John Benjamins Publishing Company

This is a contribution from *Interaction Studies 8:1*
© 2007. John Benjamins Publishing Company

This electronic file may not be altered in any way.
The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.
Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute.
For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).
Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Grounding symbols in the physics of speech communication

S.F. Worgan and R.I. Damper
School of Electronics and Computer Science, University of Southampton

The traditional view of symbol grounding seeks to connect an *a priori* internal representation or 'form' to its external referent. But such a 'form' is usually itself systematically composed out of more primitive parts (i.e., it is 'symbolic'), so this view ignores its grounding in the physics of the world. Some previous work simulating multiple talking/listening agents has effectively taken this stance, and shown how a shared discrete speech code (i.e., vowel system) can emerge. Taking the earlier work of Oudeyer, we have extended his model to include a dispersive force intended to account broadly for a speaker's motivation to increase auditory distinctiveness. New simulations show that vowel systems result that are more representative of the range seen in human languages. These simulations make many profound abstractions and assumptions. Relaxing these by including more physically and physiologically realistic mechanisms for talking and listening is seen as the key to replicating more complex and dynamic aspects of speech, such as consonant-vowel patterning.

Keywords: origins of speech sounds, symbol grounding, signal grounding, multi-agent simulation, self-organisation, emergent phenomena

Introduction

The computational metaphor that underpins cognitive science, much of artificial intelligence and functionalist philosophy of mind sees intelligent behaviour as the product of the workings of a formal symbol manipulation system (e.g., Newell, 1973; Minsky, 1974; Fodor, 1975; Newell and Simon, 1976; Newell, 1980, 1990; Pylyshyn, 1984; Dietrich, 1990). But this view faces a formidable problem, famously articulated by Harnad (1990) as: "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?" (p. 335). Harnad calls this the *symbol*

Interaction Studies 8:1 (2007), 7–30.
ISSN 1572–0373 / E-ISSN 1572–0381 © John Benjamins Publishing Company

grounding problem (SGP) and comments: "The handicap has been noticed in various forms since the advent of computing" (p. 338). The earliest reference that we know is that of Mays (1951), who writes "if we grant that these machines [*i.e., digital computers*] are complex pieces of symbolism,… to acquire a significance the symbols need to be linked with a set of referents" (p. 249). So if the computational metaphor is to offer any purchase in modelling and understanding cognition, the SGP poses a challenge that cannot be neglected (Cangelosi, Greco and Harnad, 2002). We take this challenge seriously, because the long-term goal of our research is to understand, via computer modelling and simulation, how speech sound categories (broadly, 'phonemes') could have emerged during language evolution, and then how these could be combined systematically to lead to utterances with semantic content.

To some the SGP is symptomatic of an incorrect view of AI and cognitive science, famously parodied as "good old-fashioned AI," or GOFAI, by Haugeland (1985). For instance, as Pfeifer and Scheirer (1999, p. 71) write, "… the symbol grounding problem is really an artifact of symbolic systems and 'disappears' if a different approach is used." The different approach they have in mind is, of course, *embodied* or *nouvelle* AI as spearheaded by Brooks (1990, 1991, 1999), which seeks to replace the central role played by symbolic representation with nonsymbolic interfacing to the physical world through cycles of perception and action, usually conceived as based on some connectionist or statistical machine learning principles. However, the complete banishment of symbolism from the scene is rather too radical for most AI scientists and cognitive psychologists, who continue to see a role for formal symbol systems, albeit in combination with some sort of connectionist component (e.g., Minsky, 1990; Harnad, 1990, 1993) in modelling and explaining the higher cognitive functions involved in, for example, using language, doing mathematics, and decision making under uncertainty, where nouvelle AI has arguably promised more than it has delivered.

Against this background, a new view of the SGP has recently arisen in which the physics of the external world plays an important and simplifying role (Sun, 2000; Vogt, 2002). Vogt (2002) coins the term *physical symbol grounding problem* and writes: "It is based on the idea that symbols should be grounded (cf. Harnad, 1990) and… they should be grounded by physical agents that interact with the world (cf. Brooks, 1990)" (p. 435). Our work is broadly consonant with this view, treating the SGP (as does Vogt, 2002) as a technical problem by way of computer simulation, although we have also been influenced in our thinking by the work of Barsalou (1999).

Quite apart from the intrinsic scientific interest in studying the emergence of human speech and language for its own sake (Damper, 2000), it makes an excellent context in which to consider the SGP. First and foremost, we believe human

© 2007. John Benjamins Publishing Company
All rights reserved

communication to be the clearest, certainly best-developed, example of externally-grounded cognition. As Vogt (2002, p. 431) writes, "language through its conventions offers a basis for invariant labeling of the real world." Since human communication is a social phenomenon, we pursue an approach of multi-agent simulation, not unlike much previous work in 'language games' but with one important difference (see below).

In particular in this paper, we argue that the emergence of speech sound categories 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. Important previous work along these lines is that of Steels (1997, 1998, 1999, 2003), de Boer (2000, 2001, 2005), and Oudeyer (2005a, 2005b, 2005c), who have explored grounded speech-category formation by computer simulation of multi-agent systems, with agents equipped with rudimentary articulatory and auditory systems and associated 'neural' processing. Broadly speaking, this line of work had its beginnings in the early and influential efforts of Lindblom (1986) and his colleagues to explain the origins of vowel systems in the world's languages (Liljencrantz and Lindblom, 1972; Lindblom, MacNeilage and Studdert-Kennedy, 1984; Lindblom, 1986, 2000) based on "adaptive dispersion theory." In their numerical simulations, the clustering of vowels in some metric space was predicted by minimising an energy function designed to reflect perceptual distinctiveness. An important question is exactly how realistic the simulations have to be (e.g., in terms of faithfully modelling the articulatory/auditory systems and brain mechanisms). Hence, our longer-term goal is to answer this question, although at this stage we will restrict ourselves to relatively simple simulations such as have been used in previous work.

Although Steels (1997) argues for a "limited rationality constraint" in multi-agent simulations (i.e., agents should not have access to each other's internal states), this constraint is typically violated in language games where nonlinguistic feedback figures importantly. For instance, de Boer (2001) writes, "the initiator then communicates the success or failure to the imitator using nonlinguistic communication" (p. 52). In our view, this amounts to a form of 'mind-reading,' seriously undermining the credibility of the simulations. Hence, we wish to avoid this aspect of language games, and favour Oudeyer's alternative approach where he dispenses with nonlinguistic feedback. As he writes, "it is crucial to note that agents *do not* imitate each other… The only consequence of hearing a vocalization is that it increases the probability, for the agent that hears it, of vocalizations… similar to those of the heard vocalization" (Oudeyer, p. 443). In spite of the absence of structured, coordinated interactions between agents, he achieves two results in his simulations which mirror important aspects of real language: "on the one hand discreteness and compositionality arise thanks to the coupling between perception

© 2007. John Benjamins Publishing Company
All rights reserved

and production within agents, on the other hand shared systems of phonemic categories arise thanks to the coupling across agents" (Oudeyer, p. 445).

A related line of investigation is that of Kirby (2001) and Kirby and Hurford (2002) who describe the iterated learning model (ILM). This, however, operates at the syntactic level, that is, learning agents receive from adult agents "meaning-signal pairs" (p. 103) that act as training data. Thus, the ILM already tacitly assumes the emergence of phonetic distinctiveness. Whereas the language-game style of simulations are concerned with language change once the basic mechanisms are in place, by contrast, Oudeyer is concerned with the earliest origins of a phonemic sound system, as are we. Further, Oudeyer's model is based on horizontal cultural interaction between agents of the same generation, following the works of Steels and colleagues, whereas the ILM is based on iterated learning among agents of one generation and agents of the previous generation (so this is more vertical learning).

