Accepted Manuscript

Introduction to the Special Issue on Auditory-visual expressive speech and gesture in humans and machines

Jeesun Kim , Gérard Bailly , Chris Davis

PII: S0167-6393(18)30021-9
DOI: 10.1016/j.specom.2018.02.001
Reference: SPECOM 2532

To appear in: *Speech Communication*

Received date: 17 January 2018
Revised date: 1 February 2018
Accepted date: 1 February 2018

Please cite this article as: Jeesun Kim , Gérard Bailly , Chris Davis , Introduction to the Special Issue on Auditory-visual expressive speech and gesture in humans and machines, *Speech Communication* (2018), doi: 10.1016/j.specom.2018.02.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Introduction to the Special Issue on Auditory-visual expressive speech and gesture in humans and machines

Jeesun Kim[1], Gérard Bailly[2], Chris Davis[1]

[1]The MARCS Institute, Western Sydney University, Australia;
[2] GIPSA-Lab, Grenoble-Alpes University, France.

Abstract

We speak to express ourselves. Sometimes words can capture what we mean; sometimes we mean more than can be said. This is where our visible gestures - those dynamic oscillations of our gaze, face, head, hand, arms and bodies – help. Not only do these co-verbal visual signals help express our intentions, attitudes and emotion, they also help us engage with our conversational partners to get our message across. Understanding how and when a message is supplemented, shaped and changed by auditory and visual signals is crucial for a science ultimately interesting in the correct interpretation of transmitted meaning. This special issue highlights research articles that explore co-verbal and nonverbal signals, a key topic in speech communication since these are crucial ingredients in the interpretation of meaning. That is, the meaning of speech is calibrated, augmented and even changed by co-verbal/speech behaviours and gestures including the talker's facial expression, eye-contact, gaze-direction, arm movements, hand gestures, body motion and orientation, posture, proximity, physical contact, and so on. Understanding expressive signals is a vital step for developing machines that can properly decipher intention and engage as social agents. The special issue is divided into three parts: Auditory-visual speech perception; Characterization and perception of auditory-visual prosody; Computer-generated auditory-visual speech. Below, we introduce these papers with a brief review of relevant issues and previous studies, when needed.

1 Aim and scope

This special issue aims to provide a platform for consolidating research on human expression and interactive social robotics. Behavioural research seeks to identify the conditioning behaviors that support, structure and maintain social interactions along with auditory and visual cues that signal such things as attitude, emotion and importance. Research on interactive human-agent interaction (e.g., computer vision; machine learning, robot design, etc.) provides the methods, techniques to help quantify communicative signals and to test theories by implementing cues in interactive agents. In summary, the broad aim of the special issue is to connect human and machine research on expressive speech and gesture.

2 Processing auditory-visual expressive speech and gestures: What and why?

A traditional way of classifying the information conveyed by speech is to consider three overlapping categories of information: linguistic, paralinguistic, and non-linguistic (e.g., Fujisaki, 2003). The topic of the special issue largely concerns the paralinguistic information conveyed by expressive speech; although in our view the division between linguistic, paralinguistic, and non-linguistic information is rather porous. In defining the scope of paralinguistic information, some authors (e.g., Crystal, 1974) have excluded visual information; clearly in the current special issue dedicated to auditory and visual expressive speech we do not impose such a limitation. Why are both auditory and visual information important? The key reason why both auditory and visual information are important is simple that the information from both is relevant to capturing the expressive aspects of speech. That is, if the aim is to understand how intentions, beliefs, desires, and so forth, are signaled, then the relevant information must be in the input.

Download English Version:

https://daneshyari.com/en/article/6960611

Download Persian Version:

https://daneshyari.com/article/6960611

Daneshyari.com