Convergence and stability in speech-like communication systems with a many-to-one mapping in the production process

A comprehensive investigation of sound change in spoken language requires not only to examine the characteristics and mechanisms of change but also the conditions of stability. Without understanding stability, clarifying the underpinnings of change is in danger of remaining patch work. To use a metaphor and put it simply, with sound change, there seems to be always an elephant in the room in the form of the implicit unspoken question: Why is there not constant change?

This study investigates convergence and stability in communication systems with a many-to-one mapping between different stages in the production of communicative signals: Can such a (partially) ambiguous relationship be overcome by the self-organisation of an emerging communication system? In human speech production the articulatory-to-acoustic transformation is considered to be such a many-to-one mapping: Different (static) vocal tract shapes can produce the same acoustic output (e.g., Atal et al., 1978). There is some limited evidence that the ambiguity of the underlying vocal tract shape can be resolved by context information when looking at dynamic articulatory trajectories (e.g., Hodgen, 2007). However, for this study we assume that at least some ambiguities are left and will show that - even if they are substantial - they do not lead to problems in emerging communication systems.

We employed agent-based modelling to investigate the properties of evolving simple communication systems (see Steels, 1997, 2000, for an overview) with or without a many-to-one mapping in the production process. We limited the modelling to aspects of the lexicon omitting problems of syntax and pragmatics altogether and simplifying semantics to a small set of meanings being the same for all agents. The synthetic virtual agents had the ability to move freely on a two-dimensional plane within circular boundaries. In the simulation the agents start with an empty tiered binary 'gestural score' repertoire. When an agents happens to wander within ’perceptive proximity’ of another agent, the agents interact with each other (’speak’). The active agent chooses randomly a meaning from its lexicon and retrieves the associated gestural score. If the meaning has not yet a gestural score assigned, the agent generates one randomly avoiding, however, scores it already uses. It then produces an ’acoustic’ signal from the gestural score, a rational number, subject to either a one-to-one mapping (OO) or a many-to-one mapping (MO). The interlocutor agent tries to recognise the signal using a one-dimensional Gaussian Mixture Model for the backward conversion to the gestural score. If the interlocutor agent retrieves the same meaning as the one intended by the producing agent, they both register the success. Agents have a limited lifetime, thus, they will disappear sooner or later. Every time an agent ceases from existence, a new agent is added with an empty gestural score repertoire.

A trial consisted of three simulations runs of 100,000 cycles with 80 agents in each of the two mapping conditions - one-to-one mapping (OO) and many-to-one mapping (MO) - using different levels of simulated Gaussian articulatory and acoustic noise (doubled each time). The lifetime of the agents was set to vary randomly between 10,000 and 100,000 cycles, thus at the end of the simulation all agents were second generation. Ten trials were conducted.

The results show that - given the simple but general model of the communicative process applied in the study - a many-to-one mapping does not prevent convergence, though the average number of converged meanings across all agents has been found to be lower in MO than OO. This difference, however, decreases significantly at the highest noise level. The emerging systems appear to approach asymptotically a stable maximally converged state which, however, consists of several 'dialects' - variations shared by a big number of agents but not across all agents.