However, Oudeyer's work has its own drawback in that he ignores the tenets of dispersion theory. "There are no internal forces which act as a pressure to have a repertoire of different discrete sounds," he writes (p. 443). But to cite de Boer (2001, p. 61), a successful vowel system has "its vowel clusters… dispersed (for low energy) and compact (for high imitative success)." These ideas are broadly consistent with notions of H&H theory (Lindblom, 1990) and the dispersion-focalisation theory (DFT) of Schwartz et al. (1997). Although Oudeyer tries to argue that the lack of a dispersion force is a virtue of his simulations (it is one less assumption), he also seems to recognise that it causes problems for the emergence of sound systems with realistically large numbers of vowels, writing, "Functional pressure to develop efficient communication systems might be necessary here" (p. 447).

Accordingly, the principal purpose of the present paper is to introduce ideas of H&H theory and DFT into Oudeyer-style simulations in the belief that more realistic vowel systems (i.e., more representative of those seen in a variety of human languages) will result. We will do this by extending the topological spaces in the neural maps used to couple auditory and articulatory processing as a vastly-simplified form of brain. We call these extensions *contour spaces*. The work is intended to form a baseline for future work in which we will study the impact of increased realism of the agents' articulatory and auditory capabilities, as well as extending our simulations beyond prediction of static vowel systems to the emergence of connected speech sounds with appropriate consonant-vowel patterning.

The remainder of this paper is structured as follows. In the next section, we set out our conception of physical symbol grounding, which we call *signal grounding*, and relate this to more traditional views of symbol grounding. Then, as a baseline for later discussion of our own work, we briefly describe Oudeyer's simulations of the emergence of vowel systems shared between a population of agents. We then

© 2007. John Benjamins Publishing Company
All rights reserved

introduce our extension to these simulations in the form of contour spaces and illustrate the beneficial effects of this extension in terms of emergence of more realistic vowel systems. Finally, we discuss the implication of these findings and conclude by arguing for the use of more realistic articulatory/auditory modelling as necessary to move beyond production of static vowel systems and account for the dynamic consonant-vowel patterning of speech.

Signal and Symbol Grounding

Before proceeding, it is necessary to discuss our relatively wide view of 'symbol grounding' and how it relates to the traditional, rather-narrower symbol grounding paradigm. Traditionally, the SGP has been seen as the problem of linking an internal symbolic representation like *cat* to the external (distal) object 'cat'. For instance, Figure 1 (reproduced from the influential text of Pfeifer and Scheirer, 1999) depicts a scenario linking the symbol *cup* with its external referent 'cup'. But this traditional view already assumes the existence of some sort of internal representation, which is more or less symbolic (or at least compositional). In our view, any solution to the SGP must also explain how this internal representation gets composed from elementary parts, which we take to be close to the notion of 'icons' in the terminology of Harnad (1990) or 'perceptual symbols' in the terminology of Barsalou (1999). Because these elementary parts result from sensory–motor

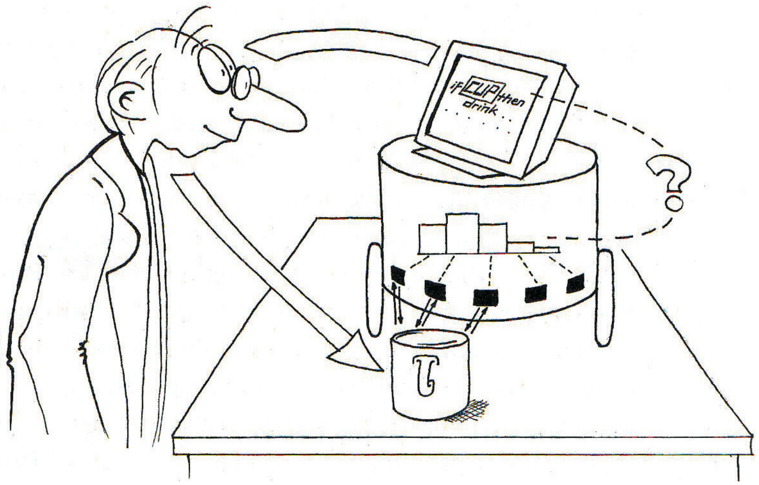


Figure 1. The traditional view of symbol grounding links an *a priori* internal representation (*cup*) to its external referent cup. Reproduced from Figure 3.4, p. 70 of Pfeifer and Scheirer (1999).

© 2007. John Benjamins Publishing Company
All rights reserved

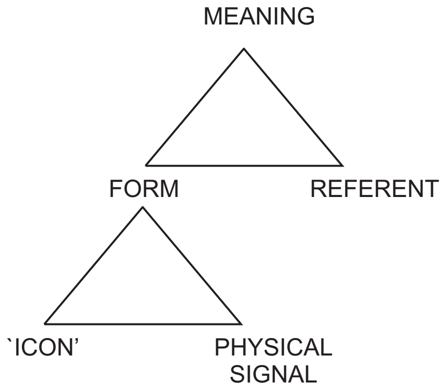


Figure 2. (a) The 'semiotic triangle,' reproduced from Figure 1, p. 433 of Vogt (2002). (b) A more complete picture of symbol grounding in which the form in (a) is grounded by interaction with the physical signal.

interaction, we cannot ignore the physics of the world. This leads us to the idea of signal grounding.

Symbol grounding is often discussed in the context of the semiotic triangle as in Figure 2(a), reproduced from Vogt (2002). But as just stated, we believe this picture to be incomplete, since the form is itself symbolic and ungrounded. A more complete view is depicted in Figure 2(b), where interaction with the physical world now grounds the form. In the case of interest here, this interaction is with the speech signal, hence the term 'signal grounding,' which can be seen either as a component part of symbol grounding, or as a specific instance of the SGP, albeit at a lower level than is usually considered. However it is viewed, we believe signal grounding is an indispensable part of symbol grounding.

For example, consider Figure 3. In this particular case of signal grounding, the distal object takes the form of an acoustic speech signal, produced by a vocal tract and perceived through the ear of a listener, linked to an arbitrary and iconic phoneme token (e.g., /æ/ using the notation of the International Phonetic Association, 1999). The form *cat* (or, equivalently, /kæt/) is then composed in a way that is systematic, but nonetheless arbitrary, from these phonemic primitives. Signal grounding then presents numerous challenges when considering the practicalities of forming an equivalence class for the phoneme /æ/. We need to map a wide range of varied signals onto the same phoneme symbol; the system needs to adapt to linguistic change over time; and the grounding of these arbitrary tokens needs to be shared among a population of speakers. These challenges will be taken up in the remainder of the paper.

