It's Time to Personalise the Voice:

Designing for the Voice in Voice Technology

Dónal Kearns
Master of Science in Interactive Digital Media

Supervised by Jack Cawley
2020

Abstract

Speech is the primary way of communicating with a human. The past six decades have seen a growing attraction towards developing technology to talk to humans. Consumers have been interacting with speech technology since the 1990s, from automated phone calls to mobile apps run primarily through voice assistants. With the advancements in artificial intelligence and technology, Voice User Interfaces are becoming globally accessible. Many users have tried voice recognition on their mobile phone, after it fails however, they never try the system again. The key to user centred design is to understand why this is the case. The primary concern of this research however is to explore the voice in voice technology.

When a device talks, what should its voice sound like? Most voice assistants share the same characteristic; a one voice fits all approach. A polite and friendly female voice being the most common. This research will demonstrate the social consequences of designing a voice and how personality, gender and language can have a major effect on how users perceive the technology. Voice alone is bounteous in social information. Voice sounds can convey various signals that humans can pick up naturally to identify, such as personality, age and gender of a talking human. The same applies when humans talk to computers, as the "Media Equation" points out that humans treat computers as real people and can have real social relationships (Reeves and Nass 1996).

Voice Cloning has recently emerged in the past year and may be on its way to being implemented into consumer devices and homes. Voice cloning is a deep-learning algorithm which can record a human voice and synthesize it to match the original voice. LyreBird, Resemble AI and iSpeech Deepsync are some applications using the voice cloning algorithm. Some applications can now mimic a human's voice within only a few minutes of recordings, which improves naturalness and similarity. As my research has pointed out, while voice cloning can create a more personalised and natural experience for the user, voice identity can be a very sensitive topic and can cause serious problems if not developed correctly. If the technology is not designed correctly; it can lead to vulnerable users who will not be confident in using the technology again. More user-centred research needs to be done to truly understand how users will react to this technology.