Modelling of sound systems

Bart de Boer
Artificial Intelligence Lab – Vrije Universiteit Brussel
Pleinlaan 2, B-1050 Brussels, Belgium
bartb@arti.vub.ac.be

Abstract

This paper is an introduction to the computer modelling of systems of speech sounds. It focuses mainly on modelling that studies (the prerequisites of) evolution, so pure speech recognition is not treated. However, both cultural and biological evolution are considered as well as aspects of learning and social interactions that might be relevant to the evolution of speech. It will be argued that speech sounds are a relatively well-studied and simple-to-model aspect of language that nevertheless has interesting aspects in common with syntax. Compositionality is probably among the most exciting of these.

1 Introduction

Speech is an aspect of language that has been well investigated in linguistics. Universals, historical change, perception, production, processing and acquisition have all been researched in great detail. However, this detailed knowledge is not reflected in the amount of computer modelling effort that has gone into understanding the evolution of linguistic sound systems. Although one of the first computer models investigating factors in language origins was about sound systems (Liljencrants & Lindblom, 1972), the more recent surge of interest in computational models of language origins (see Kirby, 2002 for an overview) has resulted in relatively few papers on modelling of sound systems. de Boer (2002) presents an overview of recent work on computer modelling of evolution of speech sounds, but this body of work is relatively small compared to work on syntax and semantics.

A lot of information is available about sound systems of human language. This is partly because sounds are concrete signals that can be recorded and measured objectively. In contrast, more abstract aspects of language such as syntax and semantics, can only be investigated in a much more indirect way. Results of simulations involving sound systems can therefore be compared much more readily with observations of real human language. It is also easier to build models of perception, production and processing that are close to how humans handle speech sounds, although, of course, there is always the issue of computational complexity. These properties make speech sounds an attractive area of computer modelling research.

It could be suggested that speech sounds are less interesting than syntax and semantics, and effectively, systems of speech sounds show less complexity than syntactic and semantic systems. One can also argue that syntax is the fundamental property of human language, and that examples of call systems similar to human speech can be found in songs of songbirds and whales. Nevertheless, speech sounds do have a complexity of their own. Just as individual words can be combined into sentences according to the rules of syntax, speech sounds can be combined into words according to definite and sometimes reasonably complex rules. These rules have to be learned by infants in a way that is very similar to the way infants have to learn syntactical rules. On the basis of these and similar observations Carstairs-MacCarthy (1999) has argued that phonology is the evolutionary precursor of syntax. Jackendoff (2002, section 8.5) also observes that phonological systems must have been the first combinatorial systems in the evolution of human language.

Different aspects of evolution of sound systems have been modelled. These include the way in which linguistic diversity can emerge (e.g. Livingstone & Fyfe, 1999), the way in which sound systems can be learned and the role of mother-child interactions in this process (e. g. de Boer & Kuhl, 2001), the emergence of sound systems in populations (*e.g.* de Boer 2000), and the emergence of phonemic coding (e. g. Lindblom, MacNeilage, & Studdert-Kennedy, 1984; Steels & Oudeyer, 2000). Different approaches have also been applied to the problem of understanding origins and evolution of speech sounds. Models have been based on straightforward optimisation, genetic algorithms or agent-based models. Speech signals have been implemented as abstract symbols or as more realistic signals. Learning of speech sounds has been implemented with connectionist as well as more classical learning mechanisms, and some approaches (especially those using genetic algorithms) do not implement learning of speech at all.

This paper is meant as an introduction to speech sound modelling for an interdisciplinary audience. Therefore it contains material that might not be of interest to all readers. Section 2 contains a crash course in phonetics and phonology and can safely be skipped by linguists. Only the sections on universals of speech sounds (2.1) might be of interest to them. Section 3 contains an overview of work on modelling of speech sounds, and is hopefully of interest to all audiences. Section 4 in turn contains some considerations of modelling and can safely be skipped by dyed-in-the-wool cognitive modellers. Casual computer scientists, on the other hand are encouraged to read this section, though. The discussion at the end of the paper contains some suggestions of future work and is, hopefully, again of interest to all audiences.

