

Modeling U Shaped Performance Curves in Ongoing Development

Anthony F. Morse (anthony.morse@plymouth.ac.uk)

Tony Belpaeme (tony.belpaeme@plymouth.ac.uk)

Angelo Cangelosi (angelo.cangelosi@plymouth.ac.uk)

Centre for Robotics and Neural Systems, University of Plymouth,
Devon, PL4 8AA, UK

Caroline Floccia (caroline.floccia@plymouth.ac.uk)

School of Psychology, University of Plymouth,
Devon, PL4 8AA, UK

Abstract

This paper details a simple and general account, and model, of the U-shaped curve phenomena apparent in many developmental psychology experiments. The model replicates both the general form of the U-shape performance in ongoing development and accounts for additional observations in the psychology literature such as the effect of noise in Switch task experiments. This leads to predictions both in psychology and neuroscience, and establishes an alternative hypothesis, which is simpler, more detailed, more predictive, and more general than those already established in the literature. This approach is also suitable for embodied robotic modeling of development.

Keywords: Cognitive modeling; neural networks; epigenetic robotics; language acquisition; development; U-shaped curve, Self-Organizing Maps; Active Hebbian learning.

Ongoing Development in Humans and Robots

This paper presents a novel neuro-computational approach to modeling cognitive development, in particular for the investigation of the phenomenon of U-shaped performance curves in development. The model is based on refinements of the associative learning mechanism recently proposed as part of the Epigenetic Robotics Architecture (ERA) (Morse, DeGreeff, Belpaeme, & Cangelosi, 2010): a neural cognitive architecture for general, scalable and embodied learning and modeling of psychological function. This architecture is particularly suited to model the role of embodiment and agent-environment interaction in development.

Modeling even a part of the process of development itself is an inherently general proposition, as humans we all go through significant physical and mental development from conception into adulthood and old age. Some of this development can be attributed to physical growth or other factors principally under genetic control. This is the case, for example, of the development of the musculoskeletal system for walking, or the mental and physical effects of puberty. Other developmental transitions are more obviously influenced by our physical and social environments, such as learning to read, or which languages you speak. But no single developmental phenomenon results wholly from nature or nurture alone (Karmiloff-Smith, 2000; Oyama, 2000a, 2000b). We are not static agents untouched by our past and we are more than the unfolding

of our genetic program. The environment always plays a role, as we shall see in the experiments herein. Recognizing this, and in contrast to a growing body of modeling work in which adaptation does not occur during the lifetime of an agent (e.g. artificial evolution), is the field of Epigenetic or Developmental Robotics (Metta & Berthouze, 2006, p. 129). While there are clear technological outcomes from endowing robots with the capacity to learn and develop, herein we focus our modeling efforts to aid and refine our understanding of human development. As the general model of U-shape learning proposed here is based on the Epigenetic Robotics Architecture (already used in development robotics experiments, Morse, Belpaeme, Cangelosi, & Smith, 2010; Morse, DeGreeff, et al., 2010), the extension of this study to new robotics experiments is facilitated.

So what is (ongoing) development in humans? From Experimental Psychology we know that much of development is not simply the linear acquisition of new skills / abilities / knowledge. Instead, the outward effects of development often happen in non-linear stage-like transitions, and rarely is it the case that some new behavior or ability is simply added to an otherwise unchanged pile. A commonly found phenomenon in developmental psychology is known as the U-shaped curve; here previously stable abilities become temporarily absent or disrupted for a period of time (sometimes months) before returning in a changed but stable form as new competencies emerge. This U-shaped pattern of behavior reoccurs again and again throughout the child development literature and is not specific to the involvement of any particular modality or physicality. This is, of course, not the only pattern of development to be found but its frequent occurrence combined with independence from any particular mode of expression or sensory modality would seem to indicate a common feature of the learning systems involved. As such, competing accounts of this U-shaped pattern of behavior can potentially have far-reaching impact on the cognitive sciences.

