Linguistic Portfolios

Volume 5 Article 3

2016

An Optimaility Theoretic Approach to Child Language Acquisition

Colette Feehan
University of Minnesota, feeh0017@umn.edu

AN OPTIMALITY THEORETIC APPROACH TO CHILD LANGUAGE ACQUISITION

COLETTE M. J. L. FEEHAN

Abstract

A central goal of linguistics is to develop models to account for how children acquire their native language. One study that has done this is Boersma, Escudero, and Hayes (2003). In this study, the authors developed a model using an Optimality Theoretic approach to account for the ways in which children acquire language and language specific F1 frequencies into phonetic categories. The model uses a set of Optimality Theoretic constraints whose rankings gradually change in response to the learner's input. The way this model works, however, utilizes high-ranked discriminatory constraints in the initial state to produce movement in the Optimality Theory hierarchy. By taking this approach, it seems that the learner initially perceives incoming speech sounds as non-speech sounds and soon thereafter learns to categorize the sounds into the appropriate phonetic categories, resulting in language acquisition. In this paper, I have modified the proposal of Boersma, Escudero, and Hayes (2003) to avoid the counter-intuitive implications of perceiving incoming speech sounds as non-speech sounds. Instead of using high-ranked discriminatory constraints in the initial state, I have reversed the model to use high-ranked "perceive" constraints in the initial state. Reversal of the initial state is attractive because it no longer assumes that children inherently do not acknowledge speech sounds at the beginning of language acquisition. This paper gives an alternative perspective on this language acquisition model and is designed to explore different ways to account for how the infant brain acquires language at the phonetic level. Additionally, the current model also acknowledges frequencies that are not present in the training data where the previous model categorizes unfamiliar frequencies as non-speech sounds.

1.0 Introduction and Background[1]

In this paper, I have set out to modify an existing Optimality Theoretic model of child language acquisition. In Boersma, Escudero, and Hayes' (2003) report, the authors created a model to show how children learn their native language by warping their perceptual space to account for only the sounds that are necessary in their language. While the previous model works well, I have simply attempted to create a different perspective that takes into account other types of research in the field.

Optimality Theory (OT) is a linguistic theory that provides a means of analyzing language in terms of ranked constraints instead of rule ordering. Some of the basic motivations behind Optimality Theory are to create a model for language analysis that is fairly generalizable and accounts for aspects of Universal Grammar (McCarthy, 2007). Instead of a rigid set of ordered, language-specific rules to account for language, Optimality Theory utilizes a hierarchy of inherently violable constraints that can be rearranged, depending on the language in use. This idea of being inherently violable allows the same hierarchy to account for "exceptions to rules"

[1] Acknowledgements: I would like to thank Ania Lubowicz and Randy Fletcher for providing the literature references that led me to choosing this topic. I also want to say thank you to Timothy Hunter for advising this project and helping me through every step of the process.

by showing that the optimal output, or optimal candidate, can still violate aspects of the grammar—it simply has violations that are not ranked highest (see figure 1). This provides linguists with an effective model to analyze languages using the same constraints in different rankings.

pat	NOCODA	PARSE
☞ pa		*
pat	*!	

Figure 1. While "pa" still has a violation mark, "pat" has a higher ranked violation leading "pa" to be deemed the optimal candidate.
Taken from *OT Learning* (Boersma, 2007).

Figure 1 shows that "pa" wins as the optimal candidate despite having a violation because the PARSE constraint is ranked lower than NOCODA. This means that anything that violates NOCODA (such as "pat") will lose to "pa."

Another advantage of Optimality Theory is that it can account for abnormalities that would otherwise be called 'exceptions to rules.' It does this by having inherently violable constraints where a candidate with high-ranked violations is undesirable, and a candidate with few or low-ranked violations can become the optimal candidate (the output) despite violating some constraints (as seen in figure 1). Another idea that some linguists propose is that the constraints are innate, meaning that humans have a basic set of constraints at birth and simply re-rank these constraints according to data from their native language during acquisition.

Optimality Theory has been very useful in the field of child language acquisition. Because the theory is built from the concept of moving and re-ranking constraints, it lends itself quite well to analyzing the prototype languages that children create and the steps they take to rearrange OT constraints on the path to developing an adult grammar. One notable study showing Optimality Theory's usage in child language acquisition is a study by Amalia Gnanadesikan (1995), in which the author detailed the steps her child took to eventually acquire an adult grammar. The basic idea behind an Optimality Theoretic approach to child language acquisition is that children begin with one constraint ranking and adjust the hierarchy as they develop, resulting in what should be an adult grammar.

Optimality Theory also ties in well with the idea of *Universal Grammar*. As defined by Steven Pinker (1994), Universal Grammar is, "the circuitry in children's brains that allows them to learn the grammar of their parents' language" (p. 515). The term Universal Grammar, however, can be somewhat misleading. By using the term, linguists are not saying that each child is born with all of the grammatical rules of every language, but rather there is a "restricted set of candidate grammars" from which children can gather grammatical rules (Nowak, Komarova, and Niyogi, 2002). The idea for Universal Grammar comes from a paradox called the *poverty of stimulus* (Nowak et al., 2002). Poverty of stimulus is the term for the fact that children in the same language community reliably grow up to speak the same language even when adults and children do not know the specific grammar rules of the language—they are able to learn a generative grammar simply by listening to grammatical sentences from adults. As Noam Chomsky explained, poverty of stimulus creates a need for a Universal Grammar and mathematical models of language acquisition support the idea that Universal Grammar is a logical necessity (Novak et al., 2002).

