EVOLUTION OF CLARITY

STEIJN KISTEMAKER
University of Amsterdam,
Plantage Muidergracht 24, 1018 TV Amsterdam,
skistema@science.uva.nl

NADYA PEEK
University of Amsterdam,
Plantage Muidergracht 24, 1018 TV Amsterdam,
npeek@science.uva.nl

For our research in the evolution of clarity we use an agent-based evolutionary model of the vowel system. This model uses genetic algorithms which are based on Darwin's ideas of natural selection. The individuals (agents) are selected on their clarity - or mutual understanding - which is tested with an imitation game. The individual vowel systems thus converge to a global common which allows the agents to best understand each other. The articulatory space that the agents use to produce the vowels also evolves to better accommodate clarity. We have found that during evolution the agents finally reach an optimal articulatory space which allows for maximal clarity without needing to increase in size.

1. Introduction

Language is both individual and global. Children begin to speak at an early age, but the utterances they make are only understood by a select group of individuals-those who speak the same language. If a child would begin to make sounds that are very distinct and easy to tell apart, then that may be considered as a more clear language, but no one in the community would understand the child. If the child was born into a community with a language consisting of sounds which are very similar, then communication might fail because the sounds are not easy to distinguish. The child cannot change the language of its community by himself, but perhaps the language will evolve itself to allow for more clarity in the community.

How do sets of phonemes evolve in the first place though? How did communities begin to distinguish sounds, and group them to form words? What is the starting point for a system of sounds which then makes up a language? In this paper we will research how a set of sounds can develop within a community of language speakers, and how the speaker may change the articulatory space they use to make the sounds.

In section 2 we will discuss an approximation of evolution using instead of a full set of phonemes, only the vowels. In section 5 we will describe the model we will use to analyze the evolution of vowel systems and the articulatory spaces of the agents, including some background on the techniques we will use: genetic algorithms and an imitation game. We will also highlight some of the most important parameter settings and algorithm descriptions we use in our particular implementation, and in section 6 we will detail some of the experiments we have run with our model. Finally we will detail our results and conclusions in sections 7, 8 and 9.

2. Vowel systems

The selection of phonemes that languages make is not a randomly distributed over the sounds that a human can produce. Some sounds have a far higher probability of occurring than others, and are represented in many more languages (Boer, 2000). The vowels within these collections of phonemes often represent a similar system, which has been hypothesized to be the case due to acoustic distinctiveness or articulatory ease- at first in (Jakobson, 1971), later expanded in (Liljencrants & Lindblom, 1972).

Vowels, unlike constants, are easily described as a combination of formants. Formants are certain acoustic resonances in the vocal tract which show up on spectrograms as dark bands. The first formant f_1 shows the acoustic resonance as produced by the opening of the vowel$-$ vowels like i and u have a relatively small f_1, whereas a has a larger opening and therefore a larger f_1. The second formant f_2 corresponds to the backness of a vowel$-$ i has a high f_2 frequency while a and u do not. The third formant can help distinguish the roundness of a vowel, but in general the first two formants are sufficient to determine the vowel sound (Rabiner & Schafer, 1978).

Because of the ease in recognizing and producing vowels, and because of their ubiquity in most languages, vowel systems are studied here as opposed to full phoneme systems in the evolution of language. We think that the study of vowel systems is a significant indication for the evolution of general speech.

3. Background

The similarity between vowel systems across languages along with the necessity to a maintain sufficient perceptual contrast is not intuitively coupled. Unsurprisingly, this has been a popular research topic within computational linguistics (Ke, Ogura, & Wang, 2003). The initial theory that maximal perceptual contrast is desired in communicative societies does did not mesh with the evolution of the vocal tract of man. While there has been evidence for the increase in size of the hypoglossal canal (Kay, Cartmill, & Balow, 1998) as well as the thoracic vertebral canal (Maclarnon & Hewitt, 2004) since prehistoric man, which indicates a more varied and controlled vocal range, the speed of this anatomical evolution does not indicate an driving need for more clarity in language through increase in vocal capabilities (Worden, 1995). Lindblom therefore hypothesized that a sufficient amount of clarity must be attainable by means of man's current articulatory capabilities (Lindblom, 1986).

