Int. J. Adv. Res. Sci. Technol. Volume 1, Issue 2, 2012, pp. 23-28.

IJARST

International Journal of Advanced Research in Science and Technology

journal homepage: www.ijarst.com

ISSN 2319 – 1783 (Print)
ISSN 2320 – 1126 (Online)

Recognition of Vowels from Spoken Phonemes

Pratyush Ranjan Mohapatra and Satyajit Nayak*

Gandhi Institute For Technology, Bhubaneswar, Odisha, India.

****Corresponding Author's Email: pratyush.mohapatra@gmail.com***

ARTICLE INFO

Article history:

Received	20 Dec. 2012
Accepted	29 Dec. 2012
Available online	31 Dec. 2012

Keywords:
Human Sound Systems ,
Regularities of systems of speech sounds,
Models.

ABSTRACT

The research paper is rooted in *artificial intelligence* and *phonetics*. Within artificial intelligence, the most relevant subfield is the one that tries to model the origins of intelligence. Within phonetics it is the subfield that is interested in explaining the structure of sounds that are found in human languages. The methodology of the research— using agent-based simulations for modeling aspects of human intelligence— is that of artificial intelligence. The data on which the simulations were based and the data that were used for verifying the results were taken from the field of phonetics. The research questions were taken from both fields. The most important phonetic work on which the thesis is based is that on the functional explanation of the regularities that are found in the vowel systems of the world's languages. Phoneticians have also used computer models (see e.g. Liljencrants & Lindblom 1972, Schwartz *et al.* 1997b) for this purpose. However, all these models were based on direct optimization of functional criteria. This is unfortunate, because humans do not optimize their sound systems. Nevertheless, the models based on optimization predict the most frequently occurring human vowel systems very well. The hypothesis that is investigated in this thesis is that the optimization is the result of self-organization in the interactions between the language users.

Introduction:

The paper is most important pillar in the field of artificial intelligence is the work by Steels (1995, 1996, 1997a, 1997b, 1997c, 1998a, 1998b) on the origins of language. His work views language as a complex dynamic, open and distributed system. The term *complex dynamic* indicates that the dynamics of language— the way it changes, the way its speakers interact and the way it works— are complex and cannot be predicted by simple rules. Language is an *open system* with respect to both its community of speakers and with respect to what it can express. The population of speakers as well as what a language can express (its words, its constructions) can change without disrupting it. Finally language is a *distributed system* in Steels' view, because none of the speakers has perfect knowledge of the language nor does any of the speakers have central control over the language. Language is to a large extent independent of its community of speakers. According to Steels, coherence is maintained through self-organization, while changes of the language are caused by cultural evolution. Steels also claim that in a population of speakers that are sufficiently intelligent to learn a system as complex as language, a language will indeed spontaneously emerge. This emergence is driven by the same processes of cultural evolution and self-organization that drive language change.

Universal Tendencies of Human Sound Systems:

Most human languages use sound as their primary medium for conveying meaning. Only sign languages use vision. The stream of speech sounds is usually analyzed as consisting of a sequence of separate speech sounds that are called *phonemes*. Phonemes are defined as minimal speech sounds that can make a distinction in meaning. In English, for example, /e/ and /æ/ are phonemes, because the words /bet/ "bet" and / bæt/ "bat" have different meanings. In Dutch or French, for example, these words would be indistinguishable, so these languages are analyzed to have only have one phoneme /e/. In making a description of a language, one first has to make an inventory of which sound distinctions can make a distinction in meaning, i.e. which phonemes the language uses. Usually a rather unambiguous analysis of the set of phonemes of a language is possible.