For years, many linguists such as Steven Pinker have been strong proponents of the idea that language and its acquisition is at least partially an innate or even genetically coded process that every human is born with. Beyond theoretical conjectures, there is even evidence of a gene involved in language production and development. A mutation in the FOXP2 gene has been strongly implicated in Specific Language Impairment which inhibits a person's ability to perform several language related tasks involved in speech production and even comprehension (Lai et al., 2001). Many linguists have adopted Universal Grammar, but it remains a controversial topic, though the controversy behind it has changed. Initially, the very idea of an innate component of language was Earth shattering, but the argument has now generally shifted to ask what form Universal Grammar takes and not whether it actually exists (Nowak et al., 2002).

Some linguists, like Patricia Kuhl, have conducted research to try to find out just how much of our language abilities are innate. Clearly, humans are not born with a proper adult grammar, but, many would argue that there must be some innate portion. According to Kuhl (2000), children seem to possess surprisingly complex, innate strategies for efficiently learning language. She states, "Infants have inherent perceptual biases that segment phonetic units without providing innate descriptions of them" (Kuhl, 2000). So while infants may not have innate grammatical structures specific to any one language that already exists on Earth, Kuhl (2000) provides strong evidence that they do have some level of innate ability to perceive, organize, and acquire language quickly and efficiently.

According to Kuhl (2006), children have extensive knowledge of language far before they can actually speak. Many studies have found evidence for Universal Grammar and perceptual reorganization of the phonetic space (Werker & Tees, 1984; Kuhl et al., 2005; Tsao, Liu, & Kuhl 2006; Kuhl et al., 2006). Kuhl and her colleagues are arguably at the forefront of this field as they provide compelling evidence that children have the ability to acquire the basic units of sound in any language but lose this generalizable skill between six and twelve months of age. During this time, children undergo some process where their perceptual space becomes warped to best accommodate the language they are learning. For example, in Kuhl et al. (2006), children learning American English and children learning Japanese were given a test to see if they could distinguish between /ra/ and /la/—sounds that do not form a minimal pair in Japanese but do in American English. At 6-8 months, all the children did comparably well on the test, but at 10-12 months, the children learning Japanese were markedly worse at distinguishing the sound—and the children learning English were well above chance levels. This shows that at some point between six and twelve months children begin this warping process and begin to better perceive only the sounds that are useful for their native language. Essentially, 'warping' means that the children initially possess one continuous space of possible frequencies, and at some point, the frequencies that are most essential to the language "take over" the perceptual space resulting in more representation of essential frequencies and a decline in representation of frequencies that are deemed non-essential. In essence, after warping has occurred, the inputs remain the same continuous space of frequencies. The inputs that are near each other, however, are grouped together into a smaller number of output categories. One explanation for this phenomenon comes from Best and McRoberts (2003) who provide evidence that the specific articulatory organ of the sounds in question plays a role in this decline. In this study, Best and McRoberts argue that non-native discrimination is more likely to decline when the contrasting phonemes involve the same articulatory organ such as /s/ versus /z/ as opposed to /p/ versus /d/. For example, in Kuhl et al. (2006) the sounds /ra/ and /la/ are both alveolar liquids and would, according to Best and McRoberts (2003) be subject to decline in non-native discrimination tasks. In another study by Kuhl, Tsao, and Liu (2003), the researchers gave children who were learning American English play sessions in Chinese before 10-12 months of age. By doing this, the characteristic decline in non-native phonemic discrimination was reversed, and they were able to maintain distinctions in sounds that are utilized in Chinese only. They also found that this was more robust for play sessions involving a person speaking to the children rather than watching videos or listening to tapes—meaning that children are much more attentive to speech sounds when in the presence of other people.

Another interesting study was carried out by Querleu, Renard, Versyp, Paris-Delrue, & Crepin (1998). In this study, the researchers found that the fetus can begin to recognize speech sounds and intonation patters during the third trimester in the womb. This suggests that linguistic learning does not begin at birth but before—children are born with a very small base of linguistic knowledge. The findings of these studies combined are incredibly important in developing a model for child language acquisition because they imply that children are, in a way, taking statistics of the speech sounds they hear and are able to distinguish their native speech sounds far before they are able to produce them.

As a tool in developing models for child language acquisition, Optimality Theoretic constraints that favor ungrammatical outputs are said to be demoted through the hierarchy until they no longer have any effect on the output grammar (Tesar & Smolensky, 2000). From an Optimality Theoretic perspective, exposure to more than one language in early development, as found in Kuhl et al. (2003), could be enough to prevent some constraints from being entirely demoted, meaning that they would still maintain a minor level of activation that allows the child to maintain subtle phonological differences would not maintain, like /ra/ and /la/ in Kuhl et al. (2006).

One practical model in acquisition based OT is the Gradual Learning Algorithm developed by Paul Boersma, which is the model I will use in this paper. The Gradual Learning Algorithm (GLA) is an error-driven model, which means that it alters rankings in the OT hierarchy only when the input conflicts with the current constraint ranking (Boersma & Hayes, 1999). The GLA also assumes a continuous spectrum of constraint strictnesses rather than distinct rankings with the main intention of preventing complete, one step re-ranking of the hierarchy. Instead, the continuous scale allows the model to adjust slowly and gradually. It does this by temporarily providing a small amount of "noise," or variance, to each constraint's ranked value in order to produce variable outputs in cases where multiple constraints are tied or nearly tied. Additionally, there are many advantages of using the GLA in child language acquisition, such as the fact that it can easily generate multiple outputs, and it is not thrown off by speech errors (Boersma & Hayes, 1999).

