

Chapter 10

Modelling the Emergence of Acoustic Segments (Phonemes) in Spoken Languages

Prof. Nik Kasabov

nkasabov@aut.ac.nz

<http://www.kedri.info>

Overview

- Introduction to the issues of learning spoken languages.
- The dilemma "Innateness versus Learning", or "Nature versus Nurture", revisited.
- ECOS for modelling the emergence of acoustic segments (phonemes).
- Modelling evolving bilingual systems.

Assumptions

- Several assumptions have been hypothesised and proven through simulation:
 - » the learning system evolves its own representation of spoken language categories (phonemes) in an unsupervised mode
 - » learning words and phrases is associated with supervised presentation of meaning;
 - » it is possible to build a 'life-long' learning system that acquires spoken languages in an effective way, possibly faster than humans, provided there are fast machines to implement the evolving, learning models.

Introduction to learning spoken languages

- Concerned with the process of learning in humans and how this process can be modelled in a program.
- The following questions will be attempted:
 - » How can continuous learning in humans be modelled?
 - » What conclusions can be drawn in respect to improved learning and teaching processes, especially learning languages?
 - » How is learning a second language related to learning a first language?
- The aim is computational modelling of processes of phoneme category acquisition, using natural, spoken language as training input to an evolving connectionist system
- Basic methodology consists in the training of an evolving connectionist structure with mel-scale transformations of natural language utterances

Introduction to learning spoken languages

- In preliminary experiments, it may be advisable to study circumscribed aspects of a language's phoneme system, such as consonant-vowel syllables
- Moreover, it will be possible to simulate acquisition under a number of input conditions:
 - » input from one or many speakers;
 - » small input vocabulary vs. large input vocabulary;
 - » simplified input first (e.g. consonant-vowel syllables) followed by phonologically more complex input;
 - » different sequences of input data.

“Nature versus Nurture” revisited

- The functioning of an organism is a result of both evolution and development (*Nature versus Nurture*)
- Dilemma of Innateness vs Learning
 - » Arguments for the innateness include rapidity with which children acquire a language, the fact that explicit instruction has little effect on acquisition, and the similarity of all human languages.
 - » Arguments against also invoked: the complexity of natural languages is such, that they could not, in principle, be learned by normal learning mechanisms of induction and abstraction.
- In computational terms, contrast is between systems with a rich in-built structure, and self-organizing systems that learn from data.
- Can regard the existence of phoneme inventories as a converging solution to two engineering problems.
 - » A speaker of any language has to store a vast inventory of meaningful units (morphemes, words, or fixed phrases)
 - » The acoustic signal contains a vast amount of information

Infant Language Acquisition

- New-born infants are able to discriminate a large number of speech sounds
- By about 6 months the ability to discriminate phonetic contrasts that are not utilized in the environmental language declines
- The "perceptual space" of e.g. the Japanese- or Spanish-learning child becomes increasingly different from the perceptual space of the English- or Swedish-learning child
- It is likely that adults in various cultures, when interacting with infants, modulate their language in ways to optimise the input for learning purposes.

Phonemes

- Phonemes are theoretical entities, at some distance from acoustic events.
- Limited storage and processing capacity requires that words be broken down into constituent elements, i.e. the phonemes
- Different procedures for identifying the phones and phonemes of a language.
 - » the principle of contrast - can be used in a modelling system through a feedback from a semantic level

