Introduction: The history of the talking machines raises foundational questions on humanity and reminds us on the power of imagination

By Pertti Grönholm

Whether its speech, singing, or shouting, the sound emanating from within us is one of the most fundamental characteristics of the human species. The human voice becomes part of our perception of the environment already in the fetal stage and forms a central part of our sound environment on the important stages of brain development during the first months and years of life. The advanced use of sound, combined with hearing and linguistic skills, is human's most important means of communication, and has evolved into a very subtle device that is difficult to imitate mechanically.

People will easily notice the voice if there is something strange or suspicious in the interlocutor. The human voice tells us not only about our thoughts, but also about our emotions and state of mind, the state of alertness and health of the body, and our attitudes towards the environment, the interaction itself and the interlocutors. The voice contains information about its producer's physiology and anatomy, the gender and age, even ethnicity. In addition, many cultural idiosyncrasies reveal themselves in the rhythm, intensity, pauses, emphases, tone and pitch of the speech.

In addition, singing in its various forms is even today a fundamental means of human self-expression and, among other things, a powerful way to strengthen the sense of intimacy, belonging and unity. Despite all other means of human-machine communication, the importance of the human voice in both everyday life and long cultural evolution cannot be underestimated in the development of future speech technologies either.

The long history of imagined man-machines

In folk tales, literature, and many other cultural products, human beings have built human-like machines for centuries. The imagining of automatons (by Hephaestus) and beings created from bronze, gold, or clay in epic tales such as Golem and Kalevala, tell us that the idea of creating a human-like living being has fascinated the human for many centuries. Intelligent and sentient automata as well as "mechanical people" have been predicted in early European science fiction since Frankenstein's monster (Mary Shelley: *Frankenstein*, 1818) and Olympia (E.T.A. Hoffmann: *Der Sandmann*, 1814). On the philosophical side, many intellectuals have contemplated materiality and mechanisms of humanity utilizing the machine metaphor since the 18th century, a good example of which is Julien Offray de la Mettrie's *L'homme machine* (1747).

The idea of a human-like machine has been cultivated in fiction and entertainment for a long time, despite the limitations of actual technology. Artisans and inventors have built dolls and vending machines that play chess, smoke cigarettes, talk and move, many of which actually operated on human intelligence and muscle power. Since as early as the 18th century, watchmakers, inventors and

engineers have also tried to imitate human vocal cords and the vocal tract, first mechanically, then electro-mechanically, and eventually completely electronically since the Second World War.

The development of sound transmission, telephony and recording technologies, such as the invention of the radio, telephone, phonograph and gramophone, as well as sound film and tape recorder, has also paved the way for the mechanical human voice since the mid-19th century. Disconnected from its time, place, and speaker, the human voice was one of the great wonders of the modern world in the early 20th century. During the 20th century, electronic aids gave a voice to people who, due to their innate characteristics, accidents, or illnesses, have not been able to speak with their own vocal cords.

During the Second World War, the military experts learned filter human speech electronically, with a voice encoder (vocoder) developed by Homer Dudley in 1928. Among other things, this made it possible to send encrypted messages from one continent to another via a telephone cable. The technology of electronic coding and decoding the voice was used e.g. US President F.D. Roosevelt and British Prime Minister W. Churchill in their bilateral talks. Later, since the early 1970s, vocoders have been used extensively in music, television, and movies. Vocoder can make the musical instrument sound like a human speech or singing by modifying the instrument with human vocal formants.

Machines began to talk and sing

When the first digital computers came into use in the 1950s, it did not take many years for engineers and researchers to start writing programs enabled computers speak text in an understandable manner. Moreover, very soon computers were taught to sing, like the IBM-7094, which went down in history in 1961 with the song ‘Daisy’ it sang. The fictional HAL-9000 computer performed the song in *2001: A Space Odyssey* (1968) which explored the relationship between human species, technology and the mysteries of cosmos. HAL-9000 communicates with its users mainly through speech, but the naturalistic male sound (actor Douglas Rain) also hid behind a paranoid, fearful, and monomaniac artificial mind. Some tech journalists have concluded that HAL’s voice and persona were so discomforting to viewers that they effectively restrained the realistic male computer voice becoming more common in talking devices and services in the 1980s and 1990s.

Sound synthesis and speech coding and editing techniques were developed in the 1950s and 1960s, especially in West Germany, Britain and the United States. Also in Finland, the development of electronics and information technology was rapid, which affected culture and the arts. An example of this was radio journalist Martti Vuorenjuuri, who produced with his team a radio play adaptation of Aldous Huxley's famous novel *The Brave New World* (1932). Olavi Paavolainen, the head of YLE's Radio Theater, commissioned the work for the tenth anniversary (1958) of the radio play department. Initially, the play was to include electronic music. Vuorenjuuri himself was well acquainted with the development of European electronic music in the late 1950’s. However, music made with voltage

controlled audio oscillators, electronic filters and modulators was quite difficult to implement in Finland at that time, and in the end the whole soundscape of the radio play — including actors' lines — was created by modifying the human voice with the help of tape technology and various electronic devices. With this solution, the dramatization of the work became even more futuristic and dystopian.