In this paper, I will address an article by Boersma, Escudero, and Hayes (2003). In the article, the authors present a model for child language acquisition on the phonetic level. This means that the model is assessing language acquisition of phones that occur far before phonological learning takes place. They only address F1 values of vowel sounds, meaning that they are not full phonemic vowels but, simply, the first part of a vowel sound. In their OT grammar, the authors use three types of constraints (see section 2.2.2). While their grammar does provide an account of how children warp their perceptual space to develop their native language, I wanted to adjust their model and see if it would yield the same results. They begin with *Categorize constraints (*CATEG) ranked quite high, which means that initially, the grammar consistently tells the child to perceive the null category (/-/) in response to an incoming F1 instead of perceiving it as an actual speech sound. In the original article, the authors are not very explicit in their definition of null category. Because of this, I have come to interpret the null category as background noise—meaning that an optimal null category signifies that the child does not acknowledge the incoming frequency as a speech sound, but rather as a type of background noise. This initial state seems to suggest that children, at least for a short amount of time, do not interpret incoming speech sounds as speech sounds but rather as some type of background noise. By adapting this meaning, the model also seems to suggest that at birth, the learner must first learn to identify speech sounds and then learn to categorize speech sounds. This conclusion seems to contradict the research in fetal hearing by Querleu et al. (1998) and also creates a type of "chicken or the egg" problem in the sense that if the child's initial grammar is consistently telling the child not to acknowledge incoming speech sounds as speech sounds, why does the child ever begin acknowledging or categorizing them?

In this paper, I have developed a model using the constraints in Boersma et al. (2003) with a different initial state. By changing the initial state, this model better takes into account, previous literature on child language acquisition and does not have the "chicken or the egg" problem. Advantages of the new model, however, are not limited to this. The new model also acknowledges frequencies that do not occur in the training data. Instead of categorizing unfamiliar frequencies as the null category (or as non-speech sounds) as seen in the previous model, the new model still tries to warp these unfamiliar frequencies into the most similar category available from the training data or tries to create another warped category when the frequencies are too far away from an existing category.

Method

2.1 Purpose

The purpose of this paper is to develop a functional model for child language acquisition using Optimality Theory that reflects previous research findings in Universal Grammar, pruning of the perceptual space and fetal learning. The main goal of this project is to experiment with a different initial state from that found in Boersma et al. (2003) and to attempt to find an initial grammatical state that avoids choosing the null category as the optimal candidate, but which also produces an analogous end-state grammar. I did this by first replicating the Boersma et al. (2003) model and then creating the new model.

2.2 Materials

2.2.1 Data Files

Praat phonetic software was the main program for analysis in this paper. It was used to format Optimality Theory tableaux and run learning simulations involving training data. Python coding software was used to create the OT tableaux files and training data files. Each tableau in the file had 122 constraints with 45 candidates. The constraints consisted of 44 Perceive constraints, 44 *Categorize constraints, and 34 *Warp constraints; the output candidates were all frequency values ranging from 200 Hz to 1000 Hz increasing by steps of 20 and the null category (/-/). The full file had 45 tableaux each, with the parameters described above and an input frequency within the range of 180 Hz to 1000 Hz increasing by steps of 20.

The training data file was created by taking 400,000 random frequency values from a set of four overlapping normal distributions centered on frequencies 280 Hz, 420 Hz, 560 Hz, and 700 Hz, in accordance with the methods in Boersma et al. (2003). Once created, the training data file was ran through the OT grammar in Praat using the Gradual Learning Algorithm to show how the child would learn according to a given initial grammar.

2.2.2 Constraints

The constraints used in the OT analysis come from Boersma et al. (2003). There are three types of constraints in this model. The first type of constraints are Perceive constraints (PERCEIVE), which act as a type of faithfulness constraint (meaning that these constraints try to prevent differences between the input and the optimal candidate). PERCEIVE constraints want input frequencies to be treated as members of some category. This means that an input F1 should be perceived as some frequency, not necessarily the actual input frequency i.e., the null category always violates PERCEIVE constraints based on the input value and any frequency value is accepted by all PERCEIVE constraints (see figure 2).

[500]	Perceive[480]	Perceive[500]	Perceive[520]
☞ /480/			
☞ /500/			
☞ /520/			
/-/		*	

Figure 2. Only the null violates PERCEIVE constraints, taking the input value into consideration

Examples: PERCIEVE[520]- The null candidate will receive a violation when the input value is also [520]
PERCEIVE[700]- The null candidate will receive a violation when the input value is also [700]

The second type of constraints are *Categorize constraints (*CATEG), which function as a type of markedness constraint (these constraints try to produce differences between the input and the optimal candidate). *CATEG constraints want particular frequencies to not be treated as members of that specific category. This means that an input F1 should not be perceived as that specific frequency e.g., candidate /500/ will always violate constraint,*CATEG/500/ (see figure 3).

