Chapter 14
I Think I Have Heard That One Before: Recurrence-Based Word Learning with a Robot

Yo Sato
University of Hertfordshire, UK

Ze Ji
University of Hertfordshire, UK

Sander van Dijk
University of Hertfordshire, UK

ABSTRACT

In this chapter, the authors present a model for learning Word-Like Units (WLUs) based on acoustic recurrence, as well as the results of an application of the model to simulated child-directed speech in human-robot interaction. It is a purely acoustic single-modality model: the learning does not invoke extralinguistic factors such as possible references of words or linguistic constructs including phonemes. The main target phenomenon is the learner's perception that a WLU has been repeated. To simulate it, a Dynamic Time Warping (DTW)-based algorithm is introduced to search for recurrent utterances of similar acoustic features. The authors then extend this model to incorporate interaction, corrective feedback in particular, and assess the ameliorating effect of caregiver correction when a WLU, which is close to the real word, is uttered by the learner.

INTRODUCTION

Extracting linguistic units from raw speech sound is an essential part of language acquisition, and its importance in its own right at a very early stage (from birth to the ‘babbling’ stage), as dis-
Based word discovery rely on *phonemes*, defined as the meaning-associated categorical sounds, themselves a top-down abstraction. Against the backdrop that acoustic phoneticians and speech recognition engineers have long struggled to identify their acoustic correlates, this manner of modelling for an infant’s sound form acquisition can only be partial at best. In contrast, roboticists, with their vantage position of having embodied cognition available, generally prefer to look at the association of sounds with non-audio (usually visual) modalities through percepts, thereby resorting effectively to some notion of ‘meaning’ (or ‘reference’).

Thus, to the best of our knowledge, little work has been done to assess the effectiveness of acoustic, single-modality word learning either in computational linguistics or robotics. The work reported in the present chapter addresses this gap. We draw inspiration from an interesting line of research that has recently emerged (Park & Glass, 2008; Aimetti, 2009), and ask how word forms may be learnt by using *acoustic recurrence* in the auditory modality alone. The core idea is simplicity itself: that when the data a child is exposed to contain a limited set of words that are frequently repeated, as is observed in what is known as Child-Directed Speech (CDS), children pick them up acoustically without, and hence prior to, phonemic representations or reference association, based on their episodic memory. To put it informally, they feel, when a similar sound pattern is repeated, ‘Oh, I have heard that one before.’ We develop a perception model of speech sound on this basis, and evaluate by means of computational experiments how effective such a learning mechanism can be in order to develop the perception of words, or more aptly, *Word-Like Units* (WLUs).

One major obstacle for the single-modality auditory perception-learning algorithm is the vast space it has to search. The learner is exposed to a huge amount of speech data, at least if taken as acoustic samples, contrary to the assumption of ‘poverty of stimulus.’ While, as we shall see in the following section “*Associative word-like unit learning,*” such a challenge with single-modality is one of the motivations for cross-modality learning, it still is possible to adhere to the auditory modality alone, if we pay attention to *prosody* or sentential (intonational) accent. We assume that the learner is naturally drawn to the intonational peak of each utterance, and this simple assumption dissolves the search space problem.

Furthermore, we also investigate the interaction, namely corrective feedback from the caregiver and its effect on the learner. The basic idea is to let the learner, not just detect what it has heard before, but also actually *say* it, and see whether this triggers ‘correction’ of the caregiver, and if so, what the effect will be on learning. Now, acoustically based learning is initially imperfect most often, as the imperfection of acoustic-based speech recognition would suggest. We exploit this fact to trigger the response from the caregiver. The child’s imitative utterance, which we call *echoing*, is often partial—echoing a sub-part of a whole word—and the expectation here is that the caregiver then provides the ‘full’ version yet again: what we call *corrective feedback*. For this purpose, we conduct a further set of experiments, which use *simulated* feedback to gauge the effect of interaction over and above the ‘raw’ perception version without interaction, still within the audio modality alone.

In what follows, we present the basic perception model as well as the add-on interaction model, and report the preliminary results of computational simulations of these models. The objectives of this study are twofold: the first is to evaluate the effectiveness of the recurrence-based learning on its own, and the second is to how much it improves with interaction. The results we present here are from offline experiments with simulated interaction. Such a study should ideally be accompanied by real interaction with an articulation model, but we bypass the complicated issue surrounding the articulatory process. In addition, the assumption of single modality is somewhat artificial, because
Related Content

Developing Secure Software Using UML Patterns
[www.igi-global.com/chapter/developing-secure-software-using-uml-patterns/205799?camid=4v1a](www.igi-global.com/chapter/developing-secure-software-using-uml-patterns/205799?camid=4v1a)

Routing in Opportunistic Networks
[www.igi-global.com/article/routing-opportunistic-networks/34033?camid=4v1a](www.igi-global.com/article/routing-opportunistic-networks/34033?camid=4v1a)

Overview of Type-2 Fuzzy Logic Systems
[www.igi-global.com/article/overview-type-fuzzy-logic-systems/70754?camid=4v1a](www.igi-global.com/article/overview-type-fuzzy-logic-systems/70754?camid=4v1a)

An Efficient Coronary Disease Diagnosis System Using Dual-Phase Multi-Objective Optimization and Embedded Feature Selection