Machines with a robotic voice became a theme of its own in Western (and Japanese) popular culture during early 1970's. Especially in the wake of the disco, funk and electronic pop, vocoders with a mechanical-sounding human voice, and even speech synthesis became somewhat trendy in pop music. In the screen and television, actors started to imitate the voice of talking machines especially after the blockbuster movie *Star Wars* (1977). Often, talking machines were portrayed in science films as a threat to humanity, but the *Star Wars* robot pair R2D2 and C3PO — whether they spoke human or machine language — represented a more humane approach to technology. Allegedly, the sympathetic fictitious androids contributed to the idea of real talking machines. At the same time, human-like robots i.e. androids and gynaikoids, began to appear in commercials and television series.

In addition, the development of speech synthesis was rapid in the late 20th century. In recent decades, very different techniques have been used to produce speech. Some of them chain recorded spoken words and syllables and some produce pure electronic human-like voice: either the algorithms model the various acoustic elements of speech or they simulate the entire human vocal tract, articulation and prosody.

Speech interface becomes more common

Speech as the most fundamental form of human communication is not only the most natural interface for a person to another intelligence. However, it is also a form of communication with numerous possibilities of misunderstanding, incorrect interpretation, factual error and manipulation. There are already many signs of the evolution of user interfaces to voice-based ones. For example, the voice messages and dictation of text messages on the phone are part of our daily life.

Today, the algorithmic speech is very difficult to distinguish from human speech by the mere qualities of the voice or the rhythm of speech. Instead, the information, the articulatory and linguistic richness, and internal logic of speech are factors that still reveal the identity of the speaker, at least in a longer conversation. In the future, the questions concerning the identity and credibility of the speaker will relate more often to the quality of algorithmic speech recognition and the linguistic nuances and general intelligence of the algorithmic speaker. Despite these challenges, speaking digital assistants from software companies and phone manufacturers are already commonplace on smartphones, computers, and e-commerce sites.

In Western culture, the idea of an intelligent talking machine has long fascinated many artists, moviemakers, writers, visionaries, engineers and artisans. Both the imagined and real machines have a millennial history, in which future thinking, technical expertise, entertainment, business prospects

and utility thinking intertwine. In addition, the visions and devices have brought up wider questions about what it is like to be human, what are the boundaries of humanity, and how and why we use technology to imitate human body and its abilities?

Nowadays, the most advanced robots and other talking applications empowered by A.I. imitate human speech so well that the area in which humans experience uncanny feelings and weirdness has changed. While in the late 20th century, machine speech that was composed of spoken words and syllables raised suspicion and feeling of alienation, nowadays machine speech is much more convincing. Not only because of the drastic improvement of speech algorithms but also because of cultural changes; people in technologically most advanced societies are aware of the possibility to encounter a conversation with talking machines.

What does this mean for the development of the social machine? Does the voice no longer reveal whom the person is interacting with? It is also necessary to ask what kinds of identities are built on machines that imitate people, what kind of concepts, values, and attitudes their algorithms represent and convey, and what kind of image of humanity they constitute.

Voices from the uncanny valley

The desire for a machine that understands speech is now beginning to come true. Visions that go even further are presented in the fiction of our own time. A good example of this is the film *Her* (2013), directed by Spike Jonze, in which a lonely young man falls in love with the talking operating system Samantha installed on his computer as a prototype.

However, the better the robots and audiovisual services can imitate real spoken conversation (and human face expression), the more emotional obstacles, such as the feelings of eeriness, oddness and aversion seem to occur. The uncanny valley is a hypothesis presented by Masahiro Mori in 1970 that aims at describing the changes in the human emotional response to an artificial object that tries to imitate the appearance and behavior of a human being. The hypothesis suggests that objects, which imperfectly simulate human beings, provoke uncanny feelings within the humans they encounter. The valley means a steep slope in the observer's ability to feel affinity for the object. The uncanny valley hypothesis is fascinating, because the slope seems to appear only in the stage of a relatively advanced human-likeness. Furthermore, there have emerged numerous theories proposing explanations to the cognitive mechanisms behind the uncanny feelings. These theories range from mate selection theories to pathogen avoidance and from religious ideas on human identity to conflicting cognitive representations.

In addition to the effects of the uncanny valley, many people are critical towards new technology for various reasons and it is right here where the audible dialogue between the human and machines settles at a very important seam. New robot and A.I. technologies do not only promise positive outcomes for the people living in technologically advanced societies. Currently, we witness the

expanding use of A.I. and human-like technologies in surveillance, spying and propaganda. In addition, our own skeptical, ironic and sometimes even hostile attitudes towards human-like machines continue to alert us. These attitudes may partly stem from the late 20th century critique on the technology-driven modernization, which has been a very strong undercurrent in fiction and media in the Western world. In the 21st century control, manipulation and distortion of digital information for political purposes have become so commonplace that they also will have profound effects on our relationship with talking and listening machines.

Captions to images:

Bell Laboratories introduced the world's first electronically emulated speech synthesizer at the 1939 New York World's Fair. Controlling the device required good coordination and training from the user. (Wikimedia Commons)

Finnish Synte2 was the world's first portable speech synthesizer for the impaired. It was developed in VTT's (Technical Research Centre of Finland Ltd) Hospital Technology Laboratory and at the Department of Electronics at Tampere University of Technology in the late 1970s.