[500]	*Categ/480/	*Categ/500/	*Categ/520/
/480/	*		
/500/		*	
/520/			*
☞ /-/			

Figure 3. The null is the only candidate with no violations and thus is the optimal candidate
Only a candidate that matches the constraint value will receive a violation

Examples: *CATEG/520/- Candidate /520/ always receives a violation
*CATEG/700/- Candidate /700/ always receives a violation

The final type of constraints are *Warp constraints (*WARP), which are another type of faithfulness constraint. *WARP constraints essentially tell the learner a range in which each incoming frequency should be placed e.g., a *WARP40 constraint tells the learner that a frequency should not be placed into a category greater than or equal to 40 hertz above or below the input value (see figure 4).

[480]	*Warp20	*Warp40	*Warp60
☞ /480/			
/500/	*		
/520/	*	*	
☞ /-/			

Figure 4. Candidates with values greater than or equal to the constraint's number receive a violation i.e. *WARP20 is violated by frequencies 20 hertz or more away from the input.

*WARP constraints reflect the results from Kuhl et al. (2000) very well. The job of *WARP constraints is to adjust or warp the perceptual space and they want input frequencies to be treated as members of the most similar available category. This means that as the learner grows the frequencies that are most useful in their native language become dominant and frequencies that do not fall into these dominant categories will be perceived as a member of the most similar, dominant frequency category (or grouped into the most similar category).

2.3 Procedure

Using the constraints detailed in 2.2.2, I have created a new model with a different initial state. As stated in section 1, the Boersma et al. (2003) model has a unique problem associated with its initial state, where the initial grammar seems to assume that children are starting without any previous learning when they are born, which discounts research in fetal learning. With the *CATEG constraints initially ranked very high (see figure 5 in section 3), the grammar consistently tells the child to not perceive speech sounds as their specific frequency value leading to the question of why the child ever begins to acknowledge and categorize speech sounds in the first place. The original model also seems to imply that when a child is born, he or she must first learn to identify speech sounds (as evidenced by choosing the null category as the optimal candidate in the initial state) and then learn to categorize them. This assumption seems to contradict research that shows that children first begin to recognize speech sounds while in the womb (Querleu et al., 1988). To avoid these problems, the current model contains PERCEIVE constraints that are ranked high in the beginning, *CATEG constraints that are ranked low, and the *WARP constraints maintain the same ranking as in the original model. Tables 1 and 2 compare the initial height rankings in both the Boersma et al. (2003) model and the current model. Red cells indicate changes made in the new proposal.

Constraint	**Initial Ranking**
*WARP800	800
*WARP780	780
…	…
*WARP60	60
All *CATEG	0
All PERCEIVE	-1000
*WARP40	-10^9
*WARP20	-10^9

Table 1. Initial height ranking in Boersma et al. (2003)

Constraint	Initial Ranking
*WARP800	800
*WARP780	780
…	…
*WARP60	60
All PERCEIVE	0
All *CATEG	-1000
*WARP40	-10^9
*WARP20	-10^9

Table 2. Initial height ranking in the current study

While both models are effective at producing warping of the perceptual space, I will show that the current study better encompasses previous research in Universal Grammar and fetal learning. It does this by removing the idea that children initially perceive only the null category and reshuffle their constraints in order to begin perceiving sounds only after they grow bored of perceiving the null category. This means that the proposed model assumes that children begin learning and taking language statistics slightly before birth meaning that they can already *identify* speech sounds at birth, but they now need to learn to *categorize* speech sounds. With the PERCEIVE constraints initially high, all frequency values will tie and have an equal chance of being correctly or incorrectly deemed optimal (see figure 6 in section 3). While this does mean that the children will be accepting incorrect optimal candidates at first, once they have encountered a specific frequency a sufficient number of times in the training data, learning will take place and the constraints should be rearranged into the proper adult hierarchy.

As explained above (see table 2), in this study, the PERCEIVE constraints were placed at a height ranking of 0 in the initial grammar and the *CATEG constraints were placed at -1000. The *WARP constraints were placed at a level equal to the frequency value it addressed e.g., a *WARP800 constraint would be initially placed at 800 and *WARP240 would be initially ranked at a height of 240. The only exceptions to this pattern were *WARP20 and *WARP40 which were both placed at -1,000,000,000, in accordance with the original article.

In order to simulate learning, the plasticity rating in Praat had to be adjusted. Plasticity, as defined in Boersma et al. (2003), "is the size of the step by which rankings can change on each input." This essentially is the variable that determines how flexible the grammar is and how susceptible the grammar is to new data. Because children are so much more efficient at learning language than adults, it is said that children have a higher plasticity—this is why an adult does not completely re-rank their OT hierarchy when they listen to a different language or accent or when they encounter speech errors. A lower plasticity ranking results in more rigidity in the hierarchy and less learning. In Praat software, plasticity is the determining factor in how much 'learning' takes place after each encounter of a datum point in the training data and, in turn, how much each constraint moves in the hierarchy as a result of these encounters. In this paper, the initial plasticity level was set at 1 which is much higher than an adult plasticity level of .001, and the basic assumption is that as the child learns, their plasticity rating slowly reverts to .001 as they reach an adult grammar and they become much less likely to re-rank their hierarchy. In the current study, I could not simulate the reversion of the plasticity level to .001 making the plasticity throughout the simulation 1. I was, however, able to replicate the findings in Boersma et al. (2003) which suggests that this difference is minimal. Once the initial state, constraints, and plasticity level were determined, tableaux were coded into Praat and run through the Gradual Learning Algorithm in order to show the ultimate constraint ranking after utilizing the training data.