To conclude this section, we remark that the ideas of signal and symbol grounding developed here are strongly related to notions of *double articulation*,

© 2007. John Benjamins Publishing Company
All rights reserved

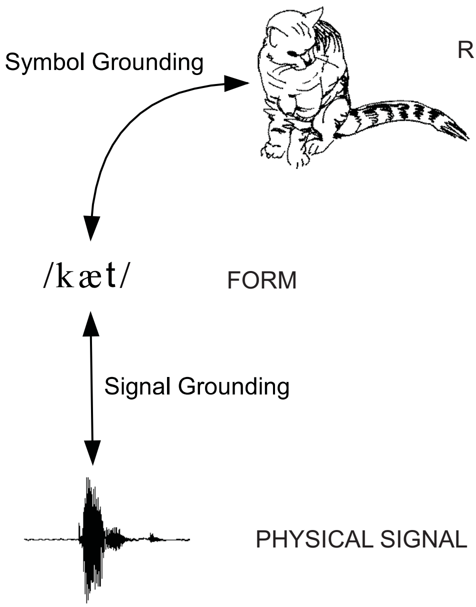


Figure 3. Illustration of signal grounding as a sub-problem of symbol grounding.

stemming from the work of de Saussure (1983), which views a linguistic system as a series of differences of sound combined with a series of differences of ideas. At the level of the first articulation, meaningful units (morphemes, words) are combined syntactically to convey ideas. At the level of the second articulation, primitive or elementary sound units (phonemes) are combined to form the meaningful units of the first articulation. The level of the second articulation is vital to human language as a fully productive system, because it is the key (loosely quoting Wilhelm von Humboldt) to achieving infinite generativity from finite machinery. Yet this is the level that is typically ignored by the traditional view of the SGP as characterised in Figure 1.

Basic agent architecture and its operation

The kind of signal grounding just described, and argued to be fundamental to human speech and language as a fully generative system, is a feature of the multi-agent simulation work of Oudeyer. We will take his work as the basis for extensions aimed at producing more realistic sound systems, by defining a *contour space* which acts as an objective function embodying measures of both articulatory effort and phonetic distinctiveness, broadly in line with both H&H theory (Lindblom, 1990) and dispersion-focalisation theory (Schwartz et al., 1997).

© 2007. John Benjamins Publishing Company
All rights reserved

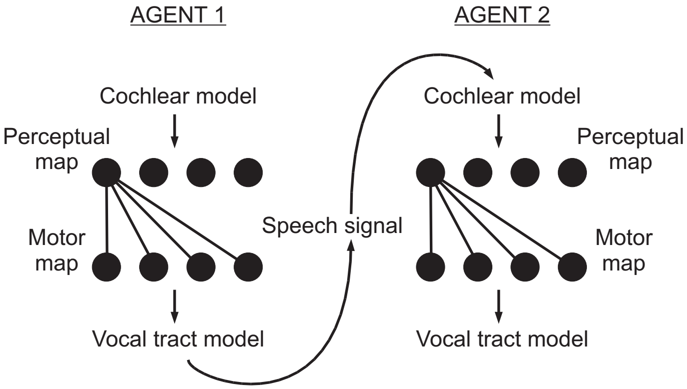


Figure 4. Architecture of the communicating multi-agent system, illustrated here for two agents. Redrawn from Figure 2, p. 439 of Oudeyer (2005c).

Figure 4 shows the basic agent architecture as used by Oudeyer and in this work. Each agent has an artificial ear (cochlear model), an artificial vocal tract, and in Oudeyer's words, an artificial 'brain.' Following Guenter and Gjaja (1996), the 'brain' features two coupled self-organising maps (SOMs, see Kohonen, 1990) — a perceptual map taking input from the auditory system and a motor map driving the articulatory system. Each agent perceives sounds produced by other agents as well as by itself. The appendix sets out details of the cochlear, vocal tract and neural models used by Oudeyer, and in our replications of his work. Note that we have used the "realistic" nonlinear articulatory/acoustic mapping (Oudeyer's Section 6.2) rather than the "abstract" linear mapping (Oudeyer's Section 6.1) throughout.

Our simulations use 10 agents (as compared to the 20 used by Oudeyer). But as he says of the number of agents, "This is a noncritical parameter of the simulations since nothing changes when we tune this parameter, except the speed of convergence of the system" (p. 443). Each 'speaking' agent is 'heard' by just one 'listening' agent (as shown in Figure 4) picked at random. Oudeyer states that "nothing changes" (p. 443) if a speaking agent is heard by more than one listener.

Initially, each agent produces utterances as dictated by its randomly-initialised 'brain' and also perceives the utterances of others. This, over some iterations, causes its SOMs to move from an unstable random configuration to a stable, converged, state of equilibrium. This process of convergence is driven by positive feedback (the basic self-organisation mechanism of the SOM), as each agent becomes increasingly likely to repeat the utterances that it has heard. Eventually, each SOM becomes partitioned into a variable number of basins of attraction as the nodes cluster around points of stability — determined by the utterances of the whole population. Any utterance which falls within the range of one of these basins of

© 2007. John Benjamins Publishing Company
All rights reserved

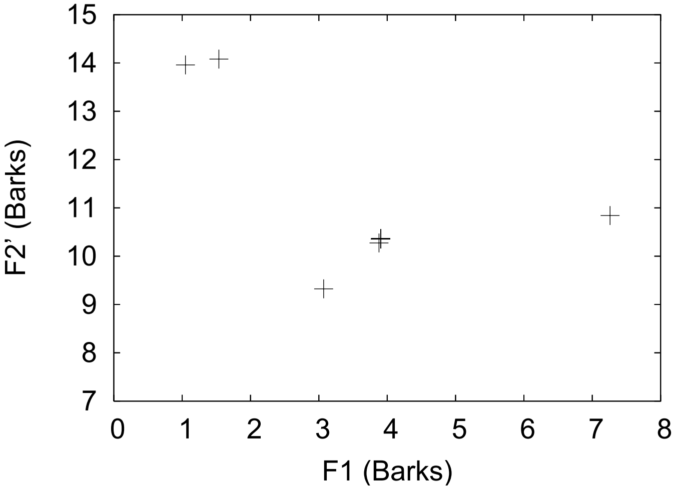


Figure 5. Convergence of Oudeyer's model to a five-vowel system with 10 agents, σ=0.05 and 2,000 iterations. Each cross represents a vector in auditory space; multiple vectors in the same region of space represent an equivalence class, or vowel. For a given equivalence class, individual vectors frequently overlay, giving the appearance of a single cross.

attraction is perceived by strong activation of the nodes around the centre point, so classifying a wide range of utterances.

The width of each SOM's gaussian function (σ in equation (5) of Appendix 3) determines the size of the basin of attraction and, therefore, in the case of the auditory map, the variety of stimuli perceived as the 'same' utterance. In Oudeyer's simulations, there is no dispersive force and, thus, as σ increases, convergence is to a single point. To quote (Oudeyer, p. 445), "if two neuron clusters… get too close, then the summation of tuning functions in the iterative process of coding/decoding smooths their distribution locally and only one attractor appears." This is not realistic behaviour within a language. However, it is clear that, with the right parameter settings, it is perfectly possible to cause the emergence of a feasible, shared, multi-vowel system. See for instance Figure 5, which depicts a typical result from our replication of Oudeyer's simulation. Here, 500 points initially distributed randomly in *F*1-*F*2 space have converged to just five clusters. In fact, in the absence of a dispersive force, the 'clusters' have actually converged (almost) to overlay at the centre of their respective basin of attraction. In the remainder of this paper, we will introduce a dispersive force and study its effect on convergence to linguistically-realistic vowel systems.