However, there are some complications. The most important one is that it is not easy to separate the actual physical speech signal into phonemes. This is because the human articulators do not produce phonemes separately, but already start producing new phonemes when they are not yet completely finished producing the previous ones. This effect is called *co-articulation*. Co-articulation causes phonemes to be realized differently in different contexts. This is called allophonic variation. However, not all allophonic variation can be explained as the effect of co-articulation. For example, the fact that the phoneme /l/ in English is produced quite differently at the beginning of a word than at the end of a word cannot easily be explained by co-articulation effects. Rather this variation is something that must be learned by a speaker. This variation can assume rather extreme forms, especially in languages with small phoneme inventories. For example in the language Rotokas, with an inventory of only 11 segments, the phoneme / / has allophones , , and all of which are apparently in free variation (Firchow & Firchow 1969). Linguistics therefore makes a distinction between the abstract elements that can distinguish meanings of words, called phonemes, and their physical realization, which is called *phonetic* realization. In parts of this paper, frequent reference will be made to the phoneme inventories of languages, without making reference to their actual phonetic realizations. The reader should therefore be aware that in the use of these inventories of phonemes one should always ask the question: “What about allophonic variation?”

Regularities of systems of speech sounds:

The phoneme inventories of the world’s languages show at the same time remarkable diversity and remarkable regularities. In the UPSID451, the UCLA Phonological Segment Inventory Database (Maddieson & Precoda 1990, the first version is described in Maddieson, 1984), a database that contains the phoneme inventories of a representative sample of 451 of the world’s languages, a total of 921 different segments occurs. Of these, 652 are consonants, 180 are vowels and 89 are diphthongs. Apparently the human vocal tract is capable of producing an amazing diversity of sounds. Still, any single language only uses a small subset of these possible sounds. In the UPSID451, the smallest inventories are those of the East-Papuan1 language Rotokas (Firchow & Firchow 1969) and the South-American language Múra-Pirahã (Everett 1982, Sheldon 1974) both with only 11 phonemes. The language with the largest inventory is the Khoisan language IX (Snyman 1970) with 141 phonemes. The typical number of phonemes, according to Maddieson (1984), lies between 20 and 37.

The phonemes that a language uses are not chosen randomly from the possible sounds the human vocal tract can make. In fact some sounds appear much more frequently than others do. According to Lindblom and Maddieson (1988) the possible sounds of the world’s languages can be divided into basic articulations, elaborated articulations and complex articulations. Apparently languages with small inventories only use basic articulations; while for larger inventories elaborated and complex articulations are used.

This implies that some systems of speech sounds will occur more frequently than others. In fact, this is even more strongly the case than would be predicted from the above mentioned symmetries. However, 1 Details of genetic affiliation and location of languages have been taken from Grimes (1996). In a previous version of the UPSID with 317 languages, the former system occurs ten times, while the latter system does not occur at all (Vall 1994, Annexure 2). This occurs in 34 of the 317 languages, (Vall 1994) much more often than any other system. Certain systems seem to be favored, while others seem to be avoided.

Explanations of regularities based on features:

Apparently sound systems of languages show great regularities. One can now ask the question where these regularities come from. Traditionally explanations have been based on innate properties of the human language capacity. These explanations (see e.g. Jakobson & Halle 1956, Chomsky & Halle 1968) assume that there are (innate) features in the human brain that determine which distinctions between sounds can be learned. These features are usually binary. An example of a feature is nasality. A sound can either be nasal or not. Some of the features and some of their values are more marked than others. This means that certain distinctions are preferred over others, so that, for instance, a language would prefer to use the distinction high/low for vowels before it would use the distinction nasal/non-nasal. Non marked values of the features are preferred over the marked ones. For example, nasality for vowels is considered to be marked. Nasal vowels will thus be rarer than non-nasal ones. In general, sounds with unmarked features and unmarked values for these features will be more frequent than ones with marked features and values.