Within the prescribed space available for vowel production in modern man, much research has been done to analyze how a distribution of vowels may come about. Nowak et al. have done this mathematically, calculating the theoretical optimum for discernibility given a static articulatory space (Nowak & Krakauer, 1999). De Boer has researched the distribution of vowels through self-organization within speaking community (Boer, 2000), giving an experimental simulation of vowel distribution. Ke, Wang and Ogura have attempted to optimize the distribution of vowels within an articulatory space with genetic algorithms (Ke et al., 2003), while also considering the focalization of the vowels as indicated in (Boë, Schwartz, & Vallée, 1994).

4. Hypothesis

Though there have been simulations of the optimal distribution of vowels in the literature like those described above, these simulations all consider a fixed size for the articulatory space of the agents. Whether the articulatory space should coevolve with the localization and focalization of the vowels has not been extensively simulated yet. In this paper we will therefore present a model for the simulation of the localization of vowels in variable articulatory spaces by a community of agents. Using this model and an evolution schema of genetic algorithms, we will experimentally determine whether the articulatory space of agents will coevolve with the localization of the vowels. We expect the articulatory space to grow as the agents evolve to more salient distributions of vowels, which require distance within the articulatory space.

5. Description of Model

Our model consists of two main parts; the communication cycle and the evolutionary cycle. In the communication cycle the agents aim to understand each other. If the communication is successful, the fitness of the communicating agents increases. Fit agents then pass successful traits on to their offspring in the evolution cycle. We will now go into more detail on both.

5.1. *Genetic Algorithms*

Genetic algorithms are a programming technique from the field of artificial intelligence based on Darwin's ideas of survival of the fittest and natural selection (Holland, 1992). In this technique individuals (in the population) with a higher fitness have higher chance of passing their genes to the next generation. Because of this, and in combination with mutations on the genes which create variations in the genes, continuously fitter agents are created (ideally) until an optimal solution is found and the evolution reaches an equilibrium.

In this technique an individual is represented by its genes. These genes (or genotype) define certain characteristics of the individual and will determine how the individual interacts with its environment. This outward, observable behaviour is called the phenotype, it is the basis for the selection procedure. The success of this outward behaviour is called the fitness. Individuals with a higher fitness have a higher chance of being selected for reproduction.

For our model we used a simple representation of the vowel space (which is depicted hierarchically in 1); two variables to define the dimension of the vowel space (width and height) which define the possible values for F_1 and F_2. For every vowel we defined a tuple (F_1 ; F_2) which defined its first to formants.

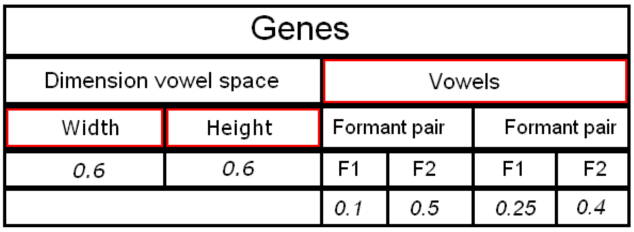

Figure 1. A hierarchical representation of a gene-set of an agent with 2 vowels

On these genes two operations can be defined; crossover and mutation. Crossover is the operation that combines the genes of two parents. Crossover takes place at the level of red-accented blocks. This means that when a child is created from two parents its width and height both have a 50% chance of coming from one of the parents. The vowels in the vowel space are inherited from one of the parents (again with equal probability) as a whole. So the vowels are not inherited separately. It is however possible that the vowels lie outside the new vowel space dimensions. In this case the vowels are bounded within new vowel space dimensions.

The second operation on these genes is mutation which causes variations on the genes. In our model every *value* in the genes (the lowest elements in 1) have a certain probability (P_m) of being mutated. Mutation is done by sampling from a Gaussian distribution with μ being its current value and $\sigma = \sigma_m$, where σ_m is a predetermined value.

The selection takes place via roulette selection where the individuals have a certain probability of being selected according to their fitness.

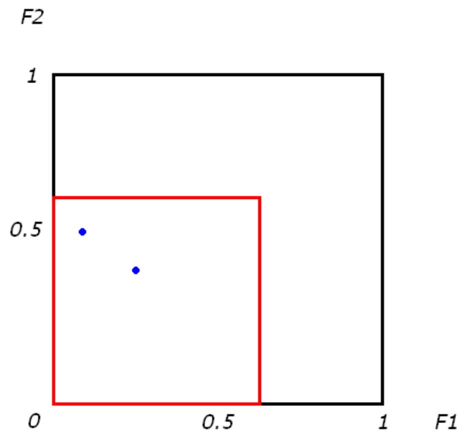

Figure 2. A graphical representations of the genes above. The blue dots are the 2 vowels within the red box vowel space.