© 2007. John Benjamins Publishing Company
All rights reserved

Contour spaces

In this section, we introduce basic ideas of H&H theory (Lindblom, 1990) and DFT (Schwartz et al., 1997) into our simulations. According to H&H theory, speakers "tune their performance according to communicative and situation demands… to vary their output along a continuum of *hyper*- and *hypospeech*" (Lindblom, 1990, p. 403). That is, in difficult communication conditions, speakers hyper-articulate in order to be understood, even though this requires additional energy be expended. In less demanding situations, energy can be conserved by hypo-articulation, always provided communication success is maintained. The 'setting' on the hyper-/hypo- continuum is determined by an on-line process in which the speaker continuously infers success of communication by monitoring linguistic and paralinguistic feedback from the listener. We assume that similar forces are at work in the process of vowel formation among a collection of communicating agents; that is, there is not only a drive towards distinctive sound categories (loosely corresponding to 'hyper'), but also an inbuilt desire to minimise energy expended by the agent (loosely corresponding to 'hypo').

Similar ideas are embodied in dispersion-focalisation theory, which encompasses more or less the same principles as H&H theory, but formulated in the auditory (rather than articulatory) domain. This theory seeks to explain the formation of vowel inventories not so much in terms of energy expended by a speaker as via competing forces of "global dispersion based on inter-vowel distances; and local focalization, which is based on intra-vowel spectral salience" (Schwartz et al., 1997, p. 255). The dispersive force thus seeks to maintain distinctiveness between sound categories. The focalisation force in DFT is a little harder to visualise and justify. Is is based on the 'compactness' of formant frequencies, formants being the resonant frequencies of the vocal tract that correspond to "concentration of acoustic energy, reflecting the way that air from the lungs vibrates in the vocal tract, as it changes its shape" (Crystal, 1980, p. 150). These concentrations of energy are reflected in peaks in the frequency spectrum; the one occurring at the lowest frequency is called the first formant, $F1$; that occurring at the next highest frequency is called the second formant, $F2$, and so on.

In the words of Schwartz et al. (1997) (note the minor difference in notation for formant frequencies):

> "a discrimination experiment involving stimuli with various F_2-F_3-F_4 patterns… demonstrated that patterns with the greatest formant convergence (namely with F_3 close to either F_2 or F_4) were more stable in auditory memory… while patterns with less convergence, namely with F_3 at an equal distance from both F_2 and F_4, were more difficult to memorize (Schwartz and Escudier, 1989)." (p. 259)

© 2007. John Benjamins Publishing Company
All rights reserved

Schwartz et al. (1997) further note, "the perceptual demonstration that formant convergence in the F_2-F_3-F_4 pattern produced more stable patterns in discrimination experiments, led us to propose that formant convergence could result in an increased 'perceptual value'… because of 'acoustic salience'" (p. 259). Hence, the focalisation force is designed to favour vowels in which the formants are close together in frequency.

Introducing Dispersive Forces

In the long term, we are seeking to minimise the articulatory effort of an utterance, at the same time maximising its perceptual distinctiveness to other agents. At this stage, however, we have no direct way to quantify articulatory effort; hence, we address the problem by using the established ideas of dispersion-focalisation theory (working in the auditory domain as opposed to the articulatory domain), as just discussed. 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. When the topological space of our self-organising maps is augmented with dispersion based on inter-vowel differences (in addition to focalisation based on intra-vowel attraction), we refer to it as a *contour space*. By introducing the proposed contour spaces, we hope to achieve a greater robustness to parameter variation and a greater level of realism in the vowel systems that are produced.

We now describe how a repulsive force acting on the perceptual neurons of the agent is introduced. For each node i of the auditory map, at time t, we define an energy functional given by

$$E\left(v_i(t), v_j(t)\right) = \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{1}{d_{ij}^2} \tag{1}$$

$$\text{where} \qquad d_{ij} = \sqrt{\left(F1_i - F1_j\right)^2 + \left(F2'_i - F2'_j\right)^2}$$

In equation (1), j is an index over all N nodes in the auditory map, $v_i = (F1_i, F2_i')$ and similarly $v_j = (F1_j, F2_j')$. (Appendix 2 for discussion of $F2'$.) This amounts to a measure of distance between the i and j vowels in the $F1$-$F2'$ auditory-map space.

Updating occurs as follows. At time t, for each neuron i in the auditory space, we generate 8 'test positions' around that neuron. These are spaced on a rectangular grid of side σ centred on i. The update equation is:

$$v_{\mathrm{i}}(t+1) = v_{\mathrm{i}}(t) + \gamma v_{\mathrm{max}} \tag{2}$$

© 2007. John Benjamins Publishing Company
All rights reserved

where $\boldsymbol{v}_{\mathrm{max}}$ is the $\boldsymbol{v}_{\mathrm{k}}(t)$ vector for which the energy $E(\boldsymbol{v}_{\mathrm{i}}(t),\boldsymbol{v}_{\mathrm{k}}(t))$ is maximised, with k being an index over the 8 neighbours of $\boldsymbol{v}_{\mathrm{i}}(t)$, and γ is a step size or learning rate. Thus, maximisation is performed by gradient ascent. In this way, we are moving the ith vowel in the direction that maximises the acoustic distinctiveness between it and all other vowels in the space.

Attractive Force: Focalisation

The articulatory space is three-dimensional, defined in terms of lip rounding r, tongue position p and tongue height h. As previously discussed, focalisation in our model follows Schwartz et al. (1997) in seeking to favour vowels with compact $F2$-$F3$-$F4$ formant patterns by defining and minimising an energy functional.

The specific energy functional used is similar to that of Schwartz et al. (1997) (see their equations (4) to (7)) modified to fit our simulations using a self-organising map:

$$E(\boldsymbol{v}_l(t) = (r_l, p_l, h_l)) = E_{12} + E_{23} + E_{34} \tag{3}$$

$$\text{where} \qquad \begin{aligned} E_{12} &= -(\frac{1}{(F2_l - F1_l)^2}) \\ E_{23} &= -(\frac{1}{(F3_l - F2_l)^2}) \\ E_{34} &= -(\frac{1}{(F4_l - F3_l)^2}) \end{aligned}$$

In (3), each neuron l has its associated (r_l, p_l, h_l) values, which allow computation of formant values via the vocal tract model (Appendix 1). At time t, each such neuron has its vector $\boldsymbol{v}_{\mathrm{l}}(t)$ updated according to:

$$\boldsymbol{v}_{\mathrm{l}}(t+1) = \boldsymbol{v}_{\mathrm{l}}(t) + \gamma \boldsymbol{v}_{\mathrm{min}} \tag{4}$$

where $\boldsymbol{v}_{\mathrm{min}}$ is the $\boldsymbol{v}_{\mathrm{m}}(t)$ vector for which $E(\boldsymbol{v}_{\mathrm{m}}(t))$ is minimised, m is an index over the 26 neighbours of $\boldsymbol{v}_{\mathrm{l}}(t)$ (on a grid of size σ in 3-D space), and γ is a step size or learning rate. Hence, we are minimising by gradient descent.

Note that although this mechanism of attraction is firmly based in perception, we are in fact minimising in (r, p, h) space. Hence, we view this as, effectively, a mechanism for reducing (if not actually minimising) articulatory effort in line with H&H theory.

© 2007. John Benjamins Publishing Company
All rights reserved

Results of simulations

In this section, we first show some typical illustrative results obtained using Oudeyer's model to act as a benchmark before presenting typical results from the new model based on DFT. Thereafter, more thorough results (averaged over 500 runs) are given comparing the sensitivity of the two models to variation in the gaussian width parameter, σ. The two models are also compared with respect to the emergence of realistic vowel systems (i.e., their similarity to those observed in human languages). In all simulations, the nodes of the self-organising maps are initially randomised, that is, placed at uniformly-distributed positions in the appropriate space.