2 Speech sound crash course

All human languages, except those used by deaf communities, use sound as their means to transmit signals. As human language must be able to convey a large and potentially infinite number of different messages efficiently, it needs a large and extensible repertoire of signals. The solution human language uses for this is to recombine a small set of basic sounds into larger assemblies. I will discuss properties of human speech in a slightly simplified fashion below. It should be kept in mind, however, that every aspect of human language is complex and idiosyncratic, and that for each of these aspects, including speech, many competing theories abound. The material I present reflects my understanding of the subject and also reflects what I think is necessary to successfully make computer models for studying (evolution of) human speech. Nonetheless, the reader is encouraged to study the linguistic and phonological theory independently. Any good textbook on the subject of phonology of phonetics that is not too much centred on one particular theory will do, although Ladefoged and Maddieson's book (Ladefoged & Maddieson, 1996) is especially recommended for those who like to marvel at the diversity and complexity of human languages.

Most linguists would say that phonemes are the basic unit of speech. Phonemes are defined as the smallest articulations that can change the meaning of a word. An example for English would be /l/ and /r/, as changing /l/ into /r/ can change the meaning of a word: "law" and "raw" have different meanings in English. Two words that differ only in one phoneme are called minimal pairs. By carefully studying minimal pairs in a language, linguists can determine the set of phonemes a language uses. However, there are complications. Phonemes are not always pronounced identically in all contexts. In standard English, /l/ at the beginning of a word is pronounced quite differently from /l/ at the end of a word. When /l/ occurs at the end of a word, it is velarised, at the beginning of a word it is not. However, not velarising an /l/ at the end of the word does not change its meaning, it only makes the speaker sound funny, or possibly as if he or she has an Irish accent. There are languages that do use this distinction. In Russian, for example, changing a palatalised /l/ (which sounds approximately like English /l/ at the beginning of a word) into a velarised /l/ changes the meaning of a word. For instance: мель "shoal", ending in palatalised /l/ contrasts with мел "chalk" ending in velarised /l/. Sounds that are phonemes in one language do not have to be phonemes in another, and they are therefore language-dependent.

There are other levels at which speech can be considered combinatorial. Phonemes are problematic in certain ways. First of all, their independent existence is not so clear as it might seem at first glance. Independently pronounceable phonemes, such as vowels, nasals (/m/, /n/) and fricatives (/v/, /s/) clearly do have an independent existence. But what about sounds such as /p/ and /t/? These can only be pronounced clearly in the context of a vowel. And indeed, it turns out that many speech sounds are detected mostly by the way they influence neighbouring sounds. Even independently pronounceable sounds are heavily influenced by the surrounding sounds. This effect is called co-articulation. This causes the interesting phenomenon that a speech signal cannot usually be cut up into its constituent phonemes and be recombined into another understandable utterance. Although phonemes definitely do have some status as units of recombination, it might be necessary to look at the syllable level as well.

One encounters many subtle and complex problems when studying syllables, but a simplified account would consider syllables to consist of an onset, a nucleus and a coda. Typically, the nucleus consists of the vowel part of a syllable[1]. The onset consists of the consonants preceding the nucleus and the coda consists of the consonants following the nucleus. Thus in English "sprint", /i/ is the nuclues, /spr/ is the onset and /nt/ is the coda. Syllables can also contain what is called suprasegmental material (modifications of the signal spread over the whole syllable), such as tone. It is usually possible to cut up a speech signal into syllables and recombine these, such that an intelligible result ensues. Therefore, syllables are uncontroversial as a unit of recombination. However, the number of syllables per language can become quite large, and these syllables can often easily be analysed in terms of smaller units – either phonemes or onsets, nuclei and codas. It is therefore likely that both the syllabic level and the phonemic level have cognitive significance.