Although the theory of distinctive features is quite useful as a tool for describing sound systems of languages, it does not work very well for explaining the observed patterns. First of all, it is not quite clear which features should be used or even how many features there are. There are many ways in which languages can make phoneme distinctions (Ladefoged & Maddieson 1996). Some of these distinctions are only used in very few languages. Furthermore, languages make subtle sound differences that are not used to distinguish meanings. For example, the English word “coo”, the French word “cou” (neck), the German word “kuh” (cow) and the dutch word “koe” (cow) are all pronounced differently and perceived as recognizably different by speakers. It is not clear, however, how these subtle differences would have to be represented or explained in a distinctive feature framework. Also, there is no clear markedness hierarchy.

This can be seen from the fact that phoneme inventories of languages can differ in one segment. Apparently the markedness of the features cannot predict the sequence in which phoneme inventories grow. Furthermore, if innate features and markedness play a role, it still remains to be explained why and how these particular features became innate, preferably in an evolutionary framework. Finally, and most importantly, features and their markedness are derived from observation of linguistic data. It is therefore circular to “explain” this linguistic data with innate features and markedness, which have been derived from the very same data. Rather, one would like to have a theory that is based on independent, preferably physical, physiological or psychological data (see Lindblom *et al.* 1984, Lindblom *in press*.)

Carré’s distinctive region model:

Another theory for explaining the structure of sound systems is the *distinctive region model* developed by Ren Carr (Carr 1994, 1996, Carr & Mrayati 1995). This theory considers human speech communication as a near-optimal solution to the physical problem of producing communication over an acoustic channel using an acoustic tube that can be deformed. The theory assumes that an optimal communication system can produce maximal acoustic differences with minimal articulatory movements. Minimal articulatory movements are defined as linear and orthogonal deformations of a uniform acoustic tube. Carr uses a computational model with which he calculates the deformations of the uniform tube that result in maximal the acoustic distinctions. This model finds deformations that result in an acoustic space that corresponds to the vowel space of human sound systems. The uniform tube is divided into four *distinctive regions* that correspond to the regions of the vocal tract that are used in vowel production.

The model can be extended to predicting places of articulation of consonants by looking at maximal changes in formant frequencies. The uniform tube is then divided into eight distinctive regions, each corresponding to different places of articulation for consonants. This model is able to predict the possible places of articulation, as well as the available vowel space from purely physical principles and from the assumption that speech communication is a near-optimal solution to the problem of communicating with acoustic signals produced by a deformable acoustic tube. However, this model does not directly predict which of the possible articulations will be chosen for building a sound system (although see: Car 1996). Note also that there seems to be a discrepancy between the Stevens’ theory and Carr’s theory. Given a certain articulatory movement, Stevens seems to favour minimum acoustic change whereas Carr seems to favour maximal acoustic change.

How sound systems have become optimized:

However, this account is not quite complete. Apparently, more or less optimized sound systems are found in the world’s languages, but it is not clear *how* they have become optimized. Clearly, the individual language users and language learners do not do an explicit optimization. On the contrary, they try to imitate their parents and peers as accurately as possible. This can be observed from the fact that people make and observe much finer distinctions in their sound systems than are necessary for successful communication. This makes it possible that speakers of slightly different dialects of a language can understand each other perfectly, but still perceive that the other speaks a different dialect. The question why this is the case is very interesting, but falls outside the scope of this thesis. This fact will just be accepted as a given. Apparently, the sound system of a language is optimized to a certain extent, even though the language users themselves do not do any explicit optimization. However, as has been pointed out in the first section of this chapter, there are individual variations of the language that tend towards ease of production, understanding and learning.

Apparently then, there is a global optimization in the language, due to local interactions. This is an example of self-organization. In order to investigate this phenomenon and in order to check what exactly is its role in explaining the structure of sound systems, one has to abandon the point of view of language as a purely individual behavior and assume the point of view of language as a collective, complex dynamic behavior. Due to the complexity of self-organizing phenomena, the best way to investigate them is by building computer simulations.