3.0 Results

3.1 Main Findings

One of the main differences between the current model and Boersma et al. (2003) is the initial state in the Optimality Theoretic grammar. In Optimality Theory, every time the initial grammar encounters a datum point the constraints rearrange accordingly. Once the model has seen the same datum point enough times constraints will begin to flip creating different outputs, e.g., in the previous model, after several encounters of a datum point, the constraints *CATEG480 and PERCEIVE480 will switch places and the learner will begin perceiving [480] as a speech sound. As stated above (see table 1), the initial ranking in Boersma et al. (2003) utilized high ranked *CATEG constraints and low ranked PERCEIVE constraints that choose the null category as the optimal candidate as seen in figure 5.

[500]	*Categ /480/	*Categ /500/	*Categ /520/	Perceive[480]	Perceive[500]	Perceive[520]
/480/	*					
/500/		*				
/520/			*			
☞ /-/					*	

Figure 5. High ranked *CATEG constraints result in the null category being the optimal candidate

The new proposed model switches the values of the *CATEG and PERCEIVE constraints to have very high ranked PERCEIVE constraints and very low ranked *CATEG constraints which causes a tie among all the candidates as seen in figure 6.

[500]	Perceive[480]	Perceive[500]	Perceive[520]	*Categ/480/	*Categ/500/	*Categ/520/
☞ /480/				*		
☞ /500/					*	
☞ /520/						*
/-/		*				

Figure 6. High ranked PERCEIVE constraints result in a tie among all the candidates except the null

Figure 6 shows that there is a tie among all the possible candidate frequency values meaning that the learner will choose a random output from the tie until he or she has received enough training data (heard adults speak enough) to refine the perceptual space and eventually choose the correct optimal candidate. Initially, this tie is broken by the evaluation “noise” produced by the GLA, but then begins to work autonomously when constraint rankings diverge based on data. This option is quite appealing because it means that the child’s grammar is telling it to perceive speech sounds as important, they simply do not know how to categorize them, as opposed to coming out of the womb with absolutely no linguistic knowledge other than a pile of ranked constraints. This essentially means that the child acknowledges that speaking is occurring, but she or he, understandably, does not yet know how to interpret it.

Once the training data was presented to the two initial states, I found that both models result in similar warping of the perceptual space (see tables 3 and 4)

<table>
<tr><th>Input</th><th>Output</th></tr>
<tr><td>[240]</td><td rowspan="5">/280/</td></tr>
<tr><td>[260]</td></tr>
<tr><td>[280]</td></tr>
<tr><td>[300]</td></tr>
<tr><td>[320]</td></tr>
<tr><td>[340]</td><td rowspan="2">/320/</td></tr>
<tr><td>[360]</td></tr>
<tr><td>[380]</td><td rowspan="2">/360/</td></tr>
<tr><td>[400]</td></tr>
<tr><td>[420]</td><td rowspan="5">/460/</td></tr>
<tr><td>[440]</td></tr>
<tr><td>[460]</td></tr>
<tr><td>[480]</td></tr>
<tr><td>[500]</td></tr>
</table>

Table 3. Boersma et. al. (2003) replication

<table>
<tr><th>Input</th><th>Output</th></tr>
<tr><td>[240]</td><td rowspan="3">/240/</td></tr>
<tr><td>[260]</td></tr>
<tr><td>[280]</td></tr>
<tr><td>[300]</td><td rowspan="2">/280/</td></tr>
<tr><td>[320]</td></tr>
<tr><td>[340]</td><td rowspan="2">/320/</td></tr>
<tr><td>[360]</td></tr>
<tr><td>[380]</td><td>/400/</td></tr>
<tr><td>[400]</td><td rowspan="5">/440/</td></tr>
<tr><td>[420]</td></tr>
<tr><td>[440]</td></tr>
<tr><td>[460]</td></tr>
<tr><td>[480]</td></tr>
<tr><td>[500]</td><td>/460/</td></tr>
</table>

Table 4. Warping in current model

In tables 3 and 4, it is clear that the two models provide similar warping of the perceptual space. Table 3 shows the values in my replication of the original study, and table 4 shows the values obtained from the current model. In both instances, multiple input frequencies are clumped into categories that would be similar to an F1 value of either; a high front vowel (240-280 group) or of a back vowel (440-460 group). While the frequency values are not exactly the same, they are all within a plausible range of F1 frequencies of two types of vowels, and by using randomly generated data, this sort of variation is expected. The same type of clumping seen around 240-280 Hz and 440-460 Hz can be found near each of the four distribution points mentioned in section 2.2.1. This warping shows that the left inputs that became clumped are not differentiated in the language and thus they become clumped into one general vowel category. Figures 7 and 8 are a simple comparison of the training data next to the warped output from both models. These figures simply help visually conceptualize how the warping lines up with the normal distributions of the training data.

Boersma et al.