U-shaped curve phenomena appear to be independent from any particular task or modality as the following prominent examples demonstrate: For example, Bosch and Sebastián-Gallés argue that initially, bilingual infants track statistical regularities across the two languages, leading to their temporary inability to discriminate acoustically similar

phonetic categories (Bosch & Sebastián-Gallés, 2003a, 2003b, 2005; Sebastián-Gallés & Bosch, 2009). Here the U-shaped development curve occurs in auditory word discrimination, while in another example it is apparent in imitation (visual and motor modalities). Babies initially imitate tongue protrusion (Meltzoff & Moore, 1977). However over the following months this imitation declines until at 6-months-old no tongue protrusion imitations are observed (Abravanel & Sigafoos, 1984; Fontaine, 1984; Heimann, Nelson, & Schaller, 1989). Then, by the end of the first year the imitation of tongue protrusion is back (Piaget & Cook, 1952). To include a modeling example, Plunkett and Juola (1999) model a U-shape curve in the production noun and verb errors, where initial production is error free, but is followed by a period of intermittent over-regularization of irregular nouns and verbs.

U-Shaped Patterns in the Development of Children's Phonetic Discrimination Responses

While there are many more examples we now focus more closely on one particular set of experiments in this case involving children's responses to paired visual and auditory input, and an apparent U-shaped profile of performance in the Switch task occurring around 12 months of age. Since the seminal study by Eimas et al. (1971) showing phonetic discrimination and categorical perception for consonant sounds in 1- and 4-month-old infants, it has been firmly established that infants until 6 months of age are equipped with excellent phonetic discrimination abilities, not only for speech sounds that are found in their native language but also for non-native speech sounds. By the end of the first year, these abilities undergo a "perceptual reorganisation", leading to the maintenance, or increase in native contrast discrimination, and a decrease, or maintenance in non-native contrast discrimination (Kuhl et al., 2006; Werker & Tees, 1984).

During the nineties, psycholinguists started investigating word formation and representation using various paradigms. Typically, infants produce their first words by the end of the first year, but are thought to have stored in memory a substantial amount of word forms before that age (at 8 months probably several dozens, according to Swingley 2009). This doesn't mean that they understand them, that is, the child might not have linked them all to a meaning, but she will have segmented them from continuous speech and retained them in long term memory (see also Jusczyk & Hohne, 1997).

The Switch Task

Stager and Werker (1997) first introduced the Switch procedure as a method for investigating the process of word learning in 8- and 14-month-old toddlers. The procedure is as follows: The child is presented repeatedly with a novel object A, which is labelled with a new word, for example a "neem". This is done until a habituation criterion is reached, that is, when looking times have decreased to a certain extent taken to indicate familiarity with the object A – "neem" paring. Then the child is presented with a new

object B, paired with another label "lif", again until the habituation criterion is reached. Then two test trials are introduced; the "same trial" test in which object A is presented with its original label "neem", and the "switch trial" in which the same object A is now paired with the other label "lif". The rationale is that if the children have encoded the pairing between each object and its label, and if they can recognise and discriminate the auditory and visual stimuli, then they should be surprised by (and look longer towards) a switch trial in which an object is paired with a "wrong" label, as compared to a same trial. Here 14-month-olds did look longer toward switch trials. Following this, a simpler version of the switch procedure was introduced with only one object paired with one sound, and then the sound is changed. This second version was tested with both 8- and 14-month old children and both age groups were found to look longer toward objects in switch trials. Unexpectedly, the authors did not find any significant surprise reaction in 14-month-olds when using two labels distant by only one phonetic feature ("lif" Vs "rif"), suggesting that when they are engaged in a word learning task, their phonetic discrimination abilities "suffer". This was surprising for several reasons; firstly, as mentioned above, decades of research had shown infants' excellent phonetic capacities in simple auditory tasks, secondly the 8-month-old children could do it in the simplified version, and thirdly Stager and Werker themselves demonstrated this same phonetic discrimination in 14-month-old children in the absence of a visual pairing.

Further experiments using the Switch procedure have exposed several variables affecting performance. Rost and McMurray (2009) have shown that the use of multiple exemplars of the same stimulus ("puk" vs "buk" recorded by 18 different speakers, each producing 3 tokens) increases the surprise response in 14-month-olds, a result we model herein. Werker et al. (2002) report a positive correlation between performance in the switch task and vocabulary size (comprehension and production) in 14-month-olds. This correlation tends to vanish with age; at 17 months, there is still a trend, at 20 months not at all.

