An Exemplar-Based Account of Emergent Phonetic Categories

Francisco Lacerda

Institute of Linguistics, Dept. of Phonetics, Stockholm University, S-106 91 Stockholm, Sweden

Abstract: The early stages of language acquisition and phonetic category formation are seen from the perspective of a (conceptually) very simple exemplar-based model, inspired by Edelman’s notion of Neural Darwinism. The paper briefly sketches how the emergence of phonetic categories may be accounted for by the interaction between ambient language input and limited memory-representation resources and attempts to illustrate how phonetic categorization may emerge “spontaneously” from exposure to natural infant-directed speech data. It is suggested that phonetic categories may emerge if the auditory input is statistically correlated with one (or several) other sensory inputs and that the perceptual-magnet effect can be seen as a consequence of distributed memory representations.

INTRODUCTION

This paper presents a sketch of an exemplar-based model potentially capable of accounting for some aspects of the early language acquisition process.

The model rests on some very simple assumptions. It is assumed that there is a non-deterministic relation between stimuli and their internal representations, in line with Edelman’s (1,2) notions, i.e. a stimulus generates a non-deterministic distributed activity. It is also assumed that the amount of activity generated by a stimulus decays with time and that there is activity summation in the representational space. The first assumption is here implemented by arbitrary two-dimensional Gaussian distributions of activity, of equal standard deviations but with means stochastically related to the stimulus characteristics (fig. 1). The second assumption implies that the activity generated by a stimulus can be seen as a non-permanent memory of that stimulus and the third generates a cumulative effect arising from repetitive exposure to similar stimuli.

The implication of the first assumption is that a given stimulus, or a set of similar stimuli, will generate distributions like that of figure 1. The other two assumptions tend to build up representations that automatically reflect the general characteristics of the most recurrent stimuli. In addition, the distributed representational activity generated by each stimulus allows for an implicit similarity measure, created by the amount of overlap between the representations generated by different stimuli.

Figure 1. Representation of the activity generated by a stimulus

From the model’s point of view, the language acquisition process starts by mere storage of general sensory information. In this perspective there is no need to postulate any special status or specific processing mechanisms affecting the speech signal. Sensory information storage is seen as a non-specific, non-goal oriented process that simply stores all the available information, constrained only by the finite representational resources and by the sensory system’s resolution. Thus, given the finite storage resources and the system’s
spontaneous activity decay, the storage process will ultimately function as an automatic correlator of the information available from the different sensory inputs.

In an ecologically relevant language learning situation the infant is exposed to both infant-directed speech (IDS) — characterized by attention catching exaggerated F0 excursions (3) and by a tendency to hyperarticulation of (at least) vowel segments (4) — along with all the other simultaneously available sensory information, conveyed by the visual, olfactory, tactile and gustatory senses. Since all this information is just stored in a memory that is affected by decay, the model automatically converges to representations that implicitly associate the recurrent multi-modal inputs, as long as the cooccurrence probability of the relevant sensory inputs is higher than that of spurious inputs. This, in turn, is a condition that is necessarily met when speech is used referentially. Furthermore, because this implicit association between relevant sensory inputs is an intrinsic property of language use, any recurrent simultaneous sensory inputs are good representational candidates. In the long run, when these memory representations are triggered by a particular sensory modality, they will inevitably activate the whole multimodal complex of associated sensory representations, where the best correlated modalities will, naturally, dominate.

In an attempt to illustrate these aspects with the present model, a situation was created in which “acoustic input” is presented along with non-deterministic “vowel labels” that are provided by the cooccurrence probability between the acoustic input and “other sensory information” represented by the vowel labels. The results of this exposure are shown in figure 2, depicted as responses to [a], [i] and [u].

Figure 2. Representation of the labeling associations for the vowels [a], [i] and [u].

The vertical axis shows the strength of the association between the acoustic input, in the F1 × F2 space, and each of the vowel labels. Because the association is non-deterministic, the acoustic input will often be erroneously labeled but, as long as the correct association is more likely than the others (as it must be as a consequence of the language's referential function) the dominant label will, in the long run, always be the one determined by the referential use. Furthermore, in line with a Darwinian competition process, if the dominant referential coupling is lost, the memory decay will tend to weaken that particular association and the next most dominant association will take over. This is a process that may represent, for instance, the effects of exposure to a second language.

ACKNOWLEDGEMENTS

Work supported by The Bank of Sweden Tercentenary Foundation, grant 94-0435.

REFERENCES