Figure 7. Boersma et al. (2003) warping beside the training data distribution

Warping

200	180
220	
240	240
260	
280	
300	280
320	
340	320
360	
380	400
400	440
420	
440	
460	
480	
500	460
520	520
540	580
560	
580	
600	
620	
640	600
660	700
680	
700	740
720	
740	
760	
780	
800	760
820	780

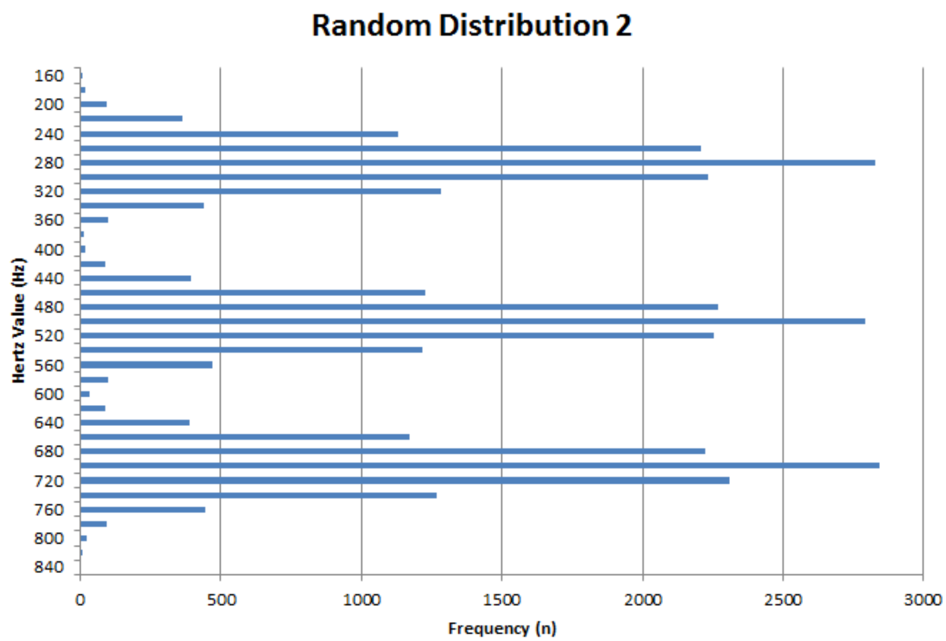
Current Study

Figure 8. Warping found in the current study beside the training data

After obtaining this result, I created more training data files using the same distribution and ran them through the model to make sure the results were replicable. From these files, I found that these findings are quite robust. While none of the adult grammars were exact replicas of each other due to variation "noise" in the GLA, similar learning and warping resulted from each training data file. In summary, both models, while using different initial states, consistently produced warping of the perceptual space and similar, competent adult grammars.

3.2 Additional Findings

After observing the results from the two models, I decided to experiment with training data that utilized different normal distributions to see if the models can still produce warping with different distributions. Figures 9 and 10 show how each model warped data from a tri-modal distribution rather than a distribution with four normal distributions.

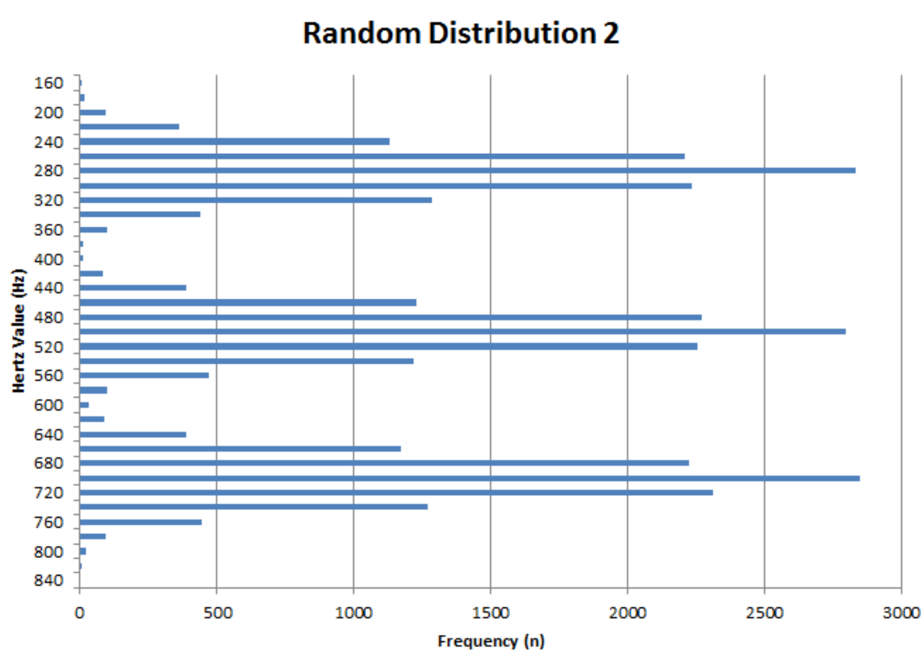


Warping

200	180
220	
240	220
260	300
280	
300	
320	
340	
360	320
380	340
400	/-/
420	420
440	480
460	
480	
500	
520	
540	500
560	560
580	
600	/-/
620	640
640	
660	
680	
700	720
720	
740	
760	
780	740

Figure 9. Boersma et. al. (2003) warping beside training data using only three normal distributions rather than four

Boersma et. al.



Warping

200	180
220	
240	260
260	
280	320
300	
320	
340	
360	
380	340
400	360
420	440
440	
460	
480	520
500	
520	
540	
560	
580	560
600	
620	660
640	
660	700
680	
700	
720	
740	
760	720
780	760

Figure 10. Warping in the current model beside training data using only three normal distributions rather than four

Current Study

Figures 9 and 10 show that the models are quite robust and do in fact produce warping when given different training data. This result could imply that the models would function with data from different languages.

An unexpected finding in the experiment, however, was how the Boersma et al. (2003) model treated frequencies for which there were no training data. Tables 5 and 6 show the results from the training data that contained four, overlapping normal distributions. For input frequencies higher than 840 Hz (frequencies that were not included in the training data), the previous model assigned the null category as the optimal candidate and the new model did not.