A third level of analysis of speech is also generally assumed. This is the level below that of phonemes. It turns out that the phoneme inventories of languages are systematic. For example, languages that use sounds like /b/, /d/ and /g/ tend to use /p/, /t/ and /k/ as well. These sounds can be paired such that the elements of each pair only differ in voicing. The sound /b/, for example is articulated with the lips and is voiced (the vocal chords need to vibrate) while the sound /p/ is articulated in the same manner, but it is voiceless. These sounds are said to differ only in the *feature* 'voicing'. Speech sounds can thus be analysed as consisting of several distinctive features. Apart from voicing, distinctive features can represent place of articulation, manner of articulation or nasalisation, for example. Almost always distinctive features are considered to be binary. Different researchers of speech sounds often use different sets of features. Many processes that have to do with how sounds are combined into words and how words are simplified in rapid, casual speech can be explained in terms of distinctive features. Although the exact nature and cognitive reality of distinctive features is hotly debated, there is no doubt that they are a useful tool for linguistic description, and speech can definitely be analysed as combinatorial on the level of distinctive features.

2.1 Universals

Systems of speech sounds of human languages are not random collections of sounds. If this were the case, there would be very little to study or model. Although humans can produce and distinguish an amazing number of different sounds (Ladefoged & Maddieson, 1996), the phoneme inventories of human languages show remarkable regularities. Most languages from a reasonably representative

---

[1] There are many languages that allow consonants as the nucleus, for example, Czech "krk" (neck, throat). Deciding what exactly the syllable nuclei are in such languages can be tricky.

sample of 451 languages (Maddieson, 1984; Maddieson & Precoda, 1990) have between 20 and 37 phonemes, while the most frequently occurring number of phonemes is 25. The minimum number of phonemes appears to be 11 for the East-Papuan language Rotokas (Firchow & Firchow, 1969) and the South-American language (Murà-) Pirahã (Everett, 1982; Sheldon, 1974) spoken by men and 10 for Piraha spoken by women (Daniel Everett, personal communication). The maximum number of phonemes for a language in UPSID is 141 for the Khoisan language !Xũ (Snyman, 1970) while over 160 phonemes have been reported for the Khoisan language !Xóõ (Traill, 1985). Because of the way clicks are analysed in these languages, these numbers might be slightly inflated, and the real maximal number of phonemes per language might be closer to 90, for example for certain North-East Caucasian languages (e. g. Catford, 1977).

Not only are there regularities in the repertoire sizes, but there are also regularities in the kinds of sounds that occur. Certain sounds occur much more often than others, and certain speech sounds also co-occur more frequently than predicted from the frequencies of the individual sounds themselves. For regularities of phoneme inventories see e.g. (Schwartz, Boë, Vallée, & Abry, 1997b) while for more general regularities see (Maddieson, 1984). Also, languages with small inventories tend to use a limited number of basic articulations, while languages with larger inventories tend to use more elaborated articulations that can often be analysed as combinations of more basic articulations (Lindblom & Maddieson, 1988). It is not always possible to distinguish between sequences of phonemes on the one hand and complex phonemes on the other hand. This is especially true for the complex clicks found in some of the Khoisan languages (as mentioned above) that are analysed as single phonemes, but that could in principle also be analysed as sequences of basic clicks and secondary articulations (Traill, 1985) (incidentally reducing the number of phonemes in these languages substantially).

The number of syllables in languages varies much more than the number of phonemes. Maddieson (1984, section 1.9) presents the numbers of possible syllables for nine languages. The numbers of theoretically possible syllables ranges from 162 for Hawaiian to 23 638 for Thai (taking into account tones as well). However, the way in which speech sounds can be combined into larger wholes (syllables, words) does show regularities. Certain speech sounds tend to occur close to the nucleus of a syllable, while other sounds tend to occur near the periphery. Speech sounds can be ordered hierarchically, such that sounds higher on the hierarchy tend to occur closer to the nucleus of a syllable when co-occurring with sounds occurring lower on the hierarchy. This is often called the sonority hierarchy. For a description, see (Vennemann, 1988). It is approximately as follows a > w > l > n > s = t (where each phoneme represents the sounds from its category). Many languages allow only very simple syllables, such as syllables consisting of a vowel only, or syllable consisting of a single consonant followed by a single vowel. Such constraints imply that phonemes cannot be combined freely. These constraints are probably due to articulatory and acoustic factors, such that speech that follows them is easier to produce and perceive.

