Chapter 2

From Wood to Bits to Silicon Chips: A History of Developments in Computer Synthesized Speech

Debbie A. Rowe
Rensselaer Polytechnic Institute, USA

ABSTRACT

This chapter lists some of the key inventions and applications in the history of computer synthesized speech (CSS). Starting with a brief look at the early synthesis machines—precursors to the computerized renditions of the 20th century—the chapter proceeds to look at the strides made by corporations, such as Bell Labs, IBM, Apple Inc., and Microsoft, in creating assistive technologies that tap into the benefits of CSS. There is also a discussion on developments in the fields of Neuroscience, Robotics, and the non-scientific fields of Composition and the Arts. Finally, the chapter explores how CSS has permeated the popular culture mediums of film and television, sometimes in parallel and sometimes as antecedents to current day inventions.

INTRODUCTION

Attempts to simulate human speech with inanimate objects have been in force well before the dawn of computers in the 20th Century. Some scholars have indicated that initial attempts go as far back as the ancient world of the Romans and the Greeks, where high priests would project their voices through statues of idols as if they were puppets (Cater, 1983; Coker et al, 1963). However, many would agree that the first legitimate attempt to have inanimate objects replicate the sounds of a human voice came in 18th Century Europe, where mechanical devices were engineered to produce certain vowel and consonant sounds. It would take approximately another 150 years for man to progress from machine-generated synthetic speech to electronically-generated versions, and then several decades more to streamline a computer-generated form. During that time, speech synthesis would go from a way to study human speech patterns, to a component of telecommunications, to an assistive technology tool, and beyond. This chapter will look at these developments.

DOI: 10.4018/978-1-61520-725-1.ch002
INDIVIDUAL INNOVATION

As is the case with many modern technologies produced en masse today, we have the inventive spirit of individuals to thank for laying the groundwork to today’s computer synthesized speech. Researchers point to 18th century Europe as the birthplace of mechanical speech synthesizers. Wolfgang Ritter von Kempelen, a Hungarian nobleman, engineer, and government official, invented one of the first synthesizer machines (Coker et al., 1963; Schroeder, 2004). According to Manfred Schroeder’s research (2004), von Kempelen began development of his earliest “speaking machine” in 1769, designing it from wood and leather to replicate the functions of human lungs, vocal cords, and vocal tracts. The mechanism was manually operated, with air blown through the various parts to produce sounds (Cater, 1983). The first rendition of the machine created vowel sounds. A later version had a “tongue” and “lips” that enabled it to produce consonants with plosive sounds, “such as the /b/ (as in bin) and /d/ (as in din)” (Schroeder, 2004, p. 26). It is unclear why von Kempelen decided to make the device. During the period of his inventions, there was a growing interest in “spring-operated automatons” that replicated human actions (Coker et al., 1963, p. 3). This could have influenced his decision to make the speaking machine, and to subsequently publish his 1791 book documenting his work. What is noted, however, is that von Kempelen’s “early forays into synthetic speech stimulated much research into the physiology of speech production and experimental phonetics” (Schroeder, 2004, p. 26).

Paralleling von Kempelen’s work was Christian Gottlieb Kratzenstein’s 1779 entry into the Imperial Russian Academy of St. Petersburg’s annual competition. Kratzenstein, a physiologist, won the competition by providing the best explanation of the physiological differences between five vowel sounds, along with his construction of a model made of resonators that could produce those sounds (Cater, 1983; Coker et al., 1963; Schroeder, 2004). Like von Kempelen’s device, Kratzenstein’s invention was modeled after the human vocal tract, and produced sounds by manipulating airflow over vibrating structures or reeds. It too was not automated, requiring someone to operate it.

By the nineteenth century, Sir Charles Wheatstone built upon ideas from von Kempelen’s machine and the theories purported by Kratzenstein and one W. Willis of Britain. Kratzenstein and Willis independently theorized that machine-generated vowels could come from “different shapes having identical resonances” (Schroeder, 2004, p. 27). Using this knowledge in combination with his musical expertise, Wheatstone made his own speaking machine—a more sophisticated rendition of the von Kempelen device (Cater, 1983).

By the middle of the nineteenth century, Joseph Faber of Vienna built a talking machine that was said to be able to “hold a conversation in normal speech, it could whisper, and it could even sing” (Coker et al., 1963, p. 4-5). According to researchers Coker, Denes, and Pinson, Faber’s machine also spoke with the use of bellows pushing air over vibrating reeds. The machine was fairly large, and required an operator who would manipulate it like an organ or piano.

According to the research of John P. Cater (1983), a young Alexander Graham Bell also entered the realm of speech synthesizer designs during this period. Bell happened to witness Wheatstone’s speaking machine in action in the mid-1800s. What Bell saw inspired him to make his own speaking machine, with the assistance of his brother and his father—an elocutionist. Bell’s machine was in the shape of a human skull “complete with rubber lips and a wooden tongue” (Schroeder, 2004, p. 28). As fate would have it, his work on his speaking machine helped to lay the groundwork for his invention of the telephone, for which he received a patent in 1876.