Chapter 8

3D Talking-Head Interface to Voice-Interactive Services on Mobile Phones

Jiri Danihelka
Czech Technical University in Prague, Czech Republic

Roman Hak
Czech Technical University in Prague, Czech Republic

Lukas Kencl
Czech Technical University in Prague, Czech Republic

Jiri Zara
Czech Technical University in Prague, Czech Republic

ABSTRACT

This paper presents a novel framework for easy creation of interactive, platform-independent voice-services with an animated 3D talking-head interface, on mobile phones. The Framework supports automated multi-modal interaction using speech and 3D graphics. The difficulty of synchronizing the audio stream to the animation is examined and alternatives for distributed network control of the animation and application logic is discussed. The ability of modern mobile devices to handle such applications is documented and it is shown that the power consumption trade-off of rendering on the mobile phone versus streaming from the server favors the phone. The presented tools will empower developers and researchers in future research and usability studies in the area of mobile talking-head applications (Figure 1). These may be used for example in entertainment, commerce, health care or education.

INTRODUCTION

Rapid proliferation of mobile devices over the past decade and their enormous improvements in terms of computing power and display quality opens new possibilities in using 3D representations for complementing voice-based user interaction. Their rendering power allows creation of new user interfaces that combine 3D graphics with speech recognition and synthesis. Likewise, powerful speech-recognition and synthesis tools are becoming widely available on mobile clients...
or readily accessible over the network, using standardized protocols and APIs. The presented 3-dimensional talking head on a mobile phone display represents a promising alternative to the traditional menu/windows/icons interface for sophisticated applications, or a more complete and natural communication alternative to purely voice-or-tone-based interaction. Such interface has proven many time to be useful as a virtual news reader (Alexa, Berner, Hellenschmidt, & Rieger, 2001), weather forecast (Kunc, & Kleindienst, 2007), healthcare communication assistant (Keskö, Balci, Aran, Sankur, & Akarun, 2007) blog enhancement (Kunc, Slavik, & Kleindienst, 2008) and can be very useful especially in developing regions where people often cannot read and write.

So far, talking-head interfaces have been used mostly on desktop PCs. Existing frameworks for talking-head development on desktop PCs (Wang, Emmi, & Faloutsos, 2007; Balci, 2005) have inspired our work. Emerging electronics such as mobile phones, pocket computers or embedded devices now possess enough power to enable a talking-head interface, but lack tools for creating such applications. In this paper we propose an effective architecture for interactive, fully-automated 3D-talking-head applications on a mobile client (Figure 1) and implement a framework for easy creation of such applications. The main contributions of this work are:

- We document that performance limits of contemporary mobile devices are sufficient for running a 3D+audio interface by practical experiments and benchmarks;
- We describe practical techniques of synchronizing the audio stream and visual animation to deliver convincing talking-head interaction on the mobile device;
- We present a platform-independent prototype implementation of a distributed framework for creating and generating the 3D-talking-head applications.

By providing a general tool for creating interactive talking-head applications on mobile platforms, we aim to spark future research in this area. It may open up space for many useful applications, such as interactive mobile virtual assistants, coaches or customer-care, e-government platforms, in-