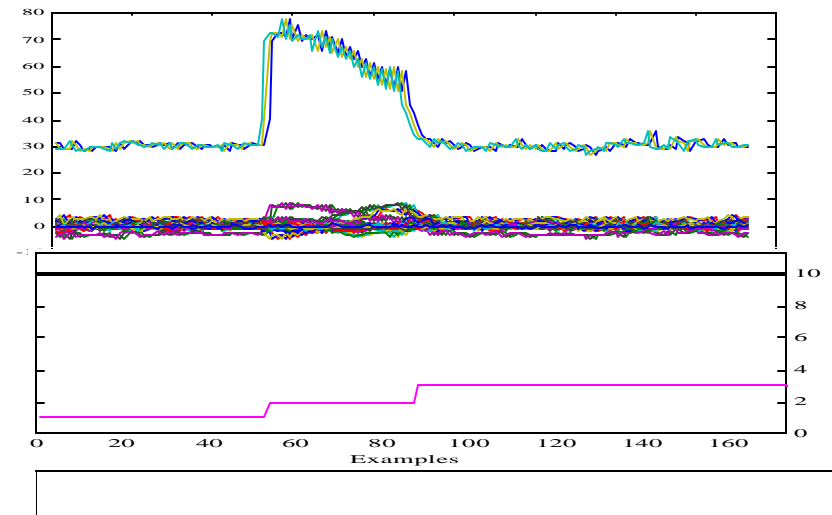
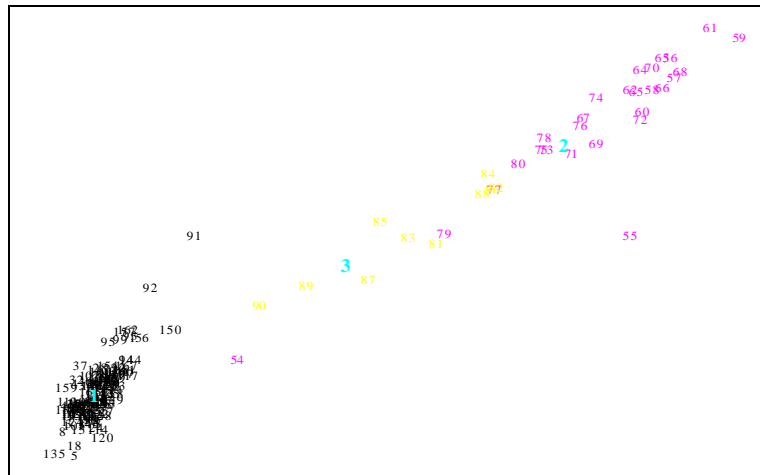
ECOS for Modelling the Emergence of Phones and Phonemes

- Can conceptualise sounding of a word as a path through multi-dimensional acoustic space.
- As the number of word types is increased, the trajectories of different words will overlap in places; these overlaps will correspond to phoneme categories.
- A minimum expectation of a learning model, is that the language input will be analysed in terms of an appropriate number of phoneme-like categories
- Basic question:
 - » *Can the system organize the acoustic input with minimal specification of the anticipated output?*

Issues for modelling of phoneme acquisition

1. Does learning require input which approximates the characteristics of "motherese" with regard to careful, exaggerated articulation, also with respect to the frequency of word types in the input language?
2. Does phoneme learning require lexical-semantic information?
3. Critical mass.
4. The speech signal is highly redundant, in that it contains vast amounts of acoustic information that is simply not relevant to the linguistically encoded message
5. Can the system organize the acoustically defined input without prior knowledge of the characteristics of the input language
6. What is the difference, in terms of acoustic space occupied by a spoken language, between simultaneous acquisition of two languages versus late bilingualism?

Phoneme modelling using Evolving Clustering Method (ECM)



- *Experimental results with an ECM model for phoneme acquisition a single pronunciation of the word “eight”.*

The two dimensional input space of the first two Mel scale coefficients of all frames taken from the speech signal of the pronounced word “eight” and numbered with the consecutive time interval and also the evolved nodes that capture each of the three phonemes of the speech unit. (fig. 10.1 (a))

ECOS model for word clusters

- Aim to develop a supervised model based on both ECM for phoneme cluster emergence, and EFuNN for word recognition.
- The outputs of the EFuNN are the words that are recognized.
- The number of words can be extended over time thus creating new outputs that is allowable in an EFuNN system.
- A sentence recognition layer can be built on top of this model. This layer will use the input from the previous layer (a sequence of recognized words over time) and will activate an output node that represents a sentence (a command, a meaningful expression, etc).
- At any time of the functioning of the system, new sentences can be introduced to the system which makes the system evolving over time.