Input	Output
820	
840	
860	
880	
900	/-/
920	
940	
960	
980	

Table 5. Boersma et. al. (2003) model addressing frequency values that were not in the training data

Input	Output
820	780
840	800
860	820
880	840
900	920
920	
940	
960	980
980	

Table 6. The current model addressing frequency values that were not in the training data

Tables 5 and 6 show the warping from training data with four normal distributions. The previous model discounted unfamiliar frequencies as non-speech sounds whereas the current model acknowledges that speech is occurring and even tries to create another warping concentration above the last normal distribution in the training data file. The current model tended to group these high frequencies into lower categories around 800-850 Hz which would roughly correspond to F1 values found in low vowels. This shows that the current model was able to categorize fairly uncommon sounds to the closest similar category (low vowel F1 values) and the previous model simply discounted them and placed them into the null category. Additionally, when the frequencies were too far from the nearest available category that has been created as a result of exposure to the training data, the new model began to create a new warping point around /980/ to try to maintain the clumping pattern. This implies that even when PERCIEVE and *CATEG constraint heights have not been affected by any training data (because there were no input values of that specific frequency) they still come into play. By having the PERCEIVE constraint inherently higher than *CATEG constraints, we do not get the null category winning. Then the *WARP constraints take over and continue to warp the perceptual space to create the new warping category in the higher frequencies.

This null category assignment was also evident in the experiment with tri-modal training data. Tables 7 and 8 show that when the training data is spread out to the point where the normal distributions do not overlap, the Boersma et al. (2003) model produces "holes" in the data where it assigns the null candidate to frequencies between the normal distributions. It does this because there are not enough data points of those frequencies to cause the *CATEG and PERCEIVE constraints to "flip," i.e., the *CATEG constraint remains higher than the PERCEIVE constraint (due to lack of data at that frequency) thus causing the null category to be chosen as optimal.

<table>
<tr><th>Input</th><th>Output</th></tr>
<tr><td>260</td><td rowspan="5">300</td></tr>
<tr><td>280</td></tr>
<tr><td>300</td></tr>
<tr><td>320</td></tr>
<tr><td>340</td></tr>
<tr><td>360</td><td>320</td></tr>
<tr><td>380</td><td>340</td></tr>
<tr><td>400</td><td>/-/</td></tr>
<tr><td>420</td><td>420</td></tr>
<tr><td>440</td><td rowspan="5">480</td></tr>
<tr><td>460</td></tr>
<tr><td>480</td></tr>
<tr><td>500</td></tr>
<tr><td>520</td></tr>
<tr><td>540</td><td>500</td></tr>
<tr><td>560</td><td rowspan="2">560</td></tr>
<tr><td>580</td></tr>
<tr><td>600</td><td>/-/</td></tr>
<tr><td>620</td><td rowspan="4">640</td></tr>
<tr><td>640</td></tr>
<tr><td>660</td></tr>
<tr><td>680</td></tr>
</table>

Table 7. Warping in the Boersma et. al. (2003) model when using a tri-modal distribution

<table>
<tr><th>Input</th><th>Output</th></tr>
<tr><td>260</td><td>260</td></tr>
<tr><td>280</td><td rowspan="5">320</td></tr>
<tr><td>300</td></tr>
<tr><td>320</td></tr>
<tr><td>340</td></tr>
<tr><td>360</td></tr>
<tr><td>380</td><td>340</td></tr>
<tr><td>400</td><td>360</td></tr>
<tr><td>420</td><td rowspan="3">440</td></tr>
<tr><td>440</td></tr>
<tr><td>460</td></tr>
<tr><td>480</td><td rowspan="5">520</td></tr>
<tr><td>500</td></tr>
<tr><td>520</td></tr>
<tr><td>540</td></tr>
<tr><td>560</td></tr>
<tr><td>580</td><td rowspan="2">560</td></tr>
<tr><td>600</td></tr>
<tr><td>620</td><td rowspan="2">660</td></tr>
<tr><td>640</td></tr>
<tr><td>660</td><td rowspan="2">700</td></tr>
<tr><td>680</td></tr>
</table>

Table 8. Warping in the current model when using a tri-modal distribution

With "foreign" frequencies, the Boersma et al. (2003) model simply deems the null category as the optimal candidate, which seems counter intuitive. If we still assume that the null category means "not acknowledging an incoming frequency as a speech sound," then this result implies that when an adult hears speech values that are not in their native language they will discount them as non-speech sounds. The current model, however, shows that while the learner has never encountered these sounds before (or has had very few encounters with them), they still acknowledge that speech is occurring and even try to warp the unfamiliar sounds into categories that they are already familiar with.

4.0 Discussion

These results show that both models produce similar adult grammars. While the frequency (Hz) categories are slightly different, they both show analogous warping of the perceptual space. One possible explanation of these differences is the evaluation "noise" that the Gradual Learning Algorithm assigns to each constraint before learning takes place. The "noise" is added as a way to break up ties and jumpstart movement in the hierarchy, but by doing this, it also creates variation in the outputs. Also, by nature of using randomly generated datum points, a certain amount of variation is expected among trials. Another possible source of the differences could be that the models need more data or "time" so to speak. The number of training data used in this study was determined by using what was found in Boersma et al. (2003) (to reduce possible confounds). It is possible, however, that the models differ because the grammars are still not entirely settled and would continue to change slightly if they received more data. Meaning that some constraints may continue to shift, resulting in a slightly different outcome.

