Modelling Interactive Language Learning:

Project Presentation

Lacerda, F., Sundberg, U.,[1] Carlson, R.[2] and Holt, L.[3]
[1]Dept. of Linguistics, Stockholm University, Stockholm
[2]Dept of Speech, Music, Hearing, KTH, Stockholm
[3]Dept of Psychology, Carnegie Mellon University, Pittsburgh, USA

Abstract

This paper describes a recently started interdisciplinary research program aiming at investigating and modelling fundamental aspects of the language acquisition process. The working hypothesis assumes that general purpose perception and memory processes, common to both human and other mammalian species, along with the particular context of initial adult-infant interaction, underlie the infant's ability to progressively derive linguistic structure implicitly available in the ambient language. The project is conceived as an interdisciplinary research effort involving the areas of Phonetics, Psychology and Speech recognition. Experimental speech perception techniques will be used at Dept. of Linguistics, SU, to investigate the development of the infant's ability to derive linguistic information from situated connected speech. These experiments will be matched by behavioural tests of animal subjects, carried out at CMU, Pittsburgh, to disclose the potential significance that recurrent multi-sensory properties of the stimuli may have for spontaneous category formation. Data from infant and child vocal productions as well as infant-adult interactions will also be collected and analyzed to address the possibility of a production-perception link. Finally, the data from the infant and animal studies will be integrated and tested in mathematical models of the language acquisition process, developed at TMH, KTH.

Background

In the advent of the experimental studies on the language acquisition process the primary focus was on the infant's ability to discriminate or produce isolated speech sounds (e.g. Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Locke, 1983). However necessary and important as initial steps, the study of isolated perceptual or articulatory phenomena soon falls short of addressing the general linguistic implications of such initial phonetic abilities. Therefore more recent attempts have been instead focused on investigating, for example, the internal structure of phonetic categories in half-year-old infants (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), the infant's ability to recognize word-like patterns from continuous speech (e.g. (Jusczyk, 1999) or to capitalize on statistical regularities in the speech signal (Saffran, Aslin, & Newport, 1996). Also research on infants' vocal production have shifted from focus on isolated speech sound productions to attempts to relate patterns of babbling with a wider biologically oriented linguistic frame (MacNeilage & Davis, 2000). Yet, many of these studies emanate from a perspective of adult-full-fledged linguistic behaviour where the infant is assumed to acquire its ambient language by engaging in a process of finding out its phonemic and other traditional linguistic constituents. Not surprisingly, the infant's capability of learning the ambient language with all its variability, have been proposed to rely on pre-wired linguistic knowledge in terms of Language Acquisition Device or notions of 'poverty of the stimulus'. In the present research program, on the other hand, the infant is assumed to have no innate language-specific predispositions and the language acquisition is viewed as a consequence of general sensory and memory processes and continuity between low-level sensory information processing and language capacity.

From the theoretical outline of the present research project it is assumed that the initial phases of human language acquisition process is a general purpose process through which auditory representations of speech sequences are linked to other co-occurring sensory stimuli as an automatic consequence of exposure to correlated multi-sensory information input. Since spoken language is often used to refer to

objects or actions in the shared outside world, there is an inevitable implicit correlation between hearing the sounds of words or phrases relating to these objects and seeing, feeling, smelling or otherwise perceiving the referents of the spoken language.

Initially, a sound sequence may be associated with any object that happen to be presented with temporal contiguity with the sound, but continued exposure to richly varying spoken language soon provides enough information to narrow the scope or review the sound-object links (Lacerda, 2003). As this happens, a more detailed representation of the initially unanalyzed chunks of spoken sounds starts to emerge, enabling differentiation between identifiable (acquired) sound objects and the remaining unknown sounds.

In the present theoretical framework, this represents the emergence of a general segmentation procedure that will eventually lead to lexical representation. At this basic level, the human representation capacity is likely to be shared by some other species although humans at some level obviously depart to using sound representations as objects that can be processed as combinatorial elements while keeping their attached multi-sensory links. At this level a cognitive process is unparalleled in the animal world begins to emerge but its ontogenesis is, so far, unaccounted for. By the coordinated study of the early segmentation and sound-object association processes in human and animal subjects and detailed simulation on the experimental conditions with operative mathematical models, the current project expects to make a contribution to the understanding of this unique human capacity.

Research plan

Through a joint effort from three areas of research – early human language development, animal learning and speech technology - data will be gathered and used to generate and calibrate mathematical models that may account for language learning in terms of general multisensory input, memory processes and the infant's interaction with its environment.

Infants as models of speech and language development

Human infants typically engage in speech communication activities within their first years of life and usually achieve a well-developed linguistic competence by about 10-12 years of age. Retrospective analyses of the language acquisition process often present language learning as an effortless and smooth development towards the adult language. This approach misses the important information conveyed by difficulties and the errors that children actually do and detaches the speech signal from its ecologic context. Indeed, traditional language development research fails to address some fundamental questions such as:

- Why does the sound of speech seem to be more attractive for the newborn infant than other sounds? In this neonatal preference a sign of human specific genetic mechanisms that under normal circumstances generate behaviour that is taken for an innate preference for listening to spoken language?

- How do representations of phonemes, syllables and words arise and develop in the human infant? Can these traditionally assumed units, the very core of the adult linguistic structure, be derived from continuous speech in the absence of pre-wired linguistic knowledge?

- How do adults conceive the infant's linguistic competence and what phonetic and linguistic modifications do they introduce to accommodate the infant?

- Are production preferences reflected in perceptual biases? For example, how is the perception of phonemic contrasts affected by the child's production abilities? How does the interplay between perception and production develop during the first years of life? How does the combinatorial pressure of lexical representations relate to the child's phonological awareness?