Tone systems, the ways in which languages use pitch and pitch contours to make distinctions between words, also has its universals. These have to do with the size of the inventories, with the frequencies with which tones occur (falling tones appear to be preferred over rising tones) and the way in which tones combine into tone inventories. Unfortunately, I am quite ignorant on the subject of universals of tone, but the interested reader is referred to (Maddieson, 1978).

These regularities observed in human languages and which are usually called *phonological universals* are the factual basis that successful computational models should explain. If a model gives results that are significantly different from what is observed in real human languages, it will not be interesting for linguists, even though the model might have other merits. Also, it is important for researchers coming from a different domain to have a good knowledge of the linguistic facts, in order to gain the necessary scientific street credibility to have ones ideas accepted in the linguistic community.

2.2 Other sources

Linguistic universals are probably the most interesting source of data for researchers interested in modelling language evolution. Results of experiments can be "grounded" in this kind of data. However, when one wants to build more realistic agents, or if one is interested in modelling individual's language behaviour, other areas of linguistics need to be taken into account. General linguistics is of course the area in which most formal models of language are generated, and Jackendoff (2002) provides a good starting place. Jackendoff has interesting things to say about computer modelling as well as language evolution. However, for use in computer modelling, models of general linguistics might be too general, and therefore difficult if not impossible to implement.

For factual data about language learning and language performance, studies of infant development and psycholinguistics might provide interesting material. Infants learn to speak amazingly quickly and according to a rather fixed pattern. Also the parent-child interactions show cross-cultural similarities (i. e. Ferguson, 1964; Fernald et al., 1989). Good books that provide an introduction to the field of infant phonological development are Vihman (1996) and Jusczyk (1997). These provide detailed background information on the stages infants go through when learning how to speak, while being relatively theory-free. As for psycholinguistics, an easy-to-understand introduction to the field is provided by Field (2003) but admittedly, this is the only book I read on the subject, so there might be better ones around. A background in these subjects is relatively easy to acquire and will certainly increase the quality of the modelling work, as well as the ability of the modeller to convince an audience of linguists.

3 Overview of speech modelling[2]

Probably the first attempt at making a computer model to explain universals of speech sounds was made by Liljencrants & Lindblom (1972). This model optimized randomly initialized vowel systems with a fixed number of vowels. The optimization was based on a function that modelled the potential energy of repelling magnets (this potential energy is higher whenever the magnets are closer together). By shifting the individual vowels in the system, the energy function was minimized. Liljencrants and Lindblom found that vowel systems that were optimized in this way were remarkably similar to vowel systems found in human languages, although there were some discrepancies. Later re-implementations that used modified distance functions (e. g. Schwartz, Boë, Vallée, & Abry, 1997a; Vallée, 1994) have succeeded in making progressively better approximations of human vowel systems.

Subsequently, Lindblom et al. (1984) have tried to use an optimising model for explaining phonemic (that is combinatorial) coding of syllables. The syllables consisted of a simple consonant followed by a vowel. Although the systems that emerged were phonemically coded, their model has not had the success of the model for vowels, because there are many more parameters in it and because it is much more difficult to replicate the results.

Only in the mid-nineties did work on explaining sound systems with computer models get a new impulse with systems that were based on populations of sound systems and populations of agents. The first to make an agent-based implementation to investigate the emergence of vowel systems was Glotin (Berrah, Glotin, Laboissière, Bessière, & Boë, 1996; Glotin, 1995) of the Institut de Communication Parlée (ICP) in Grenoble, the same institute were Schwartz et al. (1997a) do their research. He made a model in which a population of talking agents tries to develop a shared repertoire of (a fixed number of) vowels. His agents have an acoustic as well as an articulatory representation of the

[2] This is a slightly reworked and updated version of the material presented in the history of modelling section of (de Boer, 2002).

vowels, and adapt their vowel systems on the basis of their interactions. The agents are also subject to a genetic algorithm, which is (according to Glotin, personal communication) not meant to be a model of actual biological evolution of the agents, but rather of the way sound systems are transferred from parents to children. This might be considered a weak point of the research, as the influence of the genetic algorithm and the interactions between the agents are difficult to separate. Another problem with the model was that it was computationally too complex, and that therefore only few simulations with small populations and small numbers of vowels could be run. In a way, this work was ahead of the computing power of the time.