5.2. *Imitation game*

To determine the fitness of the agents we use an imitation game as defined in (Boer, 2000). In this imitation game two agents try to communicate a vowel to each other. This happens by selecting one of the vowels in the agent's vowel space randomly. Its F_1 and F_2 values are then communicated with noise. Communicating with noise happens by sampling from a Gaussian distribution with μ being F_1 or F_2 and $\sigma = \sigma_c$, where σ_c is a predefined value called the communication error.

The second agent then finds the vowel in its own vowel space closest to the communicated value (using Euclidean distance). It then communicates this vowel back with noise. The first agent then also finds the vowel in its own vowel space closest to the one he 'heard' from the second agent. If this vowel is the same as the one communicated in the beginning communication was a success, otherwise communication was a failure. A graphical representation of a successful and failing communication can be seen in figure 3.

By playing these imitation games the fitness of an agent can be determined. The agents fitness is equal to the percentage of its successful communications.

6. Experiments

In our research we performed four types of experiments. These experiments were designed to gain insight in the development of vowel spaces by evolution. We also compare our experiments to the work of (Nowak & Krakauer, 1999), (Lindblom, 1986) and (Liljencrants & Lindblom, 1972).

In all experiments described below the agents started with their vowels ran-

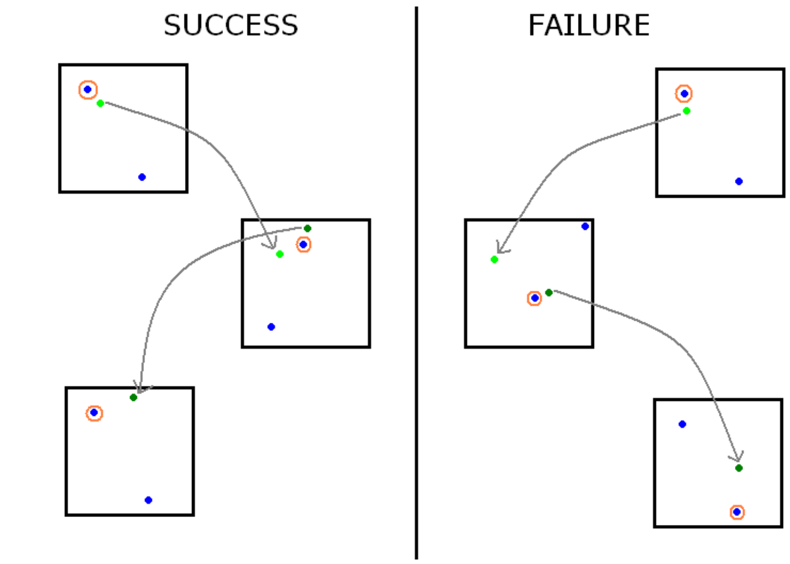

Figure 3. A graphical representation of the imitation game between two agents (per half: left squares represent agent 1, right square is agent 2). The vowels are in blue, the communicated vowels are in green. The circled vowels are the initial vowel in the first picture and the 'heard' vowels in the second and third picture.

domly placed within the (initial) vowel space. We also performed experiments where all agents had all their vowels start in the center. The results however where the same as with the random distribution, but took many generations to develop.

In the one-dimensional experiment we used a population of 50 agents, and in the two-dimensional experiments a population of 100 agents. We used such a small number of agents to make the distribution of vowels in the space clear in a visual sense while retaining enough variation in the population pool. The number of communications every agent made to determine its fitness was set to 500.

All σ-parameters of the experiments are in proportion to the size of the vowel space (which can be assumed to be 1 or 1×1 in the two-dimensional case). All two-dimensional spaces are simply squares and thus all results are on a theoretical basis, since a more realistic vowel space would better resemble a triangle shape.

6.1. *Static one-dimensional vowel space*

In our first set of experiments we investigated the evolution of a fixed-size one-dimensional vowel space with 2 to 8 vowels. We were interested in the distribution of the vowels on the vowel space (being here a line).

In (Nowak & Krakauer, 1999) a similar experiment is done using a mathematical model to calculate the optimal distribution of vowels on a line. They obtained a somewhat counter-intuitive result where the vowels aren't evenly distributed across the the line, but with $\#vowels > 4$ multiple vowels cluster at the borders. Although this would result in an global optimum there would be no way to distinguish between the clustered vowels, and thus would render the extra vowels useless. Because evolution is not guaranteed to result in the optimal solution we have no expectations that our model will find these optimal models.

