Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 66 (2015) 182–217

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Audiovisual speech synthesis: An overview of the state-of-the-art

Wesley Mattheyses [a,*], Werner Verhelst [a,b]

[a] Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[b] iMinds, Gaston Crommenlaan 8, Box 102, 9050 Ghent, Belgium

Received 21 February 2014; received in revised form 4 November 2014; accepted 4 November 2014
Available online 11 November 2014

Abstract

We live in a world where there are countless interactions with computer systems in every-day situations. In the most ideal case, this interaction feels as familiar and as natural as the communication we experience with other humans. To this end, an ideal means of communication between a user and a computer system consists of audiovisual speech signals. Audiovisual text-to-speech technology allows the computer system to utter any spoken message towards its users. Over the last decades, a wide range of techniques for performing audiovisual speech synthesis has been developed. This paper gives a comprehensive overview on these approaches using a categorization of the systems based on multiple important aspects that determine the properties of the synthesized speech signals. The paper makes a clear distinction between the techniques that are used to model the virtual speaker and the techniques that are used to generate the appropriate speech gestures. In addition, the paper discusses the evaluation of audiovisual speech synthesizers, it elaborates on the hardware requirements for performing visual speech synthesis and it describes some important future directions that should stimulate the use of audiovisual speech synthesis technology in real-life applications.

Keywords: Audiovisual speech synthesis; Visual speech synthesis; Speech synthesis

Contents

* Corresponding author.
E-mail addresses: wmatthey@etro.vub.ac.be (W. Mattheyses), wverhels@etro.vub.ac.be (W. Verhelst).

http://dx.doi.org/10.1016/j.specom.2014.11.001
0167-6393/

1. Introduction to audiovisual speech synthesis

1.1. Motivation for speech synthesis

At the present time, cars, heavy machinery, vending machines, medical devices, and even the simplest home appliances such as fridges and central heating systems are computer controlled. For every such device a human–machine interface is needed for enabling users to control the machine and to make it possible for the device to give feedback to its users. The ultimate goal should be that the computer systems that surround us are perfectly integrated