It has been at the basis of a number of subsequent research efforts, however. In the first place those of Berrah (Berrah, 1998; Berrah & Laboissière, 1999) and myself (de Boer, 1997, 2000, 2001). Berrah's work was a direct continuation of Glotin's research. Berrah's model is a simplification of Glotin's model. The agents do no longer have an articulatory representation of the sounds they use, only an acoustic one. This reduces the computational load considerably and allows more experiments with larger populations and larger numbers of vowels to be run. Berrah extends Glotin's model by investigating what he calls the "Maximum Use of Available Features". By allowing the agents to use an extra feature (which could be length, nasalization etc. in human languages, but which he models as an extra abstract dimension of the acoustic space) he shows that this is only used whenever the number of vowels in the agents' repertoires exceeds a certain threshold. His simulations also contain a genetic component, which makes it sometimes hard to tell when a particular phenomenon is due to interactions between the agents and when it is due to the actions of the genetic algorithm.

My own work has concentrated on predicting vowel systems from interactions in a population. The agents have both an articulatory as well as an acoustic representation of their vowels, but use a much simpler articulatory model than the one used by Glotin. Also, the agents do not evolve, although experiments have been done with changing populations (de Boer & Vogt, 1999). They interact through language games (in this experiment called imitation games) only. It has been shown that vowel systems of human languages, and the relative frequencies with which they occur can be predicted quite accurately with this model.

Daniel Livingstone and Colin Fyfe of the university of Paisley have investigated the origins of linguistic diversity (Livingstone & Fyfe, 1999). They model a population of agents that has a spatial structure and monitor how linguistic diversity changes over time. The research question they are addressing is how it is possible that there are many different languages, and under what conditions such diversity can arise. Their work builds on work performed by Daniel Nettle on the emergence of linguistic diversity (Nettle, 1999).

More recently research has started to investigate syllable systems with genetic algorithms and population models. This work relates in a similar way to the optimizing simulation used by Lindblom et al. (1984) as Glotin's, Berrah's and my own work relates to Liljencrants' and Lindblom's (1972) model. Redford and colleagues of the university of Texas, Austin (Redford, Chen, & Miikkulainen, 2001) have made a model that is based on a genetic algorithm. The population consists of words, which in turn consist of a closed set of phonemes. Redford et al. use a number of rules that determine how hard it is to produce and perceive different combinations and sequences of phonemes. On the basis of this a fitness for all the words in the population is calculated and selection and recombination take place. They try out different combinations of rules and investigate which rules are most important to predict syllables that are like those found in human languages.

Other work on predicting properties of more complex utterances is being conducted and published at the moment. Pierre-yves Oudeyer of the Sony computer science laboratory in Paris, France is working on predicting repertoires of syllables using more realistic signals (e. g. Oudeyer, 2002; Steels & Oudeyer, 2000). Professor William Wang of the electronic engineering department of the City University of Hong Kong and co-workers Jinyun Ke and Mieko Ogura are working on modelling tone systems within the framework of genetic algorithms (Ke, Ogura, & Wang, 2003).

My own most recent work, in cooperation with Patricia Kuhl of the University of Washington is in investigating the role of mother-child interactions in the transfer and evolution of language. We have investigated with a computer model how infant-directed speech can facilitate learning (de Boer, to appear; de Boer & Kuhl, 2001) and how it can facilitate transfer of vowel systems from generation to generation (de Boer & Kuhl, to appear).