For this experiment we used a fixed-size vowel space with a communication error $(\sigma_c) = 0.15$, a mutation deviation $(\sigma_m) = 0.01$ and a mutation probability $(P_m) = 0.1$.

6.2. *Static two-dimensional vowel space*

In our second set of experiments we investigated the evolution of a fixed-size two-dimensional vowel space with 2 to 5 vowels. We were interested in the distribution of the vowels in the two dimensional space. This experiment was designed to be a basis for the following experiments and to demonstrate that there is room for multiple solutions in this vowel space.

For this experiments we used as parameters: $\sigma_c = 0.25, \sigma_m = 0.01$ and $P_m = 0.2$.

6.3. *Variable two-dimensional vowel space*

In this third set of experiments we not only looked at the evolution of the vowels, but also of the space itself which was made variable to allow it to grow and shrink (and thus limit the space for the vowels to exist in). We expected that the size of the vowel space would grow as to give the vowels as much space to distance from each other.

We use the same settings as in the static two-dimensional experiments to be able to compare the results. We used 2 to 4 vowels in this experiment. The vowel space was initially set to 0.5, but could grow to 1 as a maximum and 0 as a minimum.

6.4. *Variable two-dimensional vowel space with a small communication error*

In the last set of experiments we investigated a theory in (Lindblom, 1986) which states that there is a limit to to where the vowels in a vowels system will drift apart, given a constant communication error. According to Lindblom, this is due to the fact that with a certain distance within the articulatory space, sufficient perceptual contrast is achieved to be able to communicate the vowels with little error.

This is easily intuitively explained by the fact that when there is only a small communication error the vowels do not have to be very different (distant) from each other. Because the probable error on communication lies within a certain range, vowels expanding over this range will not have a *practical* communicative advantage over other agents on this range and will thus not have a higher chance of passing their genes to the next generation.

We used the same settings as before, but now used a communication error $(\sigma_c) = 0.05$. We only used $\#vowels = 2$ in this experiment.

7. Results

After running the experiments detailed in section 6, we were able to produce some results which both confirm previous work from the literature such as (Lindblom, 1986) and (Boer, 2000) and show some insights in the evolution of the articulatory space in speaking agents.

7.1. *Static one-dimensional vowel space*

As stated in section 6 we performed the one-dimensional experiments because we were interested in the distribution of vowels in a one-dimensional vowel space. Nowak et al. already researched this and calculated optimal solutions given their mathematical model in (Nowak & Krakauer, 1999). They found that there was an upper limit of the number of vowels that could coexist and that when more than 4 vowels coexisted several vowels would take in the same place in the vowel space, and thus be indistinguishable.

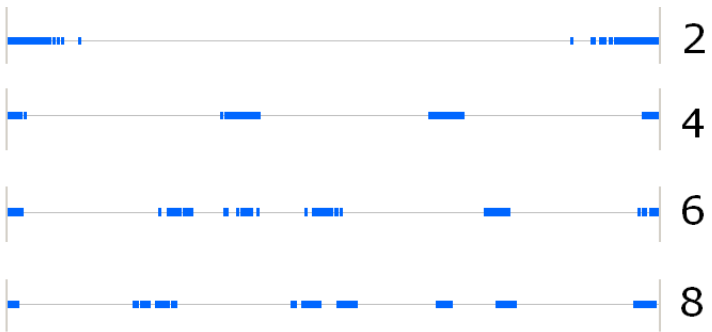

Figure 4. Distribution of vowels (in blue) in 1d vowel space for #vowels = 2,4,6,8 respectively. The vowels are represented by blue dots which can move laterally across the line which represents the vowel space.

As can be seen in figure 4 we have not obtained the same results as (Nowak & Krakauer, 1999), but find a more or less even distribution of vowels acros the vowel space. With a large number of vowels (> 6) the system takes longer to stabilize though, but the vowels still tend to be distanced from each other and not take in the same place. Since evolution is not always driven towards the optimal solution our model still holds and finds intuitive solutions to the distribution of vowels acros the vowel space.

7.2. *Static two-dimensional vowel space*

We performed our second experiment as a test to confirm our expectations about vowels in a 2-dimensional vowel space and as a comparison for further experiments.

Figure 5. Distribution of vowels of 100 agents (in blue) in a square vowel space of f_1 and f_2 with 2, 3, 4 and 5 vowels respectively.