To summarise, children at 6- and 8-months-old react to all the switch trials (in the simplified version) with surprise (increased looking), while 14-month-old children only react with surprise if the labels are distant. By 20-months-old, and earlier if vocabulary size is large, surprise is reinstated for all switch trials.

Accounting for the Phenomenon

Werker et al (2002) suggest that a temporary problem with cognitive resource allocations in tasks like the Switch task could be responsible for the phenomena just discussed. That is, what children have to do in order to succeed in that task takes so much resource in working memory/mental space that "something has to give", which happens to be phonetic processing. However, it is not clear why this would happen in the 14-month-old children and not in the 8- and 20-month-olds. Such resource allocation accounts would seem

to be a relic of the computer-mind metaphor in which limited ‘brain-resources’ are dynamically assigned to different tasks. Biologically speaking brain processes operate in parallel. Perhaps a more sympathetic interpretation is that interference effects could be responsible should the same brain areas be involved in conflicting tasks. Indeed this is close to the account that we will give but to give credence to such an account far more details are necessary.

Another alternative hypothesis is that children represent newly learnt words holistically e.g. (Charles-Luce & Luce, 1990), that is, as an underspecified phonetic representation. In other words, instead of having stored a phonemic string for “dog” like /d/-/O/-/g/, they have stored a global acoustic representation. This holistic stage would signal a discontinuity in the process of language development, given that in early childhood infants display excellent phonetic processing abilities. And although it is reasonable to suppose that children may in fact develop both phonetic and holistic recognition simultaneously, to account for the Switch task data the former would have to be lost or suppressed during the holistic stage, which would extend from 12 months to 18 months. Furthermore it is not clear why a holistic classification would necessarily be less sensitive to phonetic changes than phonetic classifications. Nevertheless if holistic recognition were less sensitive then the requirement for increasingly fine-grained recognition due to an increasing vocabulary would force another reorganization –back to phonetics again-, though again one may well ask why. Despite our reservations, even if this account is correct then it has little to say about similar U-shaped curves in development elsewhere.

What we propose as an alternative account is both simpler than either of these hypotheses and sufficiently detailed to provide an implemented computational model as task and modality independent as U-shaped curve phenomena are developmental psychology. The simple idea is that first the infant gains recognition abilities in an isolated way relying on local information only. Simultaneously associations are formed between classifications in different areas and they provide more information (via priming). So far, this is fairly uncontroversial. However this additional information can be used to further refine / reorganize pattern recognition abilities. During this reorganization, independent recognition performance should remain high but multi- or inter-modal tasks will suffer, as priming becomes temporarily misaligned causing interference. Using the previous example of visual object - sound association, we hypothesise that the recognition of the visual object leads to an expectation of the word “lif” via priming. However, if what is heard contrasts with what is primed then this is not a familiar pairing and the child elicits surprise. So to state our hypothesis more formally: *If priming is influential in the organization of local recognition, then there will necessarily be a lag during which priming is temporarily mis-aligned and performance suffers producing a U-shaped curve in performance.* As an example consider the following:

- 1) A, and B frequently co-occur and so are associated
- 2) the priming from A to B changes B into B’ (the thing that previously responded to B now responds to B’)
- 3) A and B’ are in a different relationship to A and B or do not co-occur
- 4) BUT A still primes B’ and so we have interference or mis-priming
- 5) Eventually the association becomes weak and A no longer primes B’, the interference goes away

Eventually this reorganization will conclude and the system will stabilise with new competencies and high performance once again in both independent and cross-modality / cross-sub-modality tasks. Furthermore the simultaneous priming of multiple recognisers in one map will draw them together while differential priming of close features will push them apart both enhancing meaningful experienced feature discrimination and potentially reducing discrimination of non-experienced features.