4 Modelling considerations

Making successful computer models of the evolution of speech is somewhat of an art. One needs to find the right balance between performance on the one hand and realism on the other. In order to do simulations that are relevant to linguists, it is necessary to make models that work with something that is close to real speech. Such models can become computationally intractable, though. It is often better to sacrifice excessive realism in order to have better performance. Thus one can do multiple repeated experiments, work with larger populations, or work with larger repertoires of sounds. It is easy to get carried away (especially for phoneticians trying to build computational models of articulation) and to want to use the most accurate models of speech production and perception available. This is not always the best strategy, in my opinion, as the bottleneck determining the realism of the model is usually not in the articulatory or perceptual model, but in the implementation of the more cognitive aspects of the model. These include, but are not limited to, the way the system recognises speech sounds, the way in which it learns them, or the way it coordinates movements of the vocal tract.

Simplifying too much is not a good idea, either. Although interesting work can be done by highly abstract models, it is always easier to convince linguists if one stays close to real speech. If one does choose to work with abstract signals, it is crucial to point out how these map to real speech and how the results need to be translated for and interpreted by linguists. Of course, showing results that are directly understandable to linguists increases the probability that one's work is accepted in their community enormously.

One should try to minimise the number of parameters in the system. When there are many parameters, the impression could arise that the model's results are caused by tuning, rather than that it captures something essential. This is probably the reason why Liljencrants and Lindblom's (1972) model on vowels was much more successful than Lindbom et al.'s (1984) later model on syllables. In any case one should try to derive parameter values from independent (physical, physiological, perceptual, articulatory or cognitive) considerations, and show that the performance of the model does not depend sensitively on the exact parameter values.

Modellers of speech can derive their inspiration from models used by engineers working in speech recognition and speech synthesis, but should be aware that the aims of the engineering approach are totally different from the aim of computer modelling of speech. Engineers are satisfied when their models achieve better recognition rates, or more realistic speech production, no matter how cognitively implausible their models are. When modelling speech from a cognitive perspective, one should always keep in mind that in the end one is trying to understand the workings of the human brain. This is not to say that one should always use models that are directly based on the architecture of the brain, such as neural networks. Often it is necessary to use higher level models in order to achieve more complex behaviour, or in order to keep the model's behaviour transparent. Nonetheless, it is always a good idea to remember that one is trying to understand cognition and not trying to build more and more fancy computer models.

There are a few other, more general things one has to keep in mind when doing computer modelling of speech. First of all, it is important that the statistical significance of the results be established. Although this may appear to be a bit of a superfluous statement, it is amazing how many computational modelling papers are written (and accepted for publication) without a statistical analysis of the results. The author pleads guilty in this respect, too. However, the fact that results are accepted on good faith in the computational modelling community does not mean that they are accepted on this basis in other communities. Secondly, as has been mentioned above, it is a good idea to do a sensitivity study of one's model. This means that all parameters are varied and that it is reported how these changes do (or do not) influence the behaviour of the model. And finally, one should always be careful about bias in the model. Many modelling studies depend on random initialisations and random changes. Sometimes it is not straightforward to make sure that such initialisations and changes are unbiased. It is therefore always a good idea to test this before enthusiastically reporting results that are caused by systematic error instead of the model's dynamics.

5 Discussion

Speech is an interesting part of language that is well-known and easy to model, but that nevertheless has the potential of providing insights on language evolution that might have repercussions on the understanding of syntax as well. Speech sounds form sequences that are combinatorial and in this respect they resemble syntax. By some researchers (e. g. Carstairs-McCarthy, 1999; Jackendoff, 2002) speech is even considered a possible precursor to syntax. Also, learning of sequences is something that is necessary for both speech and syntax. Apart from this, many interesting and linguistically relevant results have already been achieved within the domain of the study of speech proper.

