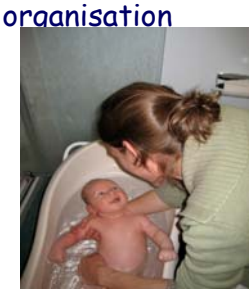
Investigating the basis for conversation between human and robot

Experiments using natural, spontaneous speech, speaking to the robot as if it were a small child
Inspired by the acquisition of language in human infants and by evidence of neuronal organisation

Proto-conversation starts early. Pre-linguistic infants engage in dynamic interaction with their carers, and are also affected by ambient language. By about 6 months baby is babbling, a key stage in language development. (e.g. Pulvermuller, 2002; Oudeyer, 2006). Our experiments start at this stage, with simulated teacher and robot. The baby's babble begins to be biased towards carers' speech. Words and holophrases start to be segmented from a stream of phonemes, but without meaning.

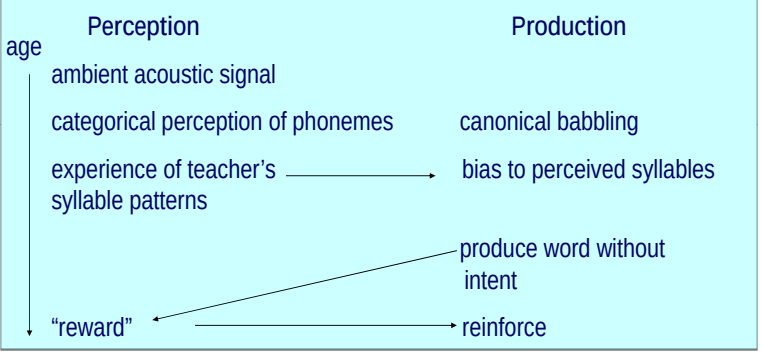


Outline of model

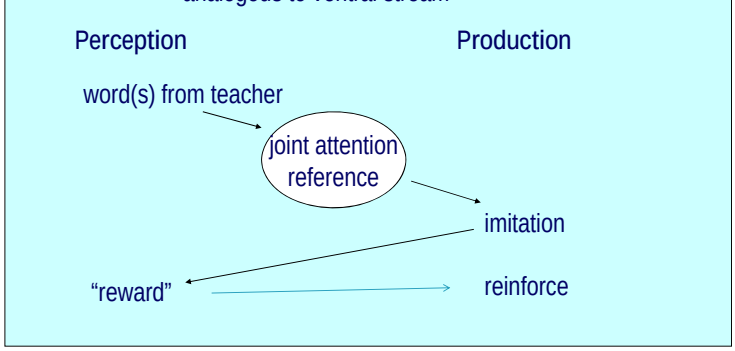
Dual System for language processing

- Implicit learning of patterns and procedures, without intentional shared reference.
- Explicit declarative learning in which there is joint attention between teacher and learner, reference to objects, actions, relationships (Hickok and Poeppel, 2004)

Part 1 Procedural, pattern learning, without meaning, analogous to dorsal processing



Part 2 Explicit learning of word and holophrase meaning, analogous to ventral stream



Experiments with shapes

The adult aims to teach robot Kaspar about different shapes. We assume that

- Kaspar has the intention to communicate
- Communicative ability is learnt through interaction with a teacher.

We investigate whether the robot will be able to extract sufficient information from these interaction episodes to ground the meaning of different shapes.

Attaching meaning to shapes and answering questions about them

The speech stream of the human, represented as phonemes with word or holophrase boundaries, is merged with the robot's sensory/motor stream. Currently this contains (i) head and vision proprioceptive senses and (ii) recognition of pre-trained shapes, so the category of the shape is available. However, the robot has to learn to associate (i) and (ii) with the speech.

Method of learning associations

- Extract significant word(s) using various methods of segmentation
- Compute "information gain" between word and sensory attributes. Store this pair in dynamically growing memory.

How Kaspar produces a response to a question

- For new speech input Kaspar polls memory to find best match of his sensory/motor attributes to significant word(s).
- If match exceeds a threshold Kaspar will reply using a speech synthesiszer.

Example of a proto conversation: "Kaspar, what do you see here?" "Box"

Caroline Lyon and Joe Saunders