SPEECH PERCEPTION AND TRANSITION OF SOUND CHANGE

Ching-Pong AU

Laboratoire Dynamique Du Langage, CNRS - Université Lumière Lyon 2
Ching-pong.au@ish-lyon.cnrs.fr

ABSTRACT

A dynamic multi-agent model was built in order to simulate language acquisition and sound change in a speech community. The simulation results provide plausible solutions that resolve some controversial issues regarding the sound change implementation such as the discrepancy between the Neogrammarian hypothesis and the lexical diffusion hypothesis.

In the simulations, the patterns described by the two seemingly contradictory hypotheses both exist in the implementation of sound changes depending on the consistency of perceptual responses of the speakers in the population.

Keywords: language acquisition, sound change, computer simulation, Neogrammarian controversy.

1. INTRODUCTION

Whether phonological change is regular or not has been debated for a long time. In the late 19th century, a German school of linguists, called Neogrammarians, proposed that the same sound in all lexical items change simultaneously and gradually [14]. In late 1960s, Wang proposed the lexical diffusion. The hypothesis suggests that sound change is initiated by a single word or a small group of words and then spreads to the other words. The change of its phonetic quality is considered to be abrupt [17].

As both Neogrammarian regularity and lexical diffusion patterns were both supported by empirical data, Labov attempted to resolve the controversy by proposing that regular sound change and lexical diffusion patterns should distribute complementarily depending on different types of speech sounds [11], [12]. However, some lexical diffusionists still believe that lexical diffusion is the basic mechanism in the implementation of sound changes [5], [19]. But some other historical linguists (for example, [4]) denied entirely the existence of lexical diffusion and argued that the irregularity in language is only due to borrowing or other reasons such as morphological or syntactic conditioning (e.g. [4], pp198-200). No matter from the points of view of linguists or the empirical evidences that support the hypotheses [2], [16], [18], there is a great deal of inconsistency between the two sides. Labov called the debate between the two parties “Neogrammarian controversy” [11].

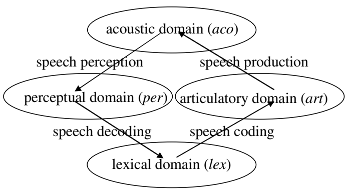
In most of the previous studies, instead of getting the whole set of data, the on-going sound change data were sampled from a population, and usually only the production data were collected. Intuitively plausible hypotheses were then built up based on the samples. However, to understand the actual process of sound change, solely observing the on-going changes in the speech sampled from the populations is not sufficient to know the whole story behind the changes, because language, such a highly complex system, involves on the one hand the interactions among cognitive subsystems of individual speakers and on the other hand the interactions among speakers. Thus the mechanisms linking up all these aspects in both acquisition and sociological levels are not always obvious to be seen from the sampled speech data from the speech communities.

To trace the connections between the cognitive and sociological levels and the mechanisms, complete sets of data are necessary. However, to collect a rather complete set of data from an on-going sound change including the perception and production measurements of all speakers is obviously impossible. In this context, computer simulation becomes a relevant tool to seek for more precise hypotheses that are not only intuitively but also operationally plausible. The model itself can also become a vivid demonstration of how sound change evolves in progress. Such a computational model was built and will be introduced in the present article.

2. A DOMAIN-BASED CIRCUITRY

Unlike many previous models, the present computational model does not follow some basic assumptions that are usually made in the traditional phonology: (1) the symmetry between perception and production, (2) the irrelevance of phonetics in the internal phonological representations, and (3) the discreteness of boundaries between sounds. Therefore, three important new properties are introduced in the model. Firstly, perception and production are stored in different cognitive subsystems (The words "perception" and "production" here do not involve the meanings. For example, the perception subsystem in the model only stores the relationship between the acoustic signals in the physical world and the perceptual feeling towards the signals). Secondly, instead of using abstract symbols to represent internal phonemes, numeric values are used to represent both the internal signals in the cognitive domains and the information stored in the cognitive subsystems. Thus phonetic details can be retained internally. Thirdly, perceptual categories are not assumed with infinitely sharp boundaries in the model. The formation of perceptual categories is driven by statistical distributions of sounds that the infants listened to (see [10], [13]). The formation process is simulated by self-organizing maps (SOM) that have been used in a number of computational models regarding the development of speech perception (e.g. [3], [8], [15]).

