Chapter 7
Speech Technologies for Augmented Communication

Gérard Bailly
CNRS/Universities of Grenoble, France

Pierre Badin
CNRS/Universities of Grenoble, France

Denis Beaufemps
CNRS/Universities of Grenoble, France

Frédéric Elisei
CNRS/Universities of Grenoble, France

ABSTRACT

The authors introduce here an emerging technological and scientific field. Augmented speech communication (ASC) aims at supplementing human-human communication with enhanced or additional modalities. ASC improves human-human communication by exploiting a priori knowledge on multimodal coherence of speech signals, user/listener voice characteristics or more general linguistic and phonological structure on the spoken language or vocabulary being exchanged. The nature of this a priori knowledge, the quantitative models that implement it and their capabilities to enhance the available input signals influence the precision and robustness of the perceived signals. After a general overview of the possible input signals characterizing speech production activity and available technologies for mapping these various speech representations between each other, three ASC systems developed at GIPSA-Lab are described in detail. Preliminary results of the evaluation of these three systems will be given and commented. A discussion on scientific and technological challenges and limitations of ASC concludes the chapter.

INTRODUCTION

Speech is very likely the most natural communication mean for humans. However, there are various situations in which audio speech cannot be used because of disabilities or adverse environmental conditions. Resorting to alternative methods such as augmented speech is a therefore an interesting approach. This chapter presents computer-mediated
communication technologies that allow such an approach (see Figure 1). Speech of the emitter may in fact:

- not be captured by available hardware communication channels – camera, microphone
- be impoverished by the quality of the hardware or the communication channel
- be impoverished because of environmental conditions or because of motor impairments of the interlocutor

On the reception side, Augmented Speech Communication (ASC) may also compensate for perceptual deficits of the user by enhancing the captured signals or adding multimodal redundancy by synthesizing new perceptual channels or adding new features to existing channels. In order to improve human-human communication ASC can make use of a priori knowledge on multimodal coherence of speech signals, user/listener voice characteristics or more general linguistic and phonological structure on the spoken language or vocabulary being exchanged. The nature of this a priori knowledge, the quantitative models that implement it and their capabilities to enhance the available communication signals influence the precision and robustness of the communication.

The chapter will first present:

- the signals that can characterise the speech production activity i.e. from electromagnetic signals from brain activity, through articulatory movements, to their audiovisual traces