In these simulations, the optimisation step size, γ of equations (2) and (4), is set equal to the gaussian width, σ of equation (5) in Appendix 3, enabling all three forces (i.e., dispersion, focalisation, self-organisation) to maintain their intended, relative level of influence. The gaussian width in the auditory space was scaled up to take account of the different range of the two maps ($[0,1]^3$ for the motor map and [0..8 Bark, 0..15 Bark] for the auditory map). All SOMs have 500 nodes, and simulations are stopped after 2,000 iterations of two-agent interaction. This stopping criterion was decided after examining how auditory dispersion (measured from the energy functional of eqn. (1)) varied during a few trials of the simulation. Figure 6 depicts a typical example. Although dispersion does not reduce monotonically, convergence is achieved well before 2,000 iterations.

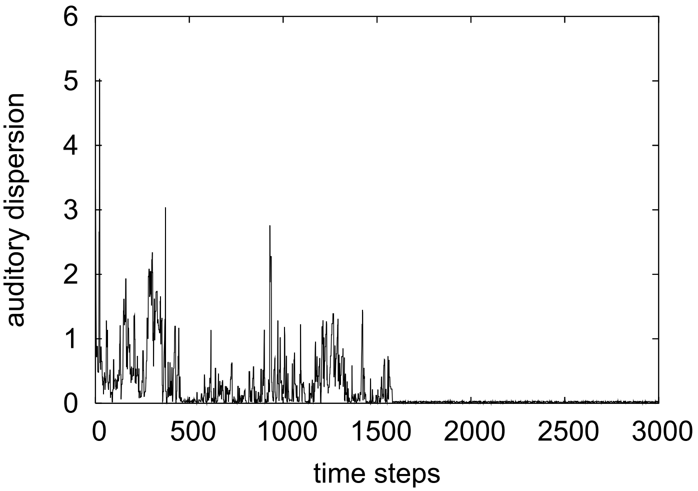
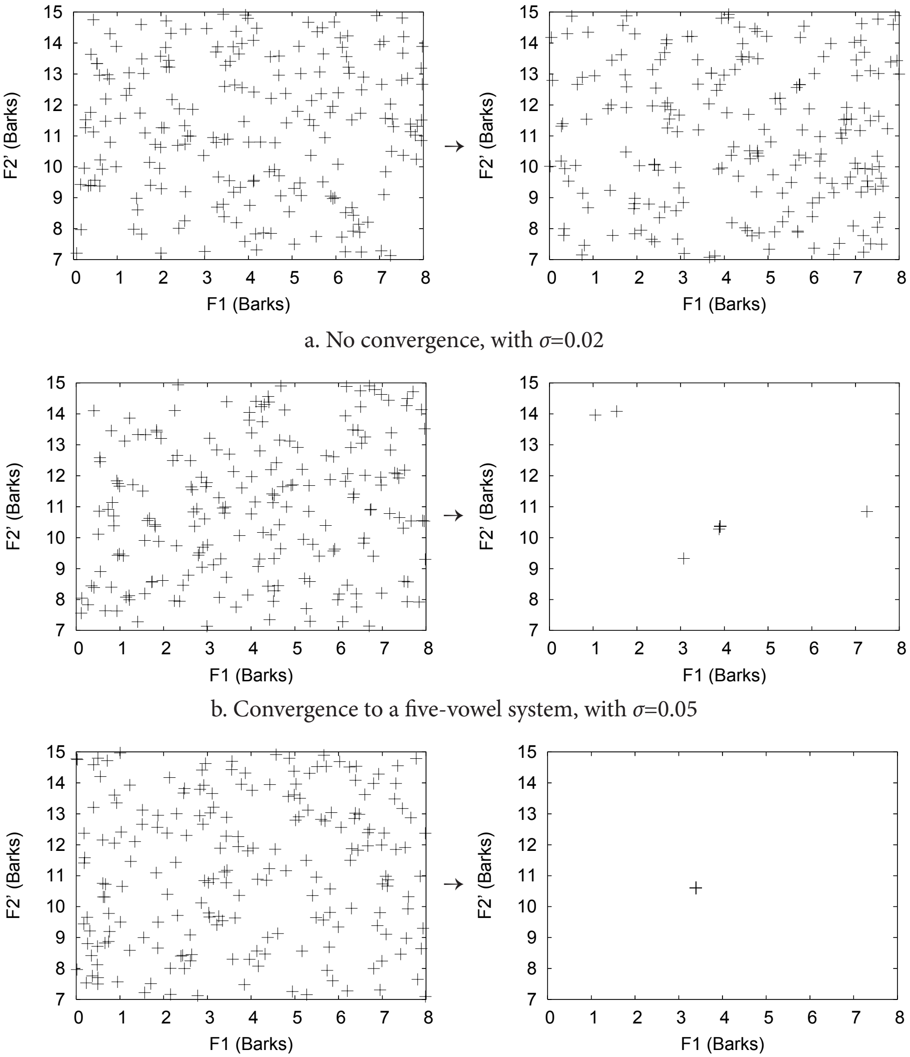


Figure 6. Typical plot of auditory dispersion versus number of iterations, showing convergence well before 2,000 steps.

© 2007. John Benjamins Publishing Company
All rights reserved

Reproduction of Oudeyer's Results

We have already shown an example of how the initial model can converge to a reasonable five-vowel system with $\sigma = 0.05$ (Figure 5 earlier). We have also detailed how, as σ increases, there is a strong tendency to converge to a single point. Figure 7 shows a composite of typical results as σ varies. It is seen that realistic vowel systems emerge only for a restricted range of σ values.



a. No convergence, with σ=0.02

b. Convergence to a five-vowel system, with σ=0.05

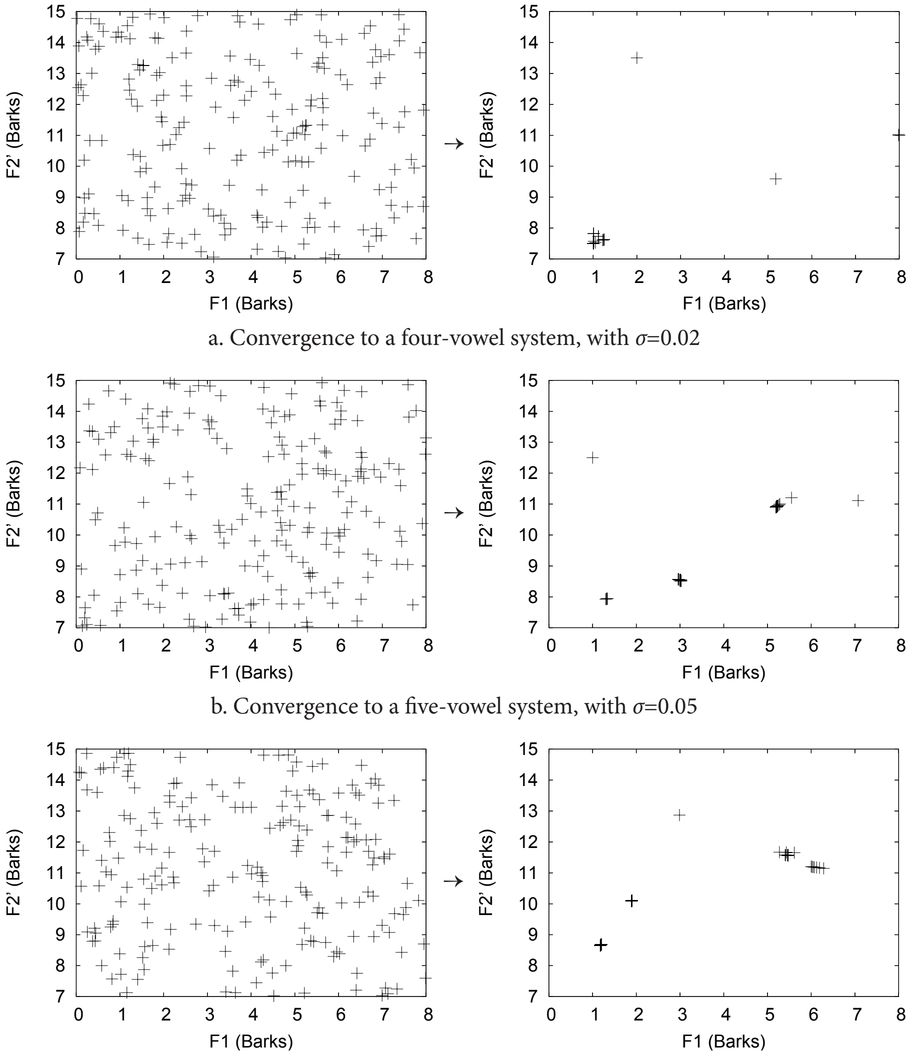
c. Convergence to a single point, with σ=0.1