Questions like these will be investigated from a broad developmental perspective using well established techniques such as the High-Amplitude Sucking technique, to test the youngest infants, and the visual preference paradigm to study the emergence of lexical and phonological structure in children up to 2 years of age. The next step will be to investigate the stability over time and

generalization ability from the established auditory-visual representations, testing the subjects' ability to cope with different kind of noise affecting the speech signal, in line with the phoneme restoration and rhyming studies.

The subject population will consist of large cross-sectional samples (about 100 subjects) to address specific perceptual issues plus a group of 20 infant-adult dyads, for additional detailed investigation of production and interaction aspects.

Animal studies of sensory processing and memory

As speech scientists began to argue about the specificity of human speech perception, the results from auditory experiments with animal subjects became a crucial contribution to the on-going debate. For instance, animal studies contributed with important result that certain acoustic continua were categorized in much the same way by both humans and non-human subjects (Kuhl & Padden, 1983; Kuhl & Padden, 1982). More recent studies involving animal subjects have further exposed general mechanisms accounting for the internal organization of categories of speech sounds, the influence of phonetic context (Lotto, Kluender, & Holt, 1997) and sensitivity to acoustic trading relations (Holt, Lotto, & Kluender, 2001; Kluender, Lotto, Holt, & Bloedel, 1998; Kluender & Lotto, 1994; Lotto et al., 1997). Using animal subjects in perception studies has proven to offer a unique possibility to exercise complete experimental control over experience. The general procedure in studies with, for instance, gerbils allows assessing detection, discrimination and identification of speech stimuli.

In the present project gerbils will be used in perception experiments since this species has been used successfully in previous studies. Methods of training these animal subjects are based on a go/no-go paradigm. The gerbils are trained with positive reinforcement to remain on a little platform in a cage on a "positive" stimulus and to jump off the platform on a "negative" stimulus. The gerbils are "at work" 15-20 min. in daily sessions and their performance is measured in terms of % correct responses and d'.

The current project attempts to widen the scope of speech perception experiments with animals by trying to create realistic animal models of certain early stages of the language acquisition process. It is expected that the animal models will perform more or less like the infant subjects up to a certain level of initial representation. By studying the levels of complexity for which the animal models fall short of the human performance the project is expected to contribute with important insights on the emergence of early cognitive processes.

Testing the hypotheses in mathematical models

Addressing the early stages of the language acquisition process from a broad perspective offers a unique opportunity to understand and design speech communication systems that hopefully may be able to capitalize on the key aspects of the human children's flexibility and learning capability rather than attempting to mimic adult stereotypes.

The ultimate goal of a speech recognition system is to be able to handle speech signals with human-like efficiency. While the performance of available systems using speech input may by reasonable under optimal communication situations, the systems' lack of flexibility and vulnerability and noise or moderately adverse communication situations is a significant hinder to a wider and safer application of such speech interfaces. There are several factors that account for the mismatch between the expected and the obtained performance of such speech communication interfaces. In addition to the difficulties inherent to the characterization of the speech signals per se, these engineered interfaces have also been developed to mimic and match the speech communication performance of experienced adult speakers, bypassing the developmental process that eventually led to the adult communicative competence. And while this in an acceptable strategy to quickly approach the ultimate communicative goal it relies on a deterministic view of the speech signal and of the whole human communication process that tends to result into rather rigid systems that can hardly cope with realistic variance in, for instance, the speech signal.

Part of the problems usually faced by speech recognition systems is caused by the strong focus on the speech signal itself, as the primary component of the speech communication process. While this is a reasonable starting point for the development of man- machine speech interfaces, it is obvious that in natural speech communication settings the speech signal is only part, albeit crucial, of the communication process (Colin et al., 2002; McGurk & MacDonald, 1976) and recent developments of language and speech systems are beginning to take advantage of multimodal information to improve the systems' communication efficiency (e.g. Beskow, 2003). However, in spite of the improvement in communication performance clearly accrued by the introduction of additional information sources, the design of these communication systems may still be limited by their attempt to engineer a full-fledged human communication system while attempting to bypass the long developmental path that eventually led to such an adult competent performance. Thus, the current project will attempt to incorporate linguistic and psychological knowledge of the early stages of the language acquisition process in order to design a prototypical system that essentially will be able to learn from multimodal information sources and integrate them to achieve flexible and realistic representations of its environment.

Acknowledgements

The project is funded by The Bank of Sweden Tercentenary Foundation K2003-0867.

References

Beskow, J., 2003. Talking heads – models and applications for multimodal speech synthesis. *PhD thesis*, TMH/KTH.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clin. Neurophysiol. 113*, 495-506.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science 171*, 303-306.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: a case of learned covariation or auditory enhancement? *JASA 109*, 764-774.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends Cogn Sci. 3*, 323-328.

Kluender, K. R. & Lotto, A. J. (1994). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. *JASA 95*, 1044-1052.

Kluender, K. R., Lotto, A. J., Holt, L. L., & Bloedel, S. L. (1998). Role of experience for language-specific functional mappings of vowel sounds. *JASA 104*, 3568-3582.

Kuhl, P. K. & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Percept Psychophys 32*, 542-550.

Kuhl, P. K. & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *JASA 73*, 1003-1010.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science 255*, 606-608.

Lacerda, F. (2003). Phonology: An emergent consequence of memory constraints and sensory input. *Reading and Writing: An Interdisciplinary Journal 16*, 41-59.

Locke, J. L. (1983). *Phonological Acquisition and Change*. New York: Academic Press

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *JASA 102*, 1134-1140.

MacNeilage, P. F. & Davis, B. L. (2000). On the Origin of Internal Structure of Word Forms. *Science 288*, 527-531.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature 264*, 746-748.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science 274*, 1926-1928.