Available online at www.sciencedirect.com

ScienceDirect

Computer Speech and Language 34 (2015) 83–86

COMPUTER SPEECH AND LANGUAGE

www.elsevier.com/locate/csl

Introduction for Speech and language for interactive robots

Abstract

This special issue includes research articles which apply spoken language processing to robots that interact with human users through speech, possibly combined with other modalities. Robots that can listen to human speech, understand it, interact according to the conveyed meaning, and respond represent major research and technological challenges. Their common aim is to equip robots with natural interaction abilities. However, robotics and spoken language processing are areas that are typically studied within their respective communities with limited communication across disciplinary boundaries. The articles in this special issue represent examples that address the need for an increased multidisciplinary exchange of ideas.

Keywords: Interactive robots; Speech localisation; Dialogue management; Multimodal interaction; Speech synthesis; Human–robot interaction

1. Aim and scope

Speech-based communication with robots faces important challenges in pushing current technology over the edge of usability in real world scenarios. In contrast to conventional interactive systems, a talking robot always needs to take its physical environment into account when communicating with users. Environments are typically unstructured, dynamic and noisy and therefore challenging for robots. The objective of this special issue is to highlight research that applies speech and language processing to robots that interact with people through speech as the main modality of interaction. For example, a robot may need to communicate with users via distant speech recognition and understand with constantly changing degrees of noise.

Alternatively, the robot may coordinate its verbal and non-verbal turn-taking behaviours as when generating speech and gestures at the same time. Speech and language technologies have huge potential to equip robots that interact naturally with their human users. However, the effectiveness of interactive robots needs to be demonstrated in real (or at least realistic) environments. This special issue presents some case studies.

2. Articles in the special issue

This special issue received a total of 20 submissions, 11 of which were originally accepted for publication. Each accepted article went through two or three rounds of reviewing, and each submission was assigned two or three reviewers. The contents of this special issue cover the following four broad topics, which are vital for interactive robots across domains:

http://dx.doi.org/10.1016/j.csl.2015.05.006
0885-2308/

2.1. Speech localisation

An important ability of interactive robots is to accurately estimate the direction of arrival of human speech, also referred to as 'sound source localisation'. This is required in order to analyse auditory scenes around robots and is an important pre-processing step followed by sound source separation and automatic speech recognition. This ability is also important to exhibit socially-interactive behaviours such as moving the robot platform and gaze to the speaker(s) in focus.

The article **A survey on sound source localisation in robotics: from binaural to array processing methods** of authors Argentieri et al. (2015) presents a survey of the state of the art in sound source localisation in robotics. It discusses topics such as embeddability, real-time, broadband environments, noise and reverberation—which are rarely taken into account simultaneously in the areas of acoustics or signal processing. The authors review binaural approaches as well as array processing techniques for localisation of human speech in robot audition.

The article **Subspace-based DOA with linear phase approximation and frequency bin selection preprocessing for interactive robots in noisy environments** of authors Lee et al. (2015) proposes a method for predicting the direction of arrival of human speech in noisy environments. The authors motivate their work in regards to the requirement of robust operation of interactive robots in the real world. The proposed method rectifies the speech signals from a microphone array affected by noise and reconstructs a representation of the received signals. The authors find that their proposed method yields improved results over conventional methods.

The article **Robust speaker localisation for real-world robots** of authors Athanasopoulos et al. (2015) describes a series of enhancements to existing acoustic localisation techniques. The authors propose novel preprocessing and time-delayed techniques for more robust localisation of human speech, which takes into account the imperfect frequency response of microphone arrays. Experimental results using a humanoid robot listening to multiple speakers report that the proposed and extended techniques improve the localisation performance in noisy and reverberant conditions.

2.2. Language understanding

Another important ability of interactive robots is to understand the meaning of human speech taking into account the entities in the spatial environment, e.g. understanding references to objects in motion. The article **Employing distance-based semantics to interpret spoken referring expressions** of authors Zukerman et al. (2015) proposes a method for interpreting spoken referring expressions. The approach considers multiple alternative recognition hypotheses at different stages including lexical, syntactic, semantics and pragmatics. At each stage, uncertainty scores are calculated and subsequently combined to reduce speech recognition errors and ambiguity. The proposed method considers the lexical similarity between the referring expression and the properties of potential referents using distance metrics. The authors report promising results from a comparison between the proposed method against humans doing the same task.

The article **Situated language understanding for a spoken dialogue system within vehicles** of authors Misu et al. (2015) studies situated language understanding in a mobile in-car system that can answer questions about the user's surroundings. The authors propose methods for understanding user queries taking into account changes in spatial relationships between the car and target buildings. They carry out an analysis of the timing in user utterances collected in a real driving setting. Language understanding takes into account the spatial relationships between car and targets, head pose of the user, and linguistic cues (expressions such as 'across the street' or 'colourful'). The analysis is then used to train probabilistic methods that can identify points of interest in referring expressions.

The article **The roles and recognition of haptic-ostensive actions in collaborative multimodal human–human dialogues** of authors Chen et al. (2015) investigates referential expressions for interactive robots for the elderly in home environments. They focus on a specific type of interaction based on multimodal actions referred to as 'Haptic-Ostensive', which not only manipulate objects but also perform a referring function in the spatial environment. The authors collect and analyse human–human dialogues in the home domain including haptic (force) actions, and train supervised models for reference resolution and dialogue act recognition. Additional experiments are reported on the recognition of actions from haptic signals measured through a sensory glove whose pressure sensors are relatively imprecise.