Details of the model

We begin by suggesting that the sensory input areas of the brain perform some kind of pattern recognition, adapting to classify the input they receive. Cognitive development can be seen as the learning of associations between emerging discrimination capacities through interaction with the world (Morse, Belpaeme, et al., 2010). In the previous example this would consist on transformation of acoustic input in one area, and visual input in another. Herein we will use Self-Organizing-Maps (SOM) (Kohonen, 1998) to provide a simple and biologically relevant classifier and simple Hebbian learning to form associations between classifications in different maps. We anticipate that our results should be independent from the classifier used (though timescales may vary). While standard SOM’s provide pattern recognition, they do not allow for active Hebbian links to participate in the map learning process, thus instead we use the active learning equations proposed by Althus (2010) as follows:

Equation 1: Initial direct activation of SOM units

$$DirA_j = \sqrt{\sum_{i=0}^{i=n} (v_i - w_{ij})^2}$$

Where $DirA_j$ is the resulting activity of each node in the map following a forward pass of the SOM, v_i is an input, and w_{ij} is the weight between that input and the current node. The winning node is the node with the smallest value for $DirA_i$

Equation 2: Initial indirect activation of SOM units

$$IndA_j = \sum_{i=0}^{i=n} x_i \times w_{ij}$$

Where $IndA_j$ is the resulting activity of each node in the map due to indirect activation via Hebbian association, x_i is the pre-gaussian activity of unit i , in the other map and w_{ij} is the Hebbian weight between it and unit j in this map.

Equation 3: Gaussian direct or indirect activation of SOM units

$$y_i = e^{\left(\frac{-\beta_i}{2\sqrt{n}}\right)}$$

Where y_i is the final activation of the i^{th} node in the map, β is the distance from node i to the winning unit (either direct or indirect), and n is the total number of nodes in the map. Note: units not within the neighborhood size are set to zero output activation, the neighborhood size and learning rate are logarithmically decreased.

Equation 4: Joint activation of SOM units

$$JoinT_i = (1 - \lambda)DirY_i + \lambda IndY_i$$

Where $JoinT_j$ is the final resulting activity of each node in the map due to the combination of direct and indirect activation, and λ is the activation mixture co-efficient (0.1).

Equation 5: SOM weight changes

$$\Delta w_{ij} = \alpha(v_i - w_{ij})JoinT_i - \zeta(JoinT_i - DirY_i)(v_i - w_{ij})$$

Where w_{ij} is the weight between input j and unit i , α is the learning rate (0.1 – 0.0), and ζ is the inhibition rate (0.001 – 0.07).

Equation 6 Positive Hebbian learning (weight changes between maps)

$$\Delta w_{ij} = \alpha_{heb} DirA_i DirA_j$$

Where w_{ij} is the weight between node j and node i , α is the hebbian learning rate (0.01), $DirA_i$ is the initial direct activity of node i , in one map and $DirA_j$ is the initial direct activity of node j , in the other map.

Using these equations, where direct and indirect activation coincide things proceed as normal, however where they don't coincide the indirect input is given precedence and the direct input falling outside the influence of the indirect input is pushed away. This reduces the number of outliers responding to a given category and has been argued as a move from topographic representation to

category representation (Althaus, 2010). While Althus (2010) used dynamically varying parameters for α , λ , and ζ , we find our results are consistent and immune to variation in parameter settings within the ranges attempted (stated in the equations above) with the exception of λ which must remain low, i.e. the direct influence on the map's activity should remain larger than the indirect (primed) influence.