Figure 1: The cognitive system in the model.

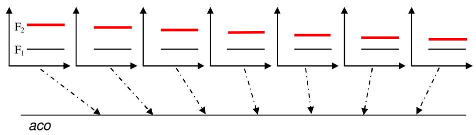


Based on the above characteristics, the cognitive structure of the agents is constructed as shown in figure 1. Each agent in the simulated population consists of four cognitive subsystems, perception, decoding, coding and production (the arrows). The perception and production subsystems are simulated by SOMs that store how the sounds are perceived and produced. The decoding and coding subsystems store the causative relationships "from sound to meaning" and "from meaning to sound" respectively. The domains of the input and output of the four subsystems are called acoustic domain (*aco*) and perceptual domain (*per*), lexical domain (*lex*) and articulatory domain (*art*).

2.1. The *aco-per* & *art-aco* relationships

The *aco* and *per* are respectively the input and the output of the speech perception subsystem. Both domains are both high-dimensional spaces that exhaustively include all the possibilities. For example, in the acoustic domain, all the possible acoustic signals can be found. These possibilities can all be arranged in the way that physically similar acoustic signals are located closer. Figure 2 shows a region in the acoustic domain (the horizontal line marked '*aco*'). Across one particular dimension, the value of the second formant varies while the other formants and acoustic parameters keep unchanged.

Figure 2: The acoustic domain. Different locations (indicated by the arrows) along the dimension represent the acoustic signals with different second formant values.



The acoustic domain is the only domain shared by all the agents in the population. Any acoustic signal that is produced or perceived by a human being (an agent in the model) can be represented by a point in this domain. That is to say, the acoustic domain encloses all possibilities of acoustic signals including both speech and non-speech.

The acoustic domain encloses all possible acoustic signals that physically exist, while the perceptual domain is an internal domain in each agent. The organization of the perceptual space differs from agent to agent because of individual auditory experience.

To understand how perceptual phenomena arise, Eimas et al [6] investigated infants' perception and found that the perceptual categories exist in a very young age. The findings lead to a hypothesis that the perceptual categories exist innately and abilities to differentiate categories that are not used in the native language will then be gone as meanings are learnt [7]. However, some recent findings suggest that the perceptual response of young infancy alters during the infants' exposure to the speech environment before they learn the words. They also suggest that the perceptual categories of very young infants from different countries emerge in different acoustic locations. Since these infants at their ages have not yet learnt the meanings of the lexical items, the experiment results lead to a new view that the development of speech perception is an automatic process driven solely by the input signals: the perceptual response will warp around an acoustic location if the speech signals around that location have been heard by the infant frequently [10], [13].

Figure 3: The emergence of perceptual categories in different linguistic environments

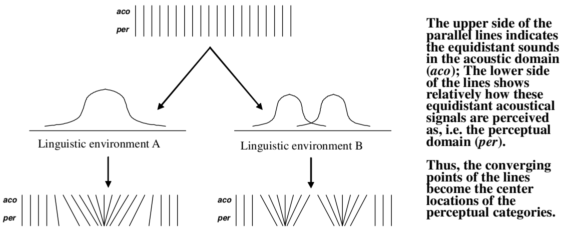


Figure 3 shows the results of the adaptations to the different linguistic environments. Initially, when an infant has not yet learnt anything from the outside world, categories do not emerge yet. Say, in environment A, the sounds heard by the infant are distributed around one acoustic location. One perceptual category gradually emerges in the perceptual domain. On the contrary, if the sounds are distributed around two distinct acoustic locations as in environment B, two distinct perceptual categories emerge gradually in the infant's perceptual domain. Thus infants grow up in environment A are less sensitive to the difference between the two sounds used in linguistic environment B.