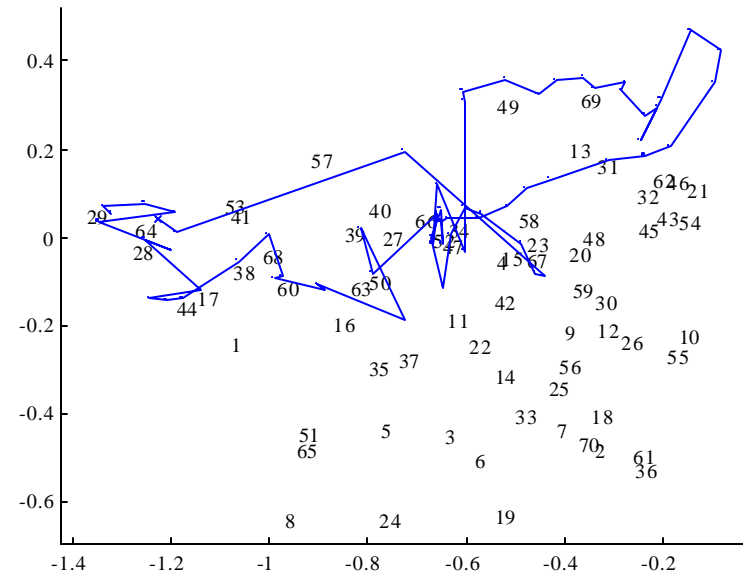
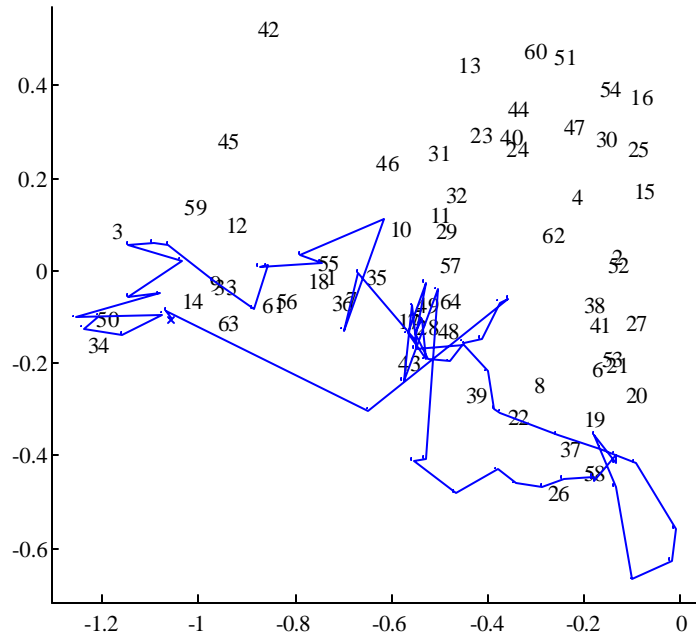
Modelling Evolving Bilingual Systems

- Once satisfactory progress has been made with modelling phoneme acquisition within a given language, a further set of research questions arises concerning the simulation of bilingual acquisition.
- We can distinguish two conditions:
 1. Simultaneous bilingualism. From the beginning, the system is trained simultaneously with input from two languages.
 2. Late bilingualism. This involves training an already trained system with input from a second language.
- Children manage bilingual acquisition with little apparent effort
- Late acquisition of a second language is typically characterized by interference from the first language
- The areas of the human brain that are responsible for the speech and the language abilities of humans evolve through the whole development of an individual

Modelling Evolving Elementary Acoustic Segments (Phones) of Two Spoken Languages

- Modeling the emergence of multiple spoken languages
- Example: Spoken words from two languages are presented to an evolving systems in three modes:
 - (a) English first and Maori second ->
 - (b) Maori first and English second
 - (b) English and Maori mixed
- The word 'zoo' is plotted as a trajectory in the two learned spaces (the 2D of the first two principal components)
- Learning two languages simultaneously leads to a more compact space
- Adding other languages on-line: Japanese; Spanish

Modelling Evolving Elementary Acoustic Segments (Phones) of Two Spoken Languages



Left - Projection of spoken word "zoo" in English+Maori PCA space (fig. 10.9)

Right - Projection of spoken word "zoo" in mixed English and Maori PCA space (fig. 10.10)

Summary

- Presented ECOS approach to modelling emergence of acoustic clusters related to phones in a spoken language, and in multiple spoken languages.
- Modelling the emergence of spoken languages is an intriguing task - not solved yet despite existing papers and books.
- Simple evolving model presented illustrates main hypothesis raised in this material, that acoustic features such as phones, are learned rather than inherited.
- The evolving of sounds, words and sentences can be modelled in a continuous learning system that is based on evolving connectionist systems and techniques ECOS.

Further Readings

- Generic readings about linguistics and understanding spoken languages (Taylor, JR, 1995, 1999; Segalowitz, 1983; Seidenberg, 1997).
- The dilemma “Innateness versus Learned” in learning languages (Chomsky, 1995; Lakoff and Johnson, 1999; Elman et al, 1997).
- Learning spoken language by children (Snow and Ferguson, 1977; MacWhinney, Fletcher and MacWhinney, 1996; Juszuk, 1997).
- The emergence of language (Brent, 1996; Culicover, 1999; Deacon, 1988; Pinker, 1994; Pinker and Prince, 1988).
- Models of language acquisition (Parisi, 1997; Regier, 1996).
- Using evolving systems for modelling the emergence of bilingual English- Maori acoustic space (Kilgour et al, 2002; Laws, 2001; Kilgour, 2002).
- Modelling the emergence of New Zealand spoken English (Kilgour, 2002).