Method

In this model, two SOMs, each of 100 units and each receiving three inputs were randomly initialised in the range 0-1. The two SOMs might be considered examples of visual (e.g. color/shape) and auditory (phonetic) recognition processes (as in the language learning model in Morse, Belpaeme, et al., 2010), though the current modelling experiment represents a general, task-independent model of developmental learning. 10 input categories were defined by uniformly distributing points (in the range 0-1) within the 3D input space so as to ensure no distribution bias. In experiment 1 no noise was used, however in the experiment 2, each example of each category included random noise around the category locus according to a noise window parameter (0.1), and both maps received different exemplars of the same category simultaneously. The network is given input examples of each of the 10 categories in turn (in random order) whilst learning. Learning is then temporarily disabled and the network is tested with a fixed category example to one map and a number of fixed inputs to the other map, corresponding to: a No-Switch trial (input is generated from the same category), a Switch trial using input generated from a neighbouring category, and a Switch trial using input equidistant between the two categories. This testing phase is analogous to the switch trials in which 'wrong' words differed from the correct paired word by differing amounts ("lif"- "neem" Vs "lif"- "rif"). Following this testing learning is re-enabled and the network is shown the 10 category examples again. This whole cycle is then repeated 1500 times and the results of testing at each cycle recorded to produce data for one individual. In each experiment this method was repeated for 20 randomly initialised networks to produce data for 20 individuals. Connection weights between the two maps where all initialised at 0.

Results

In every test instance, the input produces a winning unit in each SOM. This unit then primes (via learned Hebbian connections) a unit in the other map, which in turn primes one in the first map and so on. With few exceptions (and usually within 2 or 3 cycles) the priming from each 'winner' falls into an attractor (e.g. units that mutually prime each other). If the priming from each winner falls into the same basin of attraction then there is no interference between the maps. The expectations of activity in each map, as primed from the other map, are met and a familiarity judgment is made. If however the priming from each winner falls into different basins of attraction then there is interference

(competition between the primed and direct signals), this means that expectations have not been met, thus a novel pairing judgment is made. Figure 1 and Figure 2 show the mean scores of each 50 consecutive trials for the three conditions. In both population averaged (Figure 1) and individual (Figure 2) graphs we can see that following early learning No-Switch trials resulted in consistently low (familiar) judgments, and the Switch condition with high category difference resulted in consistently high (novel) judgments.

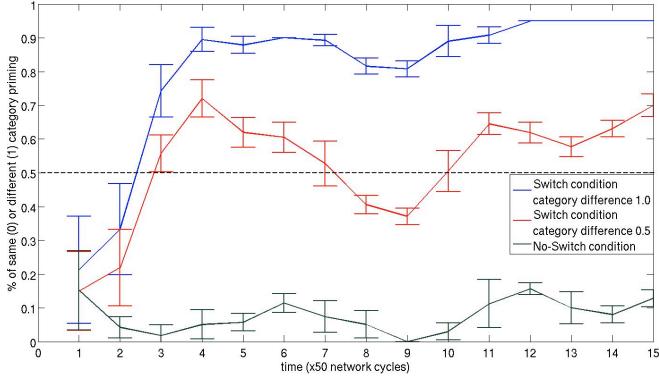


Figure 1: Showing the mean category judgment for each 50 consecutive trials across 20 individuals for 3 different conditions; Switch trial with a large difference (1), Switch trial with a small difference (0.5), and No-Switch trial

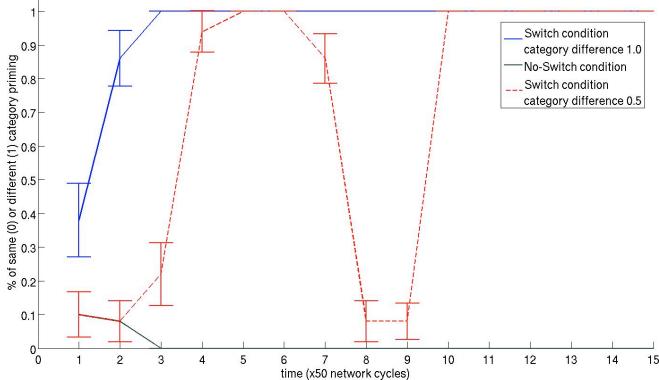


Figure 2: Showing the mean category judgment for each 50 consecutive trials for a single individual and for 3 different conditions; Switch trial with a large difference, Switch trial with a small difference, and No-Switch trial

Inspection of the data plotted on Figure 1 suggests a U-shaped curve for performance in the Switch conditions, with a clear minimum at time step 9 preceded by a maximum at time step 4. In order to evaluate the significance of this behavior, an ANOVA was conducted on individual averaged responses with condition (same items or different items) and peaks (value at time step 4 = max, followed by value at time step 9 = min) as within-participant factors. A main effect of condition was found ($F(1, 19) = 14.8, p = .001$), due to performance in the No-Switch condition being above that of the Switch conditions (.85 vs. .55). There was

also a very predictable effect of peaks ($F(1, 19) = 6.50, p = .02$), due to the fact that we selected a maximum (.81) followed by a minimum (.59). Most importantly, there was a significant interaction between conditions and peaks ($F(1, 19) = 4.46, p = .048$) due to the decrease in the Switch condition being larger than in the No-Switch condition.