A lot of problems are still open, and many have to do with extending the models to more complex speech sounds and sequences of speech sounds. Working with more complex signals involves more complex articulatory and perceptual models, as well as learning time sequences. These are problems that have only been very partially solved in the research on computer speech production and processing. However, it is not necessary to wait for progress in these fields before interesting new experiments can be done. First of all, a lot of the more sophisticated production- and processing models have not yet been used in the modelling of the evolution of language. It is also possible to do many interesting experiments into the evolution of speech sounds with simpler abstract models. The transition from holistic to phonemically coded signals, for example, is still poorly understood. Also, the role erosion of phonetic forms of words and morphemes plays in grammaticalisation could be investigated with more abstract signals. At the moment, models of emergence of grammar generally assume that language consists of strings of discrete symbols and that words are separated by silence. In real speech this is clearly not the case, and it would be interesting to investigate the combination of a grammar-learning model and a realistic phonetic component. Finally, realistic sound change has not been modelled successfully, as far as the author is aware. A model that could produce realistic chain shifts and context-dependent sound changes would be extremely interesting. To achieve this, a component of meaning probably needs to be integrated. At the moment, most models investigating sound systems only have a very rudimentary implementation of meanings of utterances.

Apart from these examples, many other experiments can be conceived that would shed light on evolution of speech sounds as well as problems that exceed the domain of speech sounds alone. Therefore it is not just interesting to continue research into speech sounds, but it is also necessary to open the dialogue between modellers of speech sounds and modellers of other aspects of language.

References

Berrah, A.-R. (1998). *Évolution Artificielle d'une Société d'Agents de Parole: Un Modèle pour l'Émergence du Code Phonétique*. Grenoble: Thèse de l'Institut National Polytechnique de Grenoble, Spécialité Sciences Cognitives.

Berrah, A.-R., Glotin, H., Laboissière, R., Bessière, P., & Boë, L.-J. (1996). From Form to Formation of Phonetic Structures: An evolutionary computing perspective. In T. Fogarty & G. Venturini (Eds.), *ICML '96 workshop on Evolutionary Computing and Machine Learning, Bari* (pp. 23–29).

Berrah, A.-R., & Laboissière, R. (1999). SPECIES : An evolutionary model for the emergence of phonetic structures in an artificial society of speech agents,. In D. Floreano, J.-D. Nicoud & F. Mondada (Eds.), *Advances in Artificial Life, Lecture Notes in Artificial Intelligence* (Vol. 1674, pp. 674–678). Berlin: Springer, 1999.

Carstairs-McCarthy, A. (1999). *The origins of complex language: an inquiry in the evolutionary beginnings of sentences, syllables, and truth*. Oxford: Oxford University Press.

Catford, J. C. (1977). Mountain of tongues: the languages of the Caucasus. *Annual Review of Anthropology, 6*, 283-314.

de Boer, B. (1997). Generating vowel systems in a population of agents. In P. Husbands & I. Harvey (Eds.), *Fourth European Conference on Arti.cial Life.* (pp. 503-510). Cambridge (MA): MIT Press.

de Boer, B. (2000). Self organization in vowel systems. *Journal of Phonetics, 28*(4), 441-465.

de Boer, B. (2001). *The origins of vowel systems*. Oxford: Oxford University Press.

de Boer, B. (2002). Evolving Sound Systems. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language* (pp. 79-97). Berlin: Springer Verlag.

de Boer, B. (to appear). Infant directed speech and the evolution of language. In M. Tallermann (Ed.), *Proceedings of the evolution of language conference (working title)*. Oxford: Oxford University Press.

de Boer, B., & Kuhl, P. (2001). Infant-directed vowels are easier to learn for a computer model. *Journal of the Acoustical Society of America, 110*(5, pt. 2), 2703.

de Boer, B., & Kuhl, P. (to appear). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*.

de Boer, B., & Vogt, P. (1999). Emergence of Speech Sounds in Changing Populations. In D. Floreano, J.-D. Nicoud & F. Mondada (Eds.), *Advances in Artifical Life, Lecture Notes in Artificial Intelligence 1674* (pp. 664-673). Berlin: Springer Verlag.

Everett, D. L. (1982). Phonetic rarities in Piraha. *Journal of the International Phonetic Association, 12*(2), 94-96.

