Chapter 1
Bridging the Gap: The Role of Speech Processing in Enhancing Human–Computer Interaction

Amena Mahmoud
https://orcid.org/0009-0006-9909-0430
Faculty of Computers and Information, Kafrelsheikh University, Egypt

Wael A. Awad
Independent Researcher, Egypt

ABSTRACT

Human-computer interaction (HCI) has evolved significantly with speech processing technologies, yet a substantial gap remains between natural human communication and machine interaction capabilities. This paper examines speech processing's role in enhancing HCI by analyzing acoustic signal processing, natural language understanding, and user interface design. The study investigates key technologies including automatic speech recognition (ASR), text-to-speech synthesis, and speaker identification across diverse applications such as voice assistants, accessibility tools, gaming environments, and interactive voice response systems. Our analysis reveals that while speech processing has successfully transformed HCI applications, critical issues including acoustic noise sensitivity, limited inclusivity for diverse speaker populations, and the "uncanny valley" effect continue impeding optimal user experience.

DOI: 10.4018/979-8-3373-3048-8.ch001

1. INTRODUCTION

Human-Computer Interaction (HCI) stands as a cornerstone of the modern digital era, defining the modalities through which humans communicate with and command computational systems. The evolution of HCI has progressed from primitive punch cards and command-line interfaces to the sophisticated graphical and touch-based systems ubiquitous today. Despite these advancements, a significant chasm persists between the fluid, intuitive nature of human-to-human communication and the often rigid, formulaic interactions characteristic of human-to-machine dialogue (Moore, 2019). This gap represents a fundamental limitation in achieving truly seamless and natural technological integration into daily life.

Speech, as the most primal and natural form of human communication, presents a compelling solution to this challenge. The field of speech processing, which encompasses the computational analysis, synthesis, and understanding of speech signals, has emerged as a critical discipline aimed at harnessing this potential (Denby, Csapó, & Wand, 2023). By enabling machines to recognize spoken words, synthesize human-like speech, and identify speakers, speech processing technologies seek to transform the HCI paradigm from one of manual input and visual output to one of conversational partnership (Polkosky, 2005). The integration of these technologies is evident in widespread applications such as voice-activated personal assistants (e.g., Siri, Alexa), automated transcription services, and interactive voice response (IVR) systems, which collectively aim to make technology more accessible, efficient, and hands-free (Kotian & Nandipi, 2024).

However, the path to realizing this vision is fraught with technical and experiential hurdles. Contemporary speech-enabled systems grapple with critical issues such as a pronounced sensitivity to acoustic noise, performance degradation for non-standard accents and dialects, and a lack of inclusivity for users with speech impairments (Shanthamallappa & Puttegowda, 2024). Furthermore, the phenomenon known as the "uncanny valley" effect, where nearly human-like synthetic voices can evoke unease, alongside profound ethical concerns regarding user privacy, data security, and embedded algorithmic biases, continues to impede widespread adoption and user trust (Aloufi, Haddadi, & Boyle, 2021), (Kulkarni, Tokareva, Qureshi, & Couceiro, 2024).

The current study presents a comprehensive examination of this dynamic intersection. The authors conduct a meticulous analysis of the core technologies underpinning speech processing—including automatic speech recognition (ASR), text-to-speech (TTS) synthesis, and speaker identification—and evaluate their application across diverse domains such as accessibility, gaming, and customer service. Through a synthesis of historical context, current applications, and case studies

Related Content