The current model was able to create proper warping of the perceptual space without the "chicken or the egg" problem present in the previous model. This means that we do not have to claim that the child gets bored of perceiving the null, but instead, he or she is already attending to speech sounds. Not only is this attractive theoretically, it is also more consistent with previous research. In the new initial state (see figure 6) the learners' grammar is confronted by a massive tie among all the frequency values, but they are aware that they should be treating the incoming sounds as a speech sounds. This is consistent with Kuhl et al. (2003), who provided robust evidence that children are continuously "taking statistics" on the sounds around them far before they are able to produce these sounds. Instead of needing to learn to start attending to speech sounds and then categorizing them as found in Boersma et al. (2003), the current model requires children to simply begin categorizing.

One very interesting and unexpected result was that the current model worked better with very high frequencies (greater than 800 Hz) i.e., the frequencies that did not occur in the training data (see tables 5 and 6). At those levels, the Boersma et al. (2003) model tended to choose the null category as the optimal candidate, however, the current model attempted to rank them into the most similar category that the learner already possessed in its learned grammar. Once the frequencies got too far from the most similar existing category, the new model even tried to continue the warping pattern and create a warped center around frequency /980/ (see table 6). The new model also filled in gaps when the training data was tri-modal (see figure 10 and table 8). This means that even though the training data provided few data points around /400/ and /600/ the new model was able to categorize them as a speech sound and the previous model produced the null category.

An interesting direction for further research would be cross-linguistic studies. It would be very intriguing to see if this type of approach functions with training data from languages other than English. In order to explore this option, I experimented with different training data distributions (see figures 9 and 10) in both models. From this, I found that both models were able to create warping centered on different normal distributions, which indicates that it is possible that these models could be used in cross-linguistic studies. It would also be interesting to see what would happen to the adult grammar if the learner was presented with two training distributions simultaneously to simulate what a child raised in a bilingual environment would encounter.

One more possible direction, would be to observe the adult grammars that result from models with different initial states such as having all the constraints at equal heights in the initial state or have the *WARP constraints all equal like the current *CATEG and PERCEIVE constraints. Once could also experiment with removing the null category entirely or creating a *CATEG(/-/) constraint.

5.0 Conclusion

Clearly, both the current proposal and the Boersma et al. (2003) study have provided a means of analyzing how children learn language on a phonetic basis. Both models result in warping of the perceptual space around the sounds that were most common in the training data. While the new model is simply a different perspective on the same topic, the current model does have a few advantages. While both depict the characteristic warping of the perceptual space described in Kuhl et al. (2006), I have shown that the current model considers and better encompasses previous research in fetal hearing and Universal Grammar. I have also shown that this model accounts for "holes" in the training data to prevent then null category from surfacing in the adult grammar.

ABOUT THE AUTHOR

Colette Feehan is now a graduate student in general linguistics at Indiana University. She worked on this paper for her undergraduate honors thesis at the University of Minnesota, Twin Cities. Since completing her BA in linguistics and psychology at Minnesota she spent a year working in the Speech Acquisition Lab at the University of Utah and has now begun her Ph.D. at Indiana University with a focus on second language phonetics and phonology. She can be reached at: feeh0017@umn.edu.

Recommendation: This senior thesis was supervised and recommended for publication by **Dr. Timothy Hunter**, Linguistics Department, University of Minnesota, Twin Cities to the University Honors Programs in partial fulfillment of the requirements for the degree of Bachelor Arts, *summa cum laude* in Linguistics.

Works Cited

Best, C., & McRoberts, G. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, *46*.

Boersma, P. (2007). Praat (3.6). OT Learning (manual).

Boersma, P., Escudero, P., & Hayes, R. (n.d.). Learning Abstract Phonological from Auditory Phonetic Categories : An Integrated Model for the Acquisition of Language-Specific Sound Categories. *Proceedings of the 15th International Congress of Phonetic Sciences,* 1013-1016.

Boersma, P., & Hayes, B. (1999). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*.

Gnanadesikan, A. (1995). Markedness and faithfulness constraints in child phonology. *Ms., University of Massachusetts, Amherst*.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 11850–7. doi:10.1073/pnas.97.22.11850

Kuhl, P., Conboy, B., Padden, D., Nelson, T., & Pruitt, J. (2005). Early Speech Perception and Later Language Development: Implications for the "Critical Period." *Language Learning and Development*, *1*(3), 237–264. doi:10.1207/s15473341lld0103&4_2

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months, *2*.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(15), 9096–101. doi:10.1073/pnas.1532872100

Lai, C. S., Fisher, S. E., Hurst, J. a, Vargha-Khadem, F., & Monaco, a P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*, *413*(6855), 519–23. doi:10.1038/35097076

McCarthy, J. University of M. A. (2007). What is Optimality Theory?, From *Selected works of John J. McCarthy.* January.

Nowak, M., Komarova, N., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, 611–617.

Pinker, S. (1994). The language instinct. New York, New York: HarperCollins Publishers.

Querleu, Denis; Renard, Xavier; Versyp, Fabienne; Paris-Delrue, Laurence; Crepin, G. (1998). Fetal hearing. *European Journal of Obsterics & Gynecology and Reproductive Biology*.

Tesar, B., & Smolensky, P. (2000). Learnability in optimality theory. *MIT Press*, *00*.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America*, *120*(4), 2285. doi:10.1121/1.2338290

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.