Ferguson, C. A. (1964). Baby talk in six languages. *American Anthropologist, 66*(6, part 2), 103-114.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*, 477-501.

Field, J. (2003). *Psycholinguistics: A resource book for students*. London: Routledge.

Firchow, I., & Firchow, J. (1969). An abbreviated phoneme inventory. *Anthropological Linguistics, 11*, 271-276.

Glotin, H. (1995). *La Vie Artificielle d'une société de robots parlants: émergence et changement du code phonétique*. Grenoble: DEA sciences cognitives-Institut National Polytechnique de Grenoble.

Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.

Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Cambridge, MA: The MIT Press.

Ke, J., Ogura, M., & Wang, W. S.-Y. (2003). Optimization Models of Sound Systems Using Genetic Algorithms. *Computational Linguistics, 29*(1), 1-18.

Kirby, S. (2002). Natural language from artificial life. *Artificial Life, 8*(2), 185-215.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Liljencrants, L., & Lindblom, B. (1972). Numerical simulatons of vowel quality systems. *Language, 48*, 839-862.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of language universals. In M. Butterworth, B. Comrie & Ö. Dahl (Eds.), *Explanations for language universals* (pp. 181-203). Berlin: Walter de Gruyter & Co.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman & C. N. Li (Eds.), *Language, speech and mind* (pp. 62-78). London: Routledge.

Livingstone, D., & Fyfe, C. (1999). Modelling the Evolution of Linguistic Diversity,. In D. Floreano, J.-D. Nicoud & F. Mondada (Eds.), *Advances in Artificial Life, Lecture Notes in Artificial Intelligence* (Vol. Volume 1674, pp. 704-708). Berlin: Springer.

Maddieson, I. (1978). Universals of Tone,. In J. H. Greenberg, C. A. Ferguson & E. A. Moravcsik (Eds.), *Universals of Human Language* (Vol. Volume 2 Phonology, pp. 335-365). Stanford: Stanford University Press.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, I., & Precoda, K. (1990). Updating UPSID. *UCLA Working Papers in Phonetics, 74*, 104-111.

Nettle, D. (1999). *Linguistic Diversity*. Oxford: Oxford University Press.

Oudeyer, P.-y. (2002). Phonemic coding might be a result of sensory-motor coupling dynamics. In J. Hallam (Ed.), *Proceedings of the International conference on the simulation of adaptive behavior (SAB)* (pp. 406-416). Edinburgh: MIT Press.

Redford, M. A., Chen, C. C., & Miikkulainen, R. (2001). Constrained emergence of universals and variation in syllable systems. *Language and Speech, 44*, 27–56.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997a). The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics, 25*, 255-286.

Schwartz, J.-L., Boë, L.-J., Vallée, N., & Abry, C. (1997b). Major trends in vowel system inventories. *Journal of Phonetics, 25*, 233-235.

Sheldon, S. N. (1974). Some morphophonemic and tone rules in Mura-Pirahã. *International Journal of American Linguistics, 40*, 279-282.

Snyman, J. W. (1970). *An introduction to the !Xu (!Kung) language*. Cape Town: Balkema.

Steels, L., & Oudeyer, P.-y. (2000). The cultural evolution of syntactic constraints in phonology. In M. A. Bedau, J. S. McCaskill, N. H. Packard & S. Rasmussen (Eds.), *Proceedings of the VIIth Artificial life conference (Alife 7)*. Cambridge (MA): MIT Press.

Traill, A. (1985). *Phonetic and phonological studies of !Xóõ bushman*. Hamburg: Helmut Buske Verlag.

Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux prédictions* (Vol. T). Grenoble: hèse préparée au sein de l'Institut de la Communication Parlée (Grenoble-URA C.N.R.S. no 368).

Vennemann, T. (1988). *Preference Laws for Syllable Structure*. Berlin: Mouton de Gruyter.

Vihman, M. M. (1996). *Phonological Development: The Origins of Language in the Child*. Cambridge MA: Blackwell.