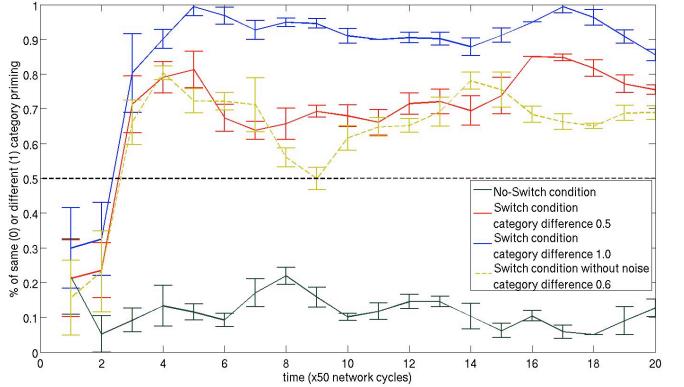


Figure 3: Showing the mean category judgment for each 50 consecutive trials across 20 individuals for 4 different conditions; For networks trained with noise; Switch trial with a large difference, Switch trial with a small difference, and No-Switch trial, and for networks trained without noise; Switch trial with a small difference (0.6).

In experiment 2 the same procedure was followed but with the addition of noise during training. Again similar results were found however the mid category decision line was pushed up indicating an increased level of surprise. The min of the dip was significantly higher ($F(1, 19) = 4.99, p = 0.038$), and the dip of the U-shape was extended (see Figure 3). This has a similar effect to increasing the distance between the categories in the noise-free experiment (the dashed line shown in Figure 3 for comparison), though the dip remains higher.

Discussion and Conclusion

The results presented here represent a novel modelling approach to developmental U-shaped curves, not only replicating a significant and general U-shaped pattern but also capturing additional details such as the increase in levels of surprise using noisy training data, akin to the use of different speakers in Rost and McMurray's (2009) experiments.

Ongoing work has already begun in which the model receives speech and vision input from a real humanoid robot (the iCub robot (Metta, Sandini, Vernon, Natale, & Nori, 2008)), with which we plan to more closely replicate the experiments of Stager and Werker (1997) discussed in the introduction. Nevertheless we have here provided evidence for our simple, task and modality independent, but detailed, account of the U-shape phenomena. This simple but effective model and simulation data allow us to make several predictions. Firstly as the drop in performance is due to interference, caused by a lag between the re-organization

of classifications and the updating of associative links, this should work both ways. That is to say, for example, the familiar Vs novel priming of phonetic classifications by visual stimuli should be accompanied by a similar familiar Vs novel priming of visual stimuli from auditory input. Arguably priming between different regions need not be uniform in strength and vision is more developed than auditory capabilities at this time, however there should still be a measurable effect, that is to say any U-shaped pattern of performance should be accompanied by another U-shaped pattern of performance in a related area. Furthermore while noise in training increases levels of surprise, it also appears to extend the duration of the U-shape (see Figure 3) though further analysis would be required to establish this.

We can also make neuroscience predictions from the model as the topology of the SOM's is related to the organisation of topographic maps in sensory regions of the brain and potentially throughout much of the cortex (see Morse, Belpaeme, et al., 2010; Morse, DeGreeff, et al., 2010). Naturally there is change in these topologies in early learning but there should be further notable changes in the organization of these topologies immediately proceeding and during the dipped phase of a U-shaped curve in development. We are not currently aware of such data from neuroscience but the prediction is certainly verifiable.