This self-organizing adaptive mechanism is implemented in the perceptual subsystem in each agent in the model. It is simulated by a self-organizing neural network called SOM. The input of the SOM is the acoustic domain while its output is the perceptual domain. Similarly, another SOM is used to simulate the production bias of the agent, because it is more difficult to produce a foreign sound than a native sound for an adult. In this SOM, the input is the articulatory domain, while the output is the acoustic domain.

2.2. The *per-lex* & *lex-art* relationships

The decoding and coding subsystems are used for storing the mapping between sounds and meanings. During the acquisition of meaning comprehension of young agents, mappings between the PERCEIVED sounds and the lexical meanings the sounds convey are stored in the decoding subsystem. The perceived sound comes from the perceptual domain while the conveyed meaning comes from the lexical domain.

Finally, the acquisition of *lex-art* relationship is the most complicated learning for the agents. Meanings are randomly picked to express during the acquisition of speaking. Each time, the agents will listen to their own pronunciations. When they find that the acoustic signals produced are not the meanings they intend to produce, they will adjust the production parameters by trial and error in order to seek for better production parameters (see [1] for further details of the model set up)

3. RESULTS AND CONCLUSION

Agents with the described new cognitive structures form a speech community, in which the agents acquire their communication systems from the other more senior agents in an automatic manner. As the same procedures repeat for many times, the sound system of this virtual population evolves slowly. Phenomena regarding sound change can be observed from the simulations and then compared with the empirical observations from the reality.

In the simulations, the patterns described by the two seemingly contradictory hypotheses both exist in the implementation of sound changes depending on the consistency of perceptual responses of the speakers in the population.

In figures 4 and 5, each rectangular block displays the three overlapping domains (acoustic, perceptual and articulatory domains) of an agent. The small circles (and rhombuses) indicate the pronunciations of the vowels in different words by the agent; the solid lines enclosing the circles/rhombuses represent a perceptual category such that the acoustic values within this area will be more likely perceived as the same vowel by this agent. Each column shows 3 sampled agents from one generation and 5 generations in total are shown in each figure.

As shown in figure 4, during a shift without fusion of sounds, the perceptual patterns of the sounds are consistent for all the speakers in the population. Hence, the spoken forms of the lexical items change regularly and gradually as described in the Neogrammarian hypothesis (figure 4). However, during a merger as shown in figure 5, at the beginning, the spoken forms display a regular pattern as in a shift. But after some time, the changing patterns become lexically irregular and phonetically abrupt as in the description of lexical diffusion. It is found in the simulations that the irregular phenomenon in spoken forms is caused by inconsistent perceptual patterns among speakers in the population when the two perceptual categories are fusing together (see [1] for further details of the simulation results).

Interestingly, for both cases in figures 4 and 5, the groups of words marked by circles scattering along the dimension in various sizes for different individuals throughout different generations. The phenomenon resembles Helgason's hypothesis that phonetic variation can be expected to expand and/or contract over time [9]. More precisely, from the simulation results, we found that the scattering sizes for the spoken forms are highly related to the regions enclosed by the perceptual categories.

Figure 4: The transition of a shift.

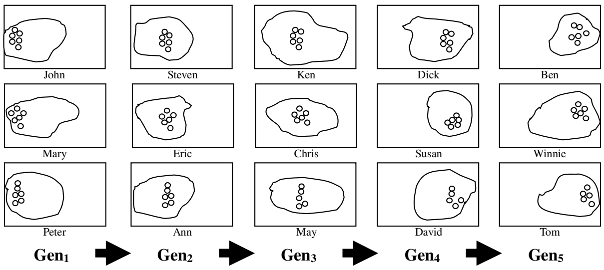
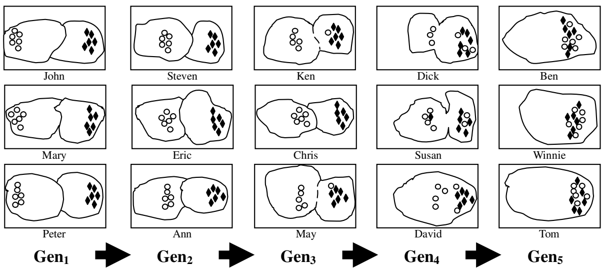


Figure 5: The transition of a merger of two sounds.