As can be seen in figure 5 our expectations about vowels being pushed towards the corners and walls of the vowel space were confirmed. We can explain these results by considering the proces of communication. The more alike the vowels sound, the harder it will be for the speakers to distinguish between each of them. Speakers with more different sounding vowels, corresponding with more distance between the vowels in the vowel space, can thus be better understood.

We can also see in figure 5 that in the two-dimensional case evolution does not always lead to the same results and multiple optimal solutions can be found.

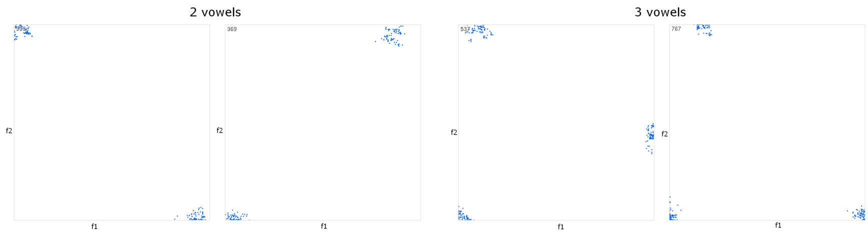

Figure 6. Two possible equilibria for a vowel space with 2 (left) and 3 (right) vowels. Again with 100 agents represented by their vowels located in 2 dimensional vowel space.

7.3. *Variable two-dimensional vowel space*

We performed our third experiment with a variable-sized vowel space to see if our agents would be pushed in evolution to obtain an increasing vowel space. As can be seen in figure 6 although the agents start with a very limited vowel space ($\frac{1}{4}$ of the maximum) the size of the vowel space increases over time and the number of generations. Eventually the boundaries of the vowel space were driven to the maximum. We confirmed these results in all test cases with 2, 3 and 4 vowels. Following the same analogy as before we can explain these results by understanding that for the agent to construct vowels that are maximally different he must have the capability to do so. Since the size of the vowel space limits his capability to produce maximally different vowel, there is an obvious function and evolutionary pressure to a large-sized vowel space.

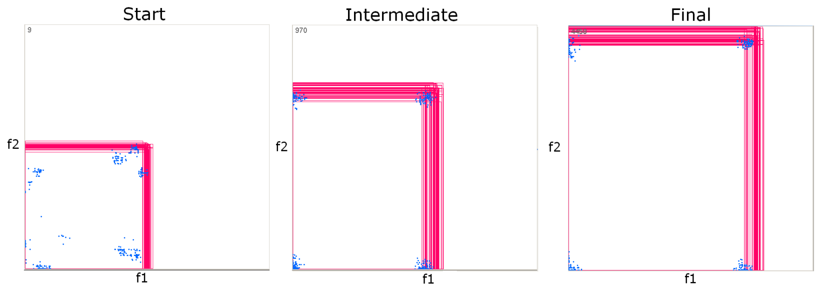

Figure 7. The evolution of a variable-sized vowel space with 4 vowels (size of vowel space in pink)

7.4. *Variable two-dimensional vowel space with a small amount of noise in communication*

Finally we were interested in the difference between sufficient perceptual contrast and maximum perceptual contrast as noted in (Liljencrants & Lindblom, 1972). In this experiment we let the agents communicate in a low-noise environment. As we can see in figure 7 we obtain very different results than before. After a large number of generations the agents still have not developed a larger vowel space. The agents' vowel spaces have even decreased.

Because there is less noise in the communication, the agents have less trouble distinguishing between different vowels. Therefore the distance between vowels can be smaller to obtain equally good communication as in the settings before. Although the noise is smaller to obtain a minimum chance at miscommunication the vowels still would have to be maximally different. We see however that the agents reach a sufficient perceptual contrast and not the maximal perceptual contrast. This can be explained by seeing that even though a larger vowel space would theoreticaly give an advantage in communication, when sufficient contrast is reached no practical advantage is reached anymore. Therefore there is no evolutionary pressure towards a larger vowel space. Because the vowel space does not have an influence on the fitness anymore (although it can not get too small) we see it growing and shrinking in equal proportions of the test runs. This is the effect of genetic drift.