Figure 7. Composite of typical results from our replication of Oudeyer's simulation as σ varies.

© 2007. John Benjamins Publishing Company
All rights reserved

Effect of the Contour Space

Figure 8 shows a composite of typical results from simulations of the new model with contour spaces with the same σ values as in Figure 7. As can be clearly seen, realistic vowel systems emerge over a much wider range of σ values. There is also,



a. Convergence to a four-vowel system, with σ=0.02

b. Convergence to a five-vowel system, with σ=0.05

c. Convergence to a five-vowel system, with σ=0.1

Figure 8. Composite of typical results from simulations of the new model with contour spaces with the same σ values as in Figure 7. Realistic vowel systems emerge over a much wider range of σ values.

© 2007. John Benjamins Publishing Company
All rights reserved

we think, less tendency for the converged points to overlay exactly than in the original work (i.e., there is more of a 'cluster').

Further comparison of the two systems

To test further the assertion that the new system featuring dispersive forces (i.e., contour spaces) will possess a greater robustness to parameter variation than Oudeyer's original, 500 repeated runs were made for different values of the gaussian width σ. The number of vowels present after convergence was then recorded for both systems. If convergence did not occur, results were discarded. Figure 9 shows the results averaged over the 500 runs; the error bars depict the standard deviation.

For the new system, a high level of variation in the number of vowels observed at convergence is seen across the whole range of σ values. We take this to be a positive feature of the new system, since human languages display a wide variety of vowel inventories (Maddieson, 1984; Ladefoged and Maddieson, 1996). By contrast, the Oudeyer system (as replicated by us) shows unrealistic convergence to a single 'vowel' with zero variability for $\sigma > 0.07$ and a total lack of convergence (to a sensibly small number of clusters) for $\sigma < 0.05$. Realistic convergence is maintained for the new system up to parameter values of 0.15. No simulations were performed for $\sigma > 0.15$.

Following Oudeyer (2005c, Figure 10, p. 446), we have also compared the two systems with data for human languages, taking vowel frequencies from Ladefoged

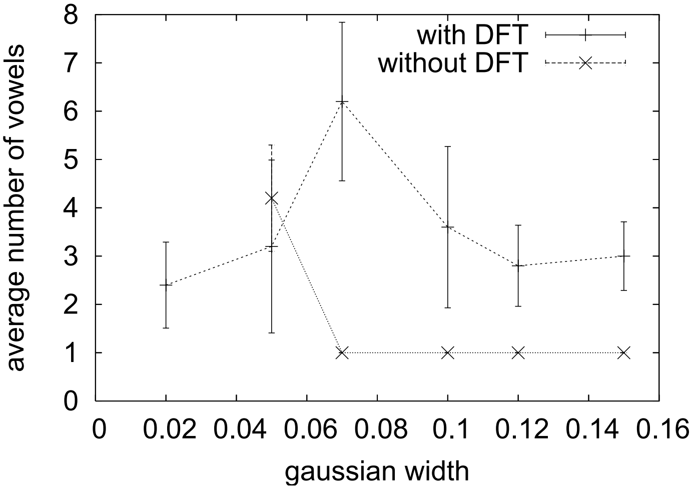


Figure 9. Comparison of our replication of Oudeyer's simulation with the new model based on DFT, illustrating the robustness to parameter variation resulting from inclusion of a dispersive force. Error bars are standard deviations over 500 runs.

© 2007. John Benjamins Publishing Company
All rights reserved

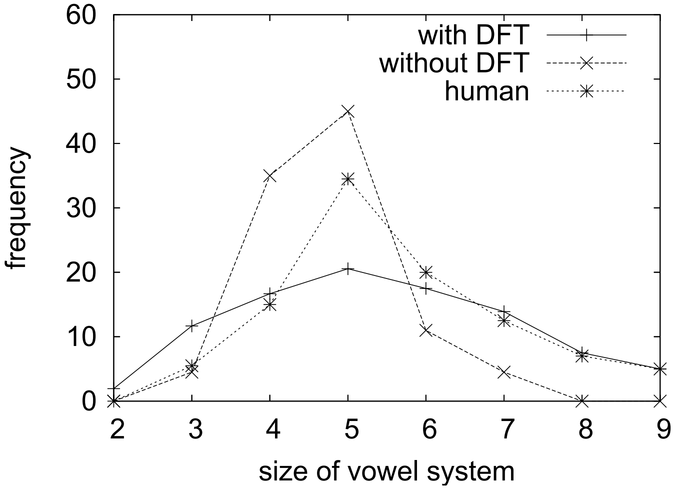


Figure 10. Comparison of vowel systems observed in human languages and those produced by computer simulation with and without DFT (i.e., with and without dispersive forces).

and Maddieson (1996). For the new computer model, σ was set to 0.05 and 500 simulations were run. Comparative data for Oudeyer's system for the same value of σ and number of iterations were taken from his original paper, rather than the simulations being replicated here. Figure 10 shows the comparison, which reveals that the system with contour spaces has a slight preference for simpler vowel systems but is able to capture the emergence of the more complex systems, which is a problem for Oudeyer. Quantitatively, the mean square error (MSE) between the curve for Oudeyer's data (labelled "without DFT") and the human data is 91.28, whereas the corresponding MSE for our simulations (labelled "with DFT") is 29.94. All three systems share a peak of five vowels. We emphasise that this comparison is made under conditions (namely, σ set at 0.05) which are maximally favourable to Oudeyer's model. This is necessary because of the sensitivity of his model to the setting of σ.

Discussion and Conclusions

The tension introduced by the addition of a dispersive force has clearly had a beneficial effect. This extension achieves an increased level of robustness to parameter variation and captures the emergence of some of the more complex vowel systems observed in human languages, in a way which Oudeyer was unable to do. Despite a slight preference for the simpler vowel systems, the distribution is

© 2007. John Benjamins Publishing Company
All rights reserved

more representative of that seen in real languages, as confirmed by the much lower mean square error (see previous section).

How have these beneficial effects come about? Boë, Schwartz and Vallée (1994) have already shown, although not in a multi-agent setting, how DFT can produce a range of vowel systems. (Rather, starting with a full set of vowel 'prototypes,' they show how DFT can be used to select realistic subsets typical of different languages.) In the present setting, the three forces of dispersion, focalisation and self-organisation act to produce convergence to attractors in the contour space. These attractors correspond to a physical grounding of the speech signals produced by the agents, as in Figure 2(b). The gradual, progressive nature of the convergence, over many interactions, ensures the final set of signal-grounded forms is shared among the population. So the physics governing a population not only potentially accounts for a wide variety of human vowel systems but also allows for this set to become established within a population.