In conclusion, the present simulation suggests that some apparently contradictory phenomena in historical linguistics are actually explainable, if the phonetic details of sounds, for both perception and production sides, are stored in the human cognitive system during acquisition.

As more and more studies on the complexity of language and its change and emergence, we knew that sound change is not always predictable. Stochasticity from the interaction in the society can cause very different consequences. By repeating the same simulation experiments for many times, we found that whether a significant sound change occurs and whether it is a merger or a chain shift are usually not predictable even the initial conditions are similar. We have generalized the tendency of when Neogrammarian pattern and lexical diffusion might occur. If the present simulations reflect the reality accurately, both Neogrammarian and lexical diffusion hypotheses should not be the most basic mechanisms of the transitions of sound changes. Instead, they are only two possibilities of transition patterns generated by the interaction among the agents' perception and production.

4. REFERENCES

[1] Au, C.-P. Accepted. *Acquisition and Evolution of Phonological Systems*. Language and Linguistics Series B Frontiers in Linguistics III, Institute of Linguistics, Academia Sinica.

[2] Bauer, R. S. 1979. Alveolarization in Cantonese: A case of lexical diffusion. *J. Chi. Ling.*, 7**,** 132-141.

[3] de Boer, B. G., Kuhl, P. K. 2001. Human and Computer Acquisition of Vowel Categories. In: Smits, R., Kingston, J., Neary, T. M., Zondervan, R. (eds), *Proc. Workshop on Speech Recognition as Pattern Classification*, 115–120.

[4] Campbell, L. 1998. *Historical linguistics: An Introduction*, MIT Press.

[5] Chen, M., Wang, S.-Y. W. 1975. Sound Change: Actuation and Implementation. *Language* 51: 255-81.

[6] Eimas P. D., Siqueland E. R., Jusczky P., Vigorito J. 1971. Speech perception in infants, *Science*, 171: 303-306.

[7] Eimas, P. D., Miller, J. L., Jusczyk, P.W. 1987. On Infant Speech Perception and the Acquisition of Language. In: Harnad, S. (ed.), *Categorical Perception: The Groundwork of Cognition*, 161-195.

[8] Guenther F. H., Gjaja M. N. 1996. The Perceptual Magnet Effect as an Emergent Property of Neural Map Formation. *JASA*, 100, 1111-1121.

[9] Helgason, P. 2002. *Preaspiration in Nordic Languages: Synchronic and diachronic aspects*. PhD Dissertation. Stockholm University.

[10] Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., Lindblom, B. 1992. Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age. *Science***,** 255, 606-608.

[11] Labov, W. 1981. Resolving Neogrammarian Controversy. *Language*, 57: 267-309.

[12] Labov, W. 1994. *Principles of Linguistic Change, vol. 1: Internal Factors*. Oxford: Blackwell.

[13] Maye, J., Werker, J. F., Gerken, L. 2002. Infant Sensitivity to Distributional Information can effect Phonetic Discrimination. *Cognition*, 82(3), B101-B111.

[14] Osthoff, H., Brugmann, K. 1878. Morphologische Untersuchungen auf dem Gebiete der indo-germanischen Sprachen, Vorwort I. iii-xx. (English Translation in Lehmann 1967)

[15] Oudeyer, P-y 2001. Coupled Neural Maps for the Origins of Vowel Systems. In Dorffner, G.., Bischof, H., Hornik, K. (eds), *Proc. International Conference on Artificial Neural Networks, LNCS 2130*, 1171-1176.

[16] Shen, Z. 1997. *Exploring the Dynamic Aspect of Sound Change, J Chi. Ling.* Monograph No. 11.

[17] Wang, W. S-Y. 1969. Competing Changes as a Cause of Residue. *Language*. 45:9-25.

[18] Wang, W. S-Y. 1977. (ed) *The Lexicon in Phonological Change.* The Hague: Mouton.

[19] Wang, W. S-Y., Lien, C. 1993. Bidirectional Diffusion in Sound Change. In Jones, C. (ed), Historical Linguistics: Problems and Perspectives. London; New York: Longman. 345-400.