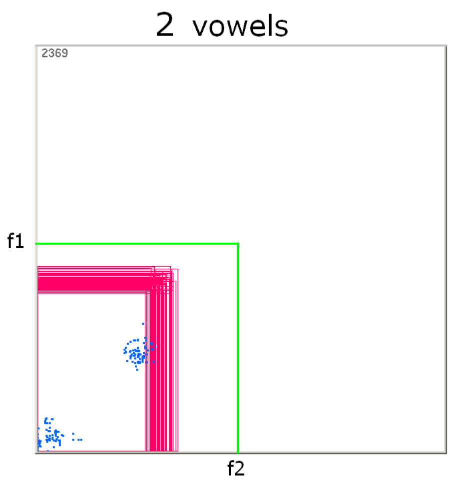

Figure 8. The evolution of a variable-sized vowel space with two vowels and low-noise communication (start size in green)

8. Conclusion

We tried to model the evolution of a simplified vowel space. We set up 4 experiments to research some theoretical properties of vowel spaces and compared our results to earlier work.

We found that our model placed the vowels evenly across a one-dimensional vowel space. Even though this doesn't correspond with the results (Nowak & Krakauer, 1999) it does not discredit our model since in evolution optimal solutions are not always found.

In a two-dimensional vowel space we have determined that vowels end up in the corners and at the walls of the vowel space. This corresponds with an optimal distance between the vowels in the vowel space. These results are easily explained in the light of noisy communication.

We also found that when the possibility for the vowel space to grow or shrink in size arises, it tends to grow. This corresponds with the speaker being able to make produce better distinguishable vowels and thus more successful communication.

There seems to be a limit on the increase of the size of vowel spaces. When sufficient perceptual contrast is reached the evolutionary pressure for a larger vowel space falls away and the vowel space stops growing (although it can continue growing due to genetic drift). We thus conclude that for better communication vowels have to be sufficiently different from each other, but not maximally per se.

9. Discussion and Future Work

Our model is quite simple and though it can be used to research basic properties of vowel spaces there is still a lot of room for improvement. Also some questions are still unanswered that could give us more insight.

Our model at the moment only works in a theoretical square vowel space, with no mapping to real vowel spaces. our model could be extended work within real F_1 and F_2 ranges that can be produced by the human sound system. Also we could improve our communication model to make it more realistic. We use a simplified linear perception scale, but perception of sounds and the differences between them are in reality on a logarithmic scale. Our model could be extended to incorporate this to make it more realistic.

At the moment our agents have a lower limit on their acoustic range (at (0,0)). This causes all agents to have an overlap in their communication possibilities. It would be interesting to see what happens when the left-lower bound of their acoustic range is also made variable. When agents have their vowels close together the overlap of vowel spaces between agents only have to consist of that specific part where the vowels lie, but when the vowels take up more room in the vowel space, the agents are most probably forced to take on more or less the same acoustic ranges.

Another interesting open would be to investigate the relation between the amount of noise and sufficient perceptual contrast. This could lead to a definition of a sufficiently sized vowel space in terms of the amount of noise.

Finally it would still be interesting to see what happens when a larger vowel space is penalized in the fitness function and how the interaction between this negative fitness and different amounts of noise in the communication would be.

References

Boë, L.-J., Schwartz, J.-L., & Vallée, N. (1994). The prediction of vowel systems: perceptual contrast and stability. *Fundamentals of speech synthesis and speech recognition: basic concepts, state-of-the-art and future challenges*, 185–213.

Boer, B. de. (2000). Emergence of vowel systems through self-organisation. *AI Communications*, *13*(1), 27-40.

Holland, J. H. (1992). *Adaptation in natural and artificial systems.* Cambridge, MA, USA: MIT Press.

Jakobson, R. (1971). *Studies on child language and aphasia.* The Hague: Mouton.

Kay, R. F., Cartmill, M., & Balow, M. (1998). The hypoglossal canal and the origin of human vocal behavior. *PNAS*, *95*(9), 5417-5419.

Ke, J., Ogura, M., & Wang, W. S.-Y. (2003). Optimization models of sound systems using genetic algorithms. *Computational Linguistics*, *29*(1), 1–18.

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*(4), 839-862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. Ohala & J. Jaeger (Eds.), *Experimental phonology* (p. 1773-1781). Orlando: Academic Press.

Maclarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evolutionary Anthropology*, *13*(5), 181–197.

Nowak, M. A., & Krakauer, D. (1999). The evolution of language. *PNAS*, *96*(14), 8028–8033.

Rabiner, L. R., & Schafer, R. W. (1978). *Digital processing of speech signals.* Prentice-Hall.

Worden, R. (1995). A speed limit for evolution. *Journal of Theoretical Biology*, *176*(1), 137–152.