In our work, grounding of the external world is via these attractors in contour space. So, rather than connecting an arbitrary *a priori* abstraction (as when *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 of the situation. Through the formation 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 can answer some of the current criticisms of the symbol grounding paradigm (e.g., Lakoff, 1993), because the attractors capture the ambiguities and 'shades of grey' that challenge more traditional views of grounding (Davidsson, 1993). This has similarities to previous work which has sought to explain grounding using connectionist models (e.g., Harnad, 1993; Damper and Harnad, 2000; Cangelosi, Greco and Harnad, 2002). These have been successful in displaying various aspects of human cognition. But, by considering grounding at the (sub-form) level of physical signals (Figures 2(b) and 3), we have developed a new framework in which this interplay between symbol grounding and connectionist systems can be further explored.

Several possibilities for future work are under consideration. At present, agents do not exactly 'hear' sounds; rather, they have direct access to formant values. From $F1$, $F2$, $F3$ and $F4$ values specifying a vocalisation, they perceive $F1$ directly and compute a perceived $F2'$. This is a very high level of abstraction, implicitly making many assumptions (e.g., about the role of formants in speech perception, and how the auditory system can extract them from the speech signal). First and foremost, therefore, we wish to move to using actual sounds as the medium of interchange between agents. This move will make it necessary to use more physically realistic vocal tract and cochlear models. It is then a matter of some importance and interest to investigate how much increased realism/complexity impacts on

© 2007. John Benjamins Publishing Company
All rights reserved

the emergence of sound systems. We know from Oudeyer and the present work that very simple, highly abstract models are adequate for the production of shared (static) vowel systems, but under rather strong assumptions. Furthermore, speech sounds do not consist entirely of vowels, but of dynamic consonant-vowel patterns forming syllables. Unfortunately, although there is general agreement among phoneticians and speech scientists that vowels can be reasonably well specified by formant values, there is no corresponding understanding of how consonant sounds can be similarly specified and distinguished.

Although Oudeyer (2005b) has extended his "abstract" linear model in the direction of "the formation of… and patterns of sound combination" (p. 328), this is done without any acoustic, perceptual space, but with agents given direct access to the relevant parameters in what we believe to be an unsatisfactory ('mind-reading') manner. By moving to simulations in which actual, physical speech sounds are exchanged between agents, we can hope to explore the emergence of speech as a dynamic phenomenon in a more realistic and satisfactory way.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–609.

Boë, L.-J., Schwartz, J.-L., and Vallée, N. (1994). The prediction of vowel systems: Perceptual contrast and stability. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (pp. 185–213). Chichester, UK: John Wiley.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*(1), 3–15.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*(1–3), 139–159.

Brooks, R. A. (1999). *Cambrian intelligence*. Cambridge, MA: Bradford Books/MIT Press.

Cangelosi, A., Greco, A., and Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi and D. Parisi (Eds.), *Simulating the evolution of language* (pp. 191–210). London, UK: Springer-Verlag.

Carlson, R., Granström, B., and Fant, G. (1970). Some studies concerning perception of isolated vowels. *STL-QPSR*, *2–3*, 19–35.

Chistovich, L. A., and Lublinskaya, V. V. (1979). The 'centre of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustic study of the perception of vowel-like stimuli. *Hearing Research*, *1*(3), 185–195.

Crystal, D. (1980). *A first dictionary of linguistics and phonetics*. London: André Deutsch.

Damper, R. I. (2000). Emergence and levels of abstraction. *International Journal of Systems Science*, *31*(7), 811–818.

Damper, R. I., and Harnad, S. R. (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?* (pp. 157–161). Raleigh, NC.

© 2007. John Benjamins Publishing Company
All rights reserved

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

de Boer, B. (2001). *The origins of vowel systems*. Oxford, UK: Oxford University Press.

de Boer, B. (2005). Evolution of speech and its acquistion. *Adaptive Behavior*, *13*(4), 281–292.

de Saussure, F. (1983). *Course in general linguistics*. London, UK: Duckworth. (Translation of 1916 edition by R. Harris)

Dietrich, E. (1990). Computationalism. *Social Epistemology*, *4*(2), 135–154.

Fodor, J. (1975). *The language of thought*. New York, NY: Crowell.

Guenter, F. H., and Gjaja, M. N. (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.

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MA: Bradford Books/MIT Press.

International Phonetic Association. (1999). *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet*. Cambridge, UK: Cambridge University Press.

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.

Kirby, S., and Hurford, J. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121–148). London, UK: Springer-Verlag.

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

Ladefoged, P., and Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, UK: Blackwell Scientific Publishers.

Lakoff, G. (1993). Grounded concepts without symbols. In *Proceedings of the fifteenth annual meeting of the cognitive society* (pp. 161–164). Boulder, CO.

Liljencrantz, J., and Lindblom, B. (1972). Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839–862.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J. J. Ohala and J. J. Jaeger (Eds.), *Experimental phonology* (pp. 14–44). Orlando, FL: Academic Press.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of H&H theory. In W. J. Hardcastle and A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Lindblom, B. (2000). Developmental origins of adult phonology: The interplay between phonetic emergents and the evolutionary adaptation of sound patterns. *Phonetica*, *57*(2–4), 297–314.

Lindblom, B., MacNeilage, P., and Studdert-Kennedy, M. (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.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge, UK: Cambridge University Press.

Mays, W. (1951). The hypothesis of cybernetics. *British Journal for the Philosophy of Science*, *2*(7), 249–250.

Minsky, M. (1974). *A framework for representing knowledge* (Tech. Rep. Nos. AIM–306). Cambridge, MA: Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

© 2007. John Benjamins Publishing Company
All rights reserved

Minsky, M. (1990). Analogical vs. logical or symbolic vs. connectionist or neat vs. scruffy. In P. H. Winston (Ed.), *Artificial intelligence at mit: Epanding frontiers* (Vol. 1, pp. 219–243). Cambridge, MA: MIT Press.

Newell, A. (1973). Artificial intelligence and the concept of mind. In R. C. Shank and K. M. Colby (Eds.), *Computer models of thought and language* (pp. 1–60). San Francisco, CA: Freeman.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*(2), 135–183.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., and Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126.

Oudeyer, P.-Y. (2005a). How phonological structures can be culturally selected for learnability. *Adaptive Behavior*, *13*(4), 269–280.

Oudeyer, P.-Y. (2005b). The self-organization of combinatoriality and phonotactics in vocalization systems. *Connection Science*, *17*(3–4), 325–341.

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

Pfeifer, R., and Scheirer, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: Bradford Books/MIT Press.

Schwartz, J.-L., Boë, L.-J., Vallée, N., and Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics*, *25*(3), 255–286.

Schwartz, J.-L., and Escudier, P. (1989). A strong evidence for the existence of a large scale integrated spectral representation in vowel perception. *Speech Communication*, *8*(3), 235–259.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, *1*(1), 1–35.

Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, *103*(1–2), 133–156.

Steels, L. (1999). *The talking heads experiment. volume 1: Words and meanings*. Antwerpen, Belgium: Laboratorium.

Steels, L. (2003). Evolving grounded communication for robots. *Trends in Cognitive Science*, *7*(7), 308–312.

Sun, R. (2000). Symbol grounding: A new look at an old idea. *Philosophical Psychology*, *13*(2), 149–172.

Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, *88*(1), 97–100.

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, *3*(3), 429–457.

Authors' addresses

S. F. Worgan and R. I. Damper
Information: Signals, Images, Systems Research Group
School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK

Emails: {sw205r|rid}@ecs.soton.ac.uk

© 2007. John Benjamins Publishing Company
All rights reserved

About the authors

Simon F. Worgan obtained his BSc in artificial intelligence and computer science in 2004 and MSc in natural computation in 2005, both from the University of Birmingham. He is currently studying for his PhD in the origins of speech sound categories at the University of Southampton.

Robert I. Damper obtained his MSc in biophysics in 1973 and PhD in electrical engineering in 1979, both from the University of London. He was appointed Lecturer in electronics at the University of Southampton in 1980, Senior Lecturer in 1989, Reader in 1999 and Professor in 2003. His research interests include speech science and speech technology, signal and pattern processing, natural language engineering, neural computing, mobile robotics, biological cybernetics, and philosophical and historical aspects of artificial intelligence. Prof. Damper has published more than 280 research articles and authored the undergraduate text *Introduction to Discrete-Time Signals and Systems*. Prof. Damper serves on the editorial boards of the *International Journal of Speech Technology* and the *International Journal of Information Technology*.

Appendix: Oudeyer's Agent Model

In Oudeyer's work, each agent has an artificial vocal tract, an artificial ear (cochlear model), and an artificial 'brain,' or neural model. These will now be detailed in turn.

1. Vocal Tract Model

Following de Boer (2001), Oudeyer uses a vocal tract simulation controlled by three parameters, namely lip rounding r, tongue height h and tongue position p. Each parameter is constrained to reflect the anatomical range of the corresponding articulator movement. We can derive formant values as follows:

$$\begin{aligned}
F1 &= ((-392+392r)h^2+(596-668r)h+(-146+166r))p^2+((348-348r)h^2\\
&\quad +(-494+606r)h+(141-175r))p+((340-72r)h^2+(-796+108r)h\\
&\quad +(708-38r))\\
F2 &= ((-1200+1208r)h^2+(1320-1328r)h+(118-158r))p^2\\
&\quad +((1864-1488r)h^2+(-2644+1510r)h+(-561+221r))p\\
&\quad +((-670+490r)h^2+(1355-697r)h+(1517-117r))\\
F3 &= ((604-604r)h^2+(1038-1178r)h+(246+566r))p^2+((-1150+1262r)h^2\\
&\quad +(-1443+1313r)h+(-317-483r))p+((1130-836r)h^2\\
&\quad +(-315+44r)h+(2427-127r))\\
F4 &= ((-1120+16r)h^2+(1696-180r)h+(500+522r))p^2+((-140+240r)h^2\\
&\quad +(-578+214r)h+(-692-419r))p+((1480-602r)h^2\\
&\quad +(-1220+289r)h+(3678-178r))
\end{aligned}$$

Although it would be possible to produce sounds (i.e., synthetic vowels) exhibiting these formant values, which were then 'heard' by the 'speaker' and other agents, this is not done in Oudeyer's simulations or in ours. Rather, a short-cut is taken in which auditory parameters are calculated from the formant values.

© 2007. John Benjamins Publishing Company
All rights reserved

2. Cochlear Model

A cochlear (ear) model, designed by Boë, Schwartz and Vallée (1994), is employed to process the formant values, placing the result in a 2-D auditory space. The model perceives the first formant directly and derives an 'effective' second formant, $F2'$ (Carlson, Granström and Fant, 1970), as follows:

$$F2' = \begin{cases} F2 & \text{if } F3 - F2 > c \\ \frac{(2-w_1)F2+w_1F3}{2} & \text{if } F3 - F2 \leq c \text{ and } F4 - F2 \geq c \\ \frac{w_2F2+(2-w_2)F3}{2} - 1 & \text{if } F4 - F2 \leq c \text{ and } F3 - F2 \leq F4 - F3 \\ \frac{(2+w_2)F3-w_2F4}{2} - 1 & \text{if } F4 - F2 \leq c \text{ and } F3 - F2 \geq F4 - F3). \end{cases}$$

where c is as a constant of value 3.5 Bark (Chistovich and Lublinskaya, 1979), and w_1 and w_2 are defined as:

$$w_1 = \frac{c-(F3-F2)}{c}$$

$$w_2 = \frac{(F4-F3)-(F3-F2)}{F4-F2}$$

The above equations assume frequency is represented on the Bark scale. Conversion to this scale from hertz frequency is done using the following conversion formula (Traunmüller, 1990):

$$f_{Bark} = \frac{26.81}{1+1960/f_{Hz}} - 0.53$$

3. Neural Model

The neural model is based on two self-organising maps (Kohonen, 1990). The self-organising map (SOM) defining the articulatory space captures the configurations of the vocal tract in terms of parameters r, h and p. The auditory space codes for the range of acoustic cues in terms of the first formant $F1$ and second 'effective' formant $F2'$. Each agent's neural model is then established by forming weighted connections between the nodes of the auditory and articulatory spaces.

When activated, the jth node in the articulatory space produces a vector $v_j = (r_j, h_j, p_j)$ forming a point in $[0,1]^3$ space coding articulatory configuration. A sequence of these vectors, $v_1, v_2, \ldots, v_n$ where n is a random number between 2 and 4, is then fed to the vocal tract model. This produces an articulatory trajectory ('utterance') of from 2 to 4 configurations. All remaining neurons are then modified according to:

$$v_k(t+1) = v_k(t) + G_k(v_j)(v_j - v_k(t)) \quad \begin{cases} k = 1..N,\ k \neq j, \\ \text{where } N \text{ is the number of neurons in each map} \end{cases}$$

Each articulatory neuron is updated by a gaussian activation function:

© 2007. John Benjamins Publishing Company
All rights reserved

$$G_k(\boldsymbol{v}_j) \quad = \exp\left(\frac{d^2_{j,k}}{2\sigma^2}\right) \tag{5}$$

$$\text{where } d^2_{j,k} = |\boldsymbol{v}_j - \boldsymbol{v}_k|^2$$

This update mechanism causes the nodes to converge on points in the articulatory space. The location of these points of convergence is determined by the agent's choice of articulation and the utterances that it is exposed to. The articulatory space can then be modified by the auditory space through the weighted connections between the two. The connections between the perceptual neuron *i* and the articulatory neuron *j* are characterised by the weight $w_{i,j}$ (initially random).

The auditory space is able to achieve a similar convergence, since on perceiving an utterance a vector containing acoustic cues $\boldsymbol{s}$ (derived from the 'speech signal') is placed in the perceptual space and the neurons updated by:

$$\boldsymbol{v}_i(t+1) = \boldsymbol{v}_i(t) + G_i(\boldsymbol{s})(\boldsymbol{s} - \boldsymbol{v}_i(t))$$

The articulatory space is then further updated through the weighted connections by characterising $d^2_{j,k}$ as:

$$d^2_{j,k} = \sum_i^N w_{i,j} G_i(\boldsymbol{s})$$

Taking the function dependence of $G()$ on $\boldsymbol{s}$ as implicit, for simplicity, the weights are updated by a Hebbian learning rule:

$$\Delta w_{i,j} = \alpha(G_i - \langle G_i \rangle)(G_j - \langle G_j \rangle)$$

where α is set to some small random number and $\langle G_j \rangle$ represents the average gaussian activation over the previous time steps.

© 2007. John Benjamins Publishing Company
All rights reserved