Grounding Symbols in the Physics of Speech Communication

S. F. Worgan and R. I. Damper
School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK.
email {sw205r|rid}@ecs.soton.ac.uk

The symbol grounding problem (i.e., 'How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads?', Harnad 1990) is crucial to cognition. Thus, it has been argued that grounding poses a challenge that cannot be neglected (Cangelosi, Greco, and Harnad 2001). We believe human communication to be the clearest, certainly best developed, example of externally grounded cognition. Despite the advantages inherent in considering speech as a grounded system, there is a danger—through simulating at too high a level of abstraction—of effectively ignoring this crucial aspect (e.g., de Boer 2000; Oudeyer 2005). But how are we to define grounding at the 'phonetic' level of speech sounds? In this paper, we argue that the emergence of speech can and should be grounded in the physics of speech communication between agents, recognising that the human's contact with the external world of sound is via their articulatory and auditory systems.

We proceed by adopting the view of speech communication offered by Lindblom, MacNeilage, and Suddert-Kennedy (1984). Specifically, we are seeking to minimise the articulatory effort of an utterance, at the same time maximising its perceptual distinctiveness to other agents. In grounding terms, the drive for perceptual distinctiveness is important in shaping the coupled production-perceptual system. The higher the perceptual distinctiveness, the clearer the meaning of the utterance. This kind of interaction has already been investigated by Kirby (2001) at the syntactic level (and so tacitly assumes the emergence of phonetic distinctiveness). Having defined the nature of phonetic grounding, we are currently implementing a system that introduces this grounding into Oudeyer's (2005) previously ungrounded investigations, Figure 1. Following Guenter and Gjaja (1996), Oudeyer's work has shown how two self-organising maps (SOMs, see Kohonen 1990)—one representing the auditory system and the other the articulatory system—can converge from producing a series of random utterances to producing a shared set of discrete speech sounds. This process is considered analogous to the emergence of early hominid speech. However, without any definition of articulatory effort or perceptual salience, this convergence process often terminates in one central point (as found by Oudeyer and confirmed by us). We propose to overcome this problem, and hopefully produce more realistic utterances, by defining a *contour space* within each SOM, i.e., an objective function which embodies measures of both effort and distinctiveness. Therefore, as well as converging to a shared language (shared between agents, that is), each SOM will attempt to optimise itself within its contour space.

This definition of contour spaces—as embodying the effort of the utterance within the articulatory system and the perceptual distinctiveness within the auditory system—provides a direct grounding to the sensory-motor process of each individual agent. The articulatory effort is measured by the muscle energy expenditure (Umberger, Karin, and Philip 2003) of an artificial vocal tract (Maeda 1982), which forms the means whereby the agent acts upon its environment, i.e., its motor process. The perceptual contour space is dictated by the human peripheral auditory system, modelled on the work of Pont and Damper (1991)—the sensors of the agent. Although this system is grounded within its environment, it does not yet form (or manipulate) any explicit symbols. However, distinct and grounded attractors do emerge during the lifetime of the agent(s), and these we count as 'symbols'.

We are still grounding the external world via these attractors, but rather than connecting an imperfect, arbitrary abstraction (as when a cat in the environment is miraculously labelled CAT in one bound), we are connecting a more complete representation of the distal object, built on the physics

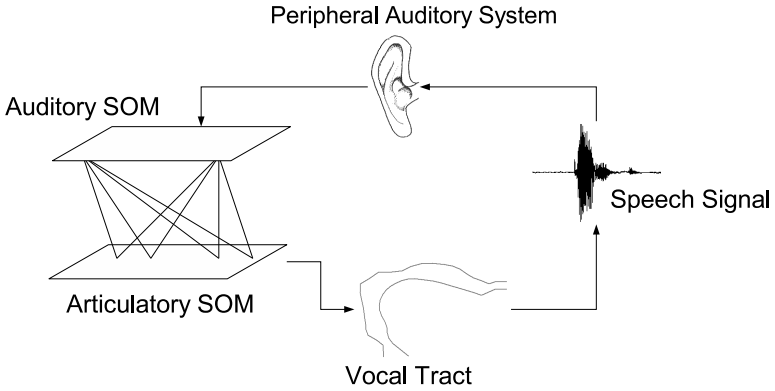

Figure 1: An agent producing and listening to its own utterances.

of the situation. Through the definition of attractors we have both a clear shared abstraction, its centre point, and a basin of attraction capturing the ambiguity and differences present in the real world. We feel that this view, based on emergence of attractors in articulatory-auditory spaces, can answer some of the current criticisms of the symbol grounding paradigm (Lakoff 1993), just because the attractors capture the ambiguities and 'shades of grey' that challenge more traditional grounded implementations (Davidsson 1993). This has precedence in other grounded implementations (e.g., Harnad 1993; Damper and Harnad 2000) that take the form of grounded, connectionist (neural network) models. These have been successful in displaying various aspects of human cognition. But, by considering grounding at the phonetic level, we have developed a new framework in which this interplay between symbolic grounding and connectionist systems can be further explored.

References

Cangelosi, A., A. Greco, and S. Harnad (2001). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi and D. Parisi (Eds.), *Simulating the Evolution of Language*, pp. 191–210. London: Springer-Verlag.

Damper, R. I. and S. R. Harnad (2000). Neural network models of categorical perception. *Perception and Psychophysics 62*(4), 843–867.

Davidsson, P. (1993). Toward a general solution to the symbol grounding problem: combining machine learning and computer vision. In *Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How?*, Raleigh, NC, pp. 157–161.

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

Guenter, F. H. and M. N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America 100*(2), 1111–1121.

Harnad, S. (1990). The symbol grounding problem. *Physica D 42*, 335–346.

Harnad, S. (1993). Grounding symbols in the analog world with neural nets. *Think 2*(1), 12–78.

Kirby, S. (2001). Spontaneous evolution of linguistic structure – an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation 5*(2), 102–110.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE 78*(9), 1464–1480.

Lakoff, G. (1993). Grounded concepts without symbols. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Society*, Boulder, CO, pp. 161–164.

Lindblom, B., P. MacNeilage, and M. Suddert-Kennedy (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, and Ö. Dahl (Eds.), *Explanations for Language Universals*, pp. 181–203. New York, NY: Mouton.

Maeda, S. (1982). A digital simulation method of the vocal-tract system. *Speech Communication 1*(3 – 4), 199–229.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology 233*(3), 435–449.

Pont, M. J. and R. I. Damper (1991). A computational model of afferent neural activity from the cochlea to the dorsal acoustic stria. *Journal of the Acoustical Society of America 89*(3), 1213–1228.

Umberger, B. R., G. M. G. Karin, and E. M. Philip (2003). A model of human muscle energy expenditure. *Computational Methods of Biomechanical and Biomedical Engineering 6*(2), 99 – 111.