Once embodied on the iCub humanoid robot we hope to use this model to capture a greater range of U-shaped phenomena and demonstrate an ability to counter noise by varying the similarity of stimuli used. Future work will also begin to explore possible reasons for the ordering of various developmental transitions. Those interested should note that the software developed to generate this data is freely available (Peniak, Morse, Larcombe, Ramirez-Contla, & Cangelosi, 2011).

Acknowledgments

Work supported by EU FP7 ITALK project (no. 214668).

References

Abrauvel, E., & Sigafoos, A. (1984). Exploring the presence of imitation during early infancy. *Child Development*, 55(2), 381-392.

Althaus, N. (2010). *Categorization in infancy and the influence of verbal learning: A feature-based account*. Birkbeck.

Bosch, L., & Sebastián-Gallés, N. (2003a). *Language experience and the perception of a voicing contrast in fricatives: Infant and adult data*. Paper presented at the International Congress of Phonetic Sciences, Barcelona.

Bosch, L., & Sebastián-Gallés, N. (2003b). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, 46(2-3), 217-243.

Bosch, L., & Sebastián-Gallés, N. (2005). *Developmental changes in the discrimination of vowel contrasts in bilingual infants*. Paper presented at the Proceedings of the 4th international symposium on bilingualism, Barcelona.

Charles-Luce, J., & Luce, P. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of child Language*, 17(01), 205-215.

Elmas, P., Siqueland, E., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171(3968), 303.

Fontaine, R. (1984). Imitative skill between birth and six months. *Infant Behavior and Development*, 7, 323-333.

Heimann, M., Nelson, K., & Schaller, J. (1989). Neonatal imitation of tongue protrusion and mouth opening: Methodological aspects and evidence of early individual differences. *Scandinavian Journal of Psychology*, 30(2), 90-101.

Jusczyk, P., & Hohne, E. (1997). Infants' memory for spoken words. *Science*, 277(5334), 1984.

Karmiloff-Smith, A. (2000). Why Babies' Brains Are Not Swiss Army Knives. In H. Rose & S. Rose (Eds.), *Alas, poor Darwin* (pp. 144-156). London: Jonathan Cape.

Kohonen, T. (1998). The self-organizing map. *Neurocomputing*, 21(1-3), 1-6.

Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13-F21.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, 198, 75-78.

Metta, G., & Berthouze, L. (2006). Epigenetic Robotics: Modelling cognitive development in robotic systems. *Interaction Studies*, 7(2), 129-134.

Metta, G., Sandini, G., Vernon, D., Natale, L., & Nori, F. (2008). *The iCub humanoid robot: an open platform for research in embodied cognition*.

Morse, A. F., Belpaeme, T., Cangelosi, A., & Smith, L. B. (2010). *Thinking With Your Body: Modelling Spatial Biases in Categorization Using a Real Humanoid Robot*. Paper presented at the Cognitive Science Conference 2010, Portland.

Morse, A. F., DeGreeff, J., Belpaeme, T., & Cangelosi, A. (2010). Epigenetic Robotics Architecture (ERA). *IEEE Transactions on Autonomous Mental Development*, 2(4), 325-339.

Oyama, S. (2000a). *Evolution's eye: A systems view of the biology-culture divide*. Duke University Press.

Oyama, S. (2000b). *The ontogeny of information: Developmental systems and evolution*. Duke University Press.

Peniak, M., Morse, A. F., Larcombe, C., Ramirez-Contla, S., & Cangelosi, A. (2011). *Aquila: An Open-Source GPU-Accelerated Toolkit for Cognitive Robotics Research*. Paper presented at the International Joint Conference on Neural Networks (IJCNN) 2011.

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children*. International Universities Press New York.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science*, 23(4), 463-490.

Rost, G., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, 12(2), 339-349.

Sebastián-Gallés, N., & Bosch, L. (2009). Developmental shift in the discrimination of vowel contrasts in bilingual infants: Is the distributional account all there is to it? *Developmental Science*, 12(6), 874-887.

Stager, C., & Werker, J. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381-382.

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3617.

Werker, J., Fennell, C., Corcoran, K., & Stager, C. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1-30.

Werker, J., & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life*. *Infant Behavior and Development*, 7(1), 49-63.