Glotin’s AGORA model:

The first model to be described that used a simulation of a population in order to explain the properties of vowel systems was the AGORA-model by Herv_ Glotin (Glotin 1995, Glotin & Laboissière 1996, Berrah *et al.* 1996). It is based on a community of talking “robots” called *carls* (Cerveau Analytique de Recherche en Linguistique/ Cooperative Agent for Research in Linguistics). Each *carl* has a repertoire of vowels, that are represented both articulatory and acoustically. It is equipped with an articulatory model, based on Maeda’s model (Maeda 1989) with which it can produce acoustic signals consisting of formant patterns. Initially, for each *carl* a fixed number of vowels is chosen at random near the position of the neutral vowel. In the simulations, two *carl*s are selected from the population at random, and they both produce a vowel that is randomly chosen from their repertoire. They then find the vowel in their repertoire that is closest to the sound they hear. They shift this vowel, so that its acoustic signal will be closer to the sound they heard, and shift all the other vowels in their repertoire away from this signal.

Depending on the amount of shifting a *carl* does, a fitness is calculated. The less shifting a *carl* does, and thus the more it confirms to the sound systems in the other *carl*s, the fitter it will be. After a number of interactions between *carl*s, the least fit *carl*s are removed from the populations, and the fittest are used to calculate a replacing *carl*, in the way of a genetic algorithm (for an introduction, see Goldberg 1998). The vowel systems of the replacing *carls* are initialized with a cross between the vowel systems of the parent *carls*.

After a while the population usually converges to a common vowel system that looks like the most common vowel system in the world's languages for the given number of vowels (usually four or five). However, convergence was not guaranteed. There are a number of disadvantages to the AGORA-model. The first is that, due to the complexity of the Maeda articulatory model, the simulations are very calculation intensive. This made it impossible to use populations of any realistic size. The population size in most of Glotin's experiments was limited to five *carls* only. Also the number of vowels was limited to four or five. Furthermore, the model had great difficulties to converge. The genetic component was added in order to get the model to converge more rapidly. However, this genetic component confuses the simulation (is the driving force natural evolution or cultural self-organization?) and makes it quite unrealistic. Apparently a new *carl* can inherit a sound system, something which obviously does not happen in humans. Glotin is aware that this is unrealistic (Glotin, personal communication) but considers it a simplification of humans learning the sound system of their parents. He says that it does not influence the outcome of the experiments much, except for making them converge more rapidly. Another problem of his model is that the agents push the vowels in their vowel systems away from each other. This makes the model equivalent to Liljencrants and Lindblom's (1972) original simulation. As the agents do a local optimization of their vowel systems, the interactions between them are not crucial for the shape of the emerging vowel systems. An agent talking to itself would get the same results.

Simulation:

Simulating the development of a system of speech sounds in a potentially large population of agents requires a computer model that is at the same time realistic and fast. These are two contradictory requirements. Realism can only be increased by doing extra calculations, which reduces speed. Consequently, it is necessary to make a compromise. The properties that are really essential for getting results that are comparable with human speech will have to be kept, and the rest will have to be sacrificed. However, if realism is sacrificed in a sensible way, the results of the simulation will still be comparable with observations of data from real human languages. Of course, in doing this, one has to keep in mind which parts of the simulation were realistic and which ones were artificial. This simulation describes the computer model that was used for investigating the formation of vowel systems in a population of agents. It also describes and defends the choices that were made between realism and speed. In order to understand the reasoning behind these choices better, a short history of the simulation is first presented.

A first complex model:

The idea of building a simulation that models interactions with human-like speech sounds was first put forward in a discussion with Luc Steels (Ardennes, October 1995). The first simulation that was tried was quite ambitious. It was equipped with a simple articulatory synthesizer, with dynamically moving articulators and with a model of perception based on formant frequencies. The articulatory synthesizer was inspired by Ladefoged's observations (Ladefoged 1981 ch. 8, fig 12) of the dependence of formant frequencies on the positions of the articulators. He observed that first formant frequency seems to increase with decreasing height of a vowel and that the distance between the first and second formant frequency seems to increase with increasing frontness of a vowel, while lip rounding results in lowered first and second formant frequencies. But Ladefoged warns: "…measurements of formant frequencies are not so simply related to the other traditional labels high— low, front— back and spread— round." Nevertheless a crude synthesizer was based on these observations, together with data on the absolute formant frequencies of vowels. The synthesizer worked as follows:

F p F h p r F h p r = + × = + × + × - × = - × - × - ×
2200 1000 800 1000 600 200 900 500 100 100 3 2 1
where *h* is the height of the tongue, *p* is the position of the tongue in the front-back dimension and *r* is the rounding of the lips. All parameters are supposed to lie between 0 and 1, where 0 is the lowest height, the most front and the least rounded, respectively. As can be learned from a cursory examination of the formula, it is not very realistic. An acoustic signal consisted of 32 frequency bins. The frequency bins whose centre frequency was close to that of a formant frequencies were filled with a higher value than those whose centre frequency was far away from the formant frequencies. The value in a frequency bin corresponded to the energy of that frequency. A number of different methods for distributing energy over neighboring frequency bins were used, all of which were found to give similar results. The agents could produce dynamic utterances; they could move their articulators while producing sound. This means that they could, in principle, produce simple consonant-like sounds. They could not produce noise, so fricatives and plosive bursts could not be produced. However, rapid formant transitions that gave a consonant- like impression could be produced. Dynamic movement of articulators worked by approaching target articulatory values in the following way:

( )-1 -1 = + - *t t t p p* a *g p* where *pt* is the position at time *t*, a is a constant indicating the speed with which an articulator moves and *g* is the goal value. The complex utterances were built up of smaller units, equivalent to phonemes. The agents therefore had to store items at two levels. They had to store a list of phonemes as well as a lexicon of possible words. Whenever an agent heard a certain sound sequence, it would analyse it in terms of the phonemes it knew. This would then result in a sequence of phonemes, which could correspond to an existing word or form a new word. Previously unheard words could be added to the lexicon with a certain probability. Perception of sounds by agents was based on a heuristic to split up the speech stream and on a neural network. This network was a simple perceptron (Hertz *et al.* chapter 5) with an output for every phoneme. The phoneme that corresponded to the output with the highest value was recognised. The network was trained by making the agents listen to themselves while they were speaking. The inputs to the network were calculated in the same way as when the agent listened to another agent. The response to be learnt was to make active the output node corresponding to the phoneme that was said at that moment, while making all other outputs inactive. The interactions between the agents worked in the same way as the imitation game, which is described in more detail below. One agent would produce a word; the other would listen to it, analyze it in terms of its own phonemes and produce an imitation. The first agent would then check if the word it heard as imitation was the same as the word it originally said. If this was the case, the imitation was successful. If not, the imitation was a failure. Depending on the outcome of the interaction, the agents could add phonemes or words to their repertoire. They also kept score of the success of phonemes and words, and could occasionally throw away bad ones.

Results of the first complex model:

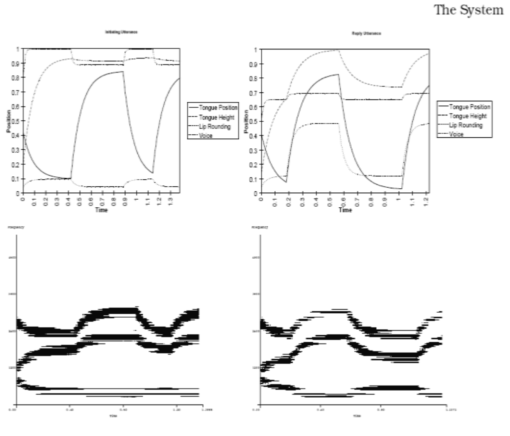

Fig. 1: Result of the first complex model

This model, which was tested only with a "population" of two agents, occasionally produced good imitations (see Fig 1) and the average success score of phonemes was around 30% (with inventories of up to five phonemes and up to a hundred words). This seems to indicate better than random performance, as with a repertoire of a hundred words one would expect a success rate of one percent when randomly selecting words. However, because of the way words of different length were compared, random performance would result in approximately 30% success, which made the simulation's performance no better than random. Any good imitations were mostly due to coincidence. This observation teaches that not all results should be trusted at face value. It is essential to analyze the performance of the agent in the case that it chooses its behavior without paying attention to the other agent's behavior (resulting in either random or simple systematic behavior). The experiments can only be considered successful if agent behavior is observed that is significantly better than random or simple systematic, input-less behavior.

A Feature-based model:

The complex simulation did not work satisfactorily. It did not result in a shared system of speech sounds. As no stable systems of sounds were reached, nothing could be said about the structure of these systems, either. There were several reasons for this failure. The model was too complex, because it worked with dynamic utterances, which had to be represented at multiple levels. Also the perception part, with its neural network, introduced too many extra parameters. At the same time, the sound production was too simplistic and unrealistic. A simpler, more controllable model was needed.

Results of the feature-based model:

This model resulted in successful imitations when used in a population of two agents. Words of average length three were created. This is not very long, but sufficient for co-articulation to play an important role. Lexicons up to 600 words and phoneme sets of up to 15 phonemes were generated. Success of word imitation went up to 70%, which in this case was much better than random behavior. Good similarities between the sound systems of the agents are found if one looks at the most used (minimally 5 times) and most successful (minimally 70% successful of all the times it was used) phonemes. A different approach to modelling speech sounds was therefore adopted. This approach was inspired by the work of Herv Glotin (Glotin 1995, Berrah *et al.* 1996, Glotin & Laboissière 1996) at the *institut de communication parl* in Grenoble. In his work a population of agents develops a set that consists of vowels only. Vowels can be modeled as single utterance, so there is no need for a separate word-level. Also, vowels are easy to synthesize. Furthermore, a lot is known about the universal tendencies of vowel systems of the world's languages. Therefore it was decided to first build a simulation to investigate the development of vowel systems in a population of agents. When this would work, and the dynamics were well understood, an extension to more complex utterances could be undertaken.

Purpose of the Simulation:

The purpose of the simulation is to investigate the emergence of a vowel system in a population of agents that learns to imitate each other as successfully as possible with an open system of vowel sounds. The agents' production, perception and learning of speech sounds should be as human-like as possible. Each agent should be able to produce, perceive and remember a set of realistic vowels. It should be able to engage in interactions with other agents and to learn and adapt its system of vowels from these interactions. The number of vowels it knows or their positions should not be determined beforehand. Once an agent has developed a vowel system that works, it should keep this system, without altering it too much. In a group, the agents should be able to generate such a system from scratch. The aim is not to model the exact way in which human vowel systems emerge and change historically, but to investigate whether a population is in principle able to develop a coherent set of vowels from scratch, and whether the sets of vowels that emerge show the same universal tendencies as human vowel systems. In order to keep the simulation tractable a number of things should *not* be modeled. First of all, the utterances of the agents do not have any meaning. They are just sounds. The goal of the agents is to imitate the other agents as well as possible. This is considered to be basic to language; only if you are capable of identifying and imitating the other speaker's sounds, can you begin to learn the meaning that is attached to the sounds. Other researchers (Steels 1997a, Steels & Vogt 1997, Gasser 1998) are investigating the origins of meaning and the way in which meanings can be coupled to words. The question why agents would want to communicate with language, and thus to imitate, is not posed either. In the work presented here, the need for communication with language is assumed as a given. Other researchers are investigating the origins of communication with language. Having the agents develop the need to imitate would complicate the model needlessly; this need is therefore pre-programmed. The drive to add new sounds to the inventory is also pre-programmed. It is needed, because the agents start out with empty sound systems, but still have an urge to imitate. It is therefore necessary to add new sounds every once in a while, in order to get the imitations started. In a natural language one can imagine that addition of new sounds would be driven by the need to distinguish as many meanings as possible, while keeping the length of utterances low. In order to make more distinctions, more sounds and an effective use of the available acoustic space is necessary. This is an example of a case where one part of language, the lexicon, exerts pressure on another part of language, the sound system.

Conclusion:

This paper investigated whether the emergence of vowel systems can be explained as the result of interactions in a population of agents that learn and use vowel sounds. It was shown with computer simulations that this was indeed possible. Moreover, it was shown that the universal properties of human vowel systems can be predicted accurately from these simulations. The frequently found vowel systems can be considered as attractors of the dynamic system that is formed by the articulatory and perceptual constraints of the agents and by the rules of the imitation game. Self-organization ensured that the resulting systems were coherent. It appears that these mechanisms must also be taken into account in the explanation of human vowel systems, and probably for other human speech sounds as well.

References:

1. Allen. Jonathan. M. Sharon Hunnicutt & Dennis Klatt (1987) From text to speech: The MITalk system. Cambridge: Cambridge University Press. Baldwin.
2. J. Mark (1896) A new Factor in Evolution. The American Naturalist 30 (June 1896) pp. 441-451, 536-555. Reprinted in RK Belew and M. Mitchell (eds.) Adaptive Individuals in Evolving Populations: Models and Algorithms.SFI Studies in the Sciences of Complexity. Proc, VoL XXVI. Addison Wesley. Reading. MA. 1996.
3. Batall, John (1998) Computational stimulations of the emergence of grammar. In Hurford et al. 1998, pp. 405-426.
4. Boe, Louis-Jean. Jean-I.uc Schwartz & Nathalie Vallee (l995), The Prediction of Vowel systems: perceptual Contrast and Stability. In Eric Keller (ed.) Fundamentals of Speech Synthesis and Speech Recognition. John Wiley, pp. 185-213
5. Boersma. Paul (1998) Functional Phonology The Hague: Holland Academic Graphtcs.
6. Browman, Catherine P. & Louis Goldstein (1995) Dynamics and Articulatory Phonology. In Robert F. Port & Timothy van Gelder (eds.) Mind as-Motion. MIT Press. Cambridge Mass. pp. 175-194.
7. Carlson. R. B. Granstrom & C. Fant (1970) Some studies concerning perception of isolated vowels. STL/QPSR (Speech Transmission Laboratory Quarterly Progress and Status Report, Department of Speech Communication and Music Acoustics. KTH. Stockholm) 2/3 pp. 19-35.
8. Carre, Rene (1994) 'Speaker' and 'Speech' Characteristics: A Deductive Approach. Phonetica 51. pp. 7-16.
9. Carre, Rene (1996) Prediction of vowel systems using a deductive approach. In Proceedings of the ICSLP 96. Philadelphia. pp. 434-437.
10. Carre, Rene & Maria Mody (1997). Prediction of Vowel and Consonant Place of Articulation. In John Coleman (ed.) Computational Phonology. Proceedings of a Workshop Sponsored by the Association for Computational Linguistics. Association of Computational Linguistics. pp. 26-32.
11. Chomsky, Noam (1965) Aspects of the Theory of Syntax. MIT Press. Cambridge, Mass.
12. Dunbar, Robin (1996) Grooming, gossip and evolution of language. London: Faber and Faber.
13. Elman, Jeffrey L. (1990) Finding Structure in Time. Cognitive Science 14, pp. 179-211.
14. Elman, Jeffrey L. & David Zipser (1988) Learning the hidden structure of speech. Journal of' the Acoustical Society of America 83(4), pp. 1615-1626.