AAB	0
AAC	0
ABA	0.12
ABB	0.08
ABC	0
ACA	0.1
ACB	0.09
ACC	0
BAA	0.13
BAB	0
BAC	0.08
BBA	0
BBB	0.07
BBC	0
BCA	0
BCB	0.11
BCC	0.09
CAA	0
CAB	0
CAC	0
CBA	0
CBB	0
CBC	0
CCA	0
CCB	0
CCC	0

Appendix C: Prime Relativity

	AAA	ABA	ABB	ACA	ACB	BAA	BAC	BBB	BCB	BCC
AAA	0.28	0.22	0.18	0.13	0.1	0.05	0.01	0.01	0.01	0.01
ABA	0.22	0.28	0.03	0.01	0.07	0.01	0.2	0.12	0.01	0.05
ABB	0.18	0.03	0.28	0.05	0.25	0.01	0.14	0.01	0.04	0.01
ACA	0.13	0.01	0.05	0.28	0.01	0.01	0.03	0.13	0.09	0.26
ACB	0.1	0.07	0.25	0.01	0.28	0.06	0.09	0.03	0.01	0.1
BAA	0.05	0.01	0.01	0.01	0.06	0.28	0.11	0.17	0.21	0.09
BAC	0.01	0.2	0.14	0.03	0.09	0.11	0.28	0.08	0.05	0.01
BBB	0.01	0.12	0.01	0.13	0.03	0.17	0.08	0.28	0.16	0.01
BCB	0.01	0.01	0.04	0.09	0.01	0.21	0.05	0.16	0.28	0.18
BCC	0.01	0.05	0.01	0.26	0.1	0.09	0.01	0.01	0.14	0.28
AAB	0	0	0	0	0	0	0	0	0	0
AAC	0	0	0	0	0	0	0	0	0	0
ABC	0	0	0	0	0	0	0	0	0	0
ACC	0	0	0	0	0	0	0	0	0	0
BAB	0	0	0	0	0	0	0	0	0	0
BBA	0	0	0	0	0	0	0	0	0	0
BBC	0	0	0	0	0	0	0	0	0	0
BCA	0	0	0	0	0	0	0	0	0	0
CAA	0	0	0	0	0	0	0	0	0	0
CAB	0	0	0	0	0	0	0	0	0	0
CAC	0	0	0	0	0	0	0	0	0	0
CBA	0	0	0	0	0	0	0	0	0	0
CBB	0	0	0	0	0	0	0	0	0	0
CBC	0	0	0	0	0	0	0	0	0	0
CCA	0	0	0	0	0	0	0	0	0	0
CCB	0	0	0	0	0	0	0	0	0	0
CCC	0	0	0	0	0	0	0	0	0	0

Grading Rubric

Both instructors will grade your work independently according to the criteria below (may not have equal weight). The final grade will be assigned by normalizing each instructor's evaluations (over all submissions) to have the same mean and variance (decided based on overall class performance), and averaging both instructors' normalized grades.

	Raw Score		
Criterion	Instructor:	Instructor:	Comments
Student worked independently without requiring too much instructor assistance.			
Motivation and research question clear and interesting from a scientific perspective.			
Model clearly explained.			
Model original and ambitious.			
Assumptions are thoroughly considered and well justified.			
Experiments are appropriate to answer research question.			
Experimental results clearly explained.			
Thoroughly explores implications of results and insights gained in regard to research question.			
The page limits were satisfied.			
Total =			Final Grade: