Tracking down speech signals

Background noise, reverberations and acoustic feedback often hamper speech comprehension when technical devices are in use. Hearing researcher Simon Doclo uses mathematical methods to tackle the problem

n the legendary Star Trek series, devices called "communicators" allowed people to communicate directly with each other anywhere in the universe in a fraction of a second. That was in the late 1960s. But although interstellar communication is still a distant dream, technology – mobile phones, teleconferencing and hearing devices – has come to play an increasingly important role in everyday communication.

Yet despite major advances, the technology that helps us communicate with each other still has its shortcomings. If you have ever tried to make a phone call from a noisy train station or listen to a person at the other end of the table in a busy restaurant, you know the problem: background noise and reverberation make it difficult to understand what is being said, even for those with good hearing, not to mention people who depend on a hearing aid. Experts call this the "cocktail party effect", a term that was coined back in the 1950s.

This is the starting point for the research of Prof. Dr. Simon Doclo, who heads the Signal Processing Group of the Department for Medical Physics and Acoustics at the University of Oldenburg. "Our aim is to improve speech communication in adverse acoustic environments when devices such as mobile phones and hearing aids are being used," he says. Doclo's team uses mathematical methods to meet the challenge that the Oldenburg researchers in the Cluster of Excellence "Hearing4all" and the Collaborative Research Centre (SFB) "Hearing Acoustics" approach from a wide range of perspectives: the team develops algorithms – in other words sequences of instructions for calculations – which can be used to cancel or at least suppress effects that reduce speech intelligibility and speech quality during communication.

The Belgian-born scientist has been researching and teaching at the university since 2009 – a time, he stresses, when hearing research in Oldenburg had already established an international reputation. Doclo's research is generally a three-step process, the first step being the design of a new audio signal processing algorithm. "In the next step we implement and optimize the new algorithm and evaluate it using computer simulations to see whether it achieves the desired effect, for instance improving speech quality by a certain percentage," the electrical engineer explains.

To obtain more than just technical results, in the final step the researchers determine whether the new algorithm actually works in practice by means of listening experiments with test persons – a time-consuming process, which is often carried out in cooperation with other groups in the SFB and the Cluster of Excellence, Doclo emphasizes. But what may sound like a routine task is actually highly complex, because working with mathematical methods to filter out intelligible speech from complex acoustic sound information is often an extremely laborious process. "Only a handful of the algorithms that we develop turn out to be good enough to use in the final stage with test persons," Doclo says.

In his work, the 46-year-old takes advantage of the fact that most devices, including hearing aids, contain several microphones. This allows the researchers to filter out more than just the so-called spectral components from an audio signal – specific frequencies with more or less background noise, for example. "The different microphone signals also provide information about the positions of the sound sources and how sound propagates within a room," Doclo explains.

"The 'holy grail' of acoustic signal processing."

To develop what he calls "smart algorithms", the team takes two different approaches: "On the one hand, we work with classical methods of digital signal processing, utilizing statistical models of speech and room acoustical properties," he says. This means that the researchers use statistical methods to describe for instance the frequencies of an acoustic signal, that is the sound's oscillations, and how these frequencies change over time.

Based on such time-frequency analyses, the optimal parameters of a mathematical objective function are then estimated, for example, aiming at extracting the clean speech signal from noisy and distorted recordings. "One of the main challenges is to design a suitable objective function that it is on the one hand relatively easy to optimize but on the other hand also incorporates psycho-acoustic properties of human hearing," Doclo says. Furthermore, the algorithms should preferably work in a fully blind way. "In well-defined experiments in the lab the positions of the sound sources or the microphone distances may be assumed to be known. But in real-life situations, these variables are often unknown," he explains.

With this approach the researchers' work of the last five years has led to major progress in several areas, most notably how to deal with reverberation. Reverberation is a term used by experts to describe the repeated reflections of an acoustic signal that occur when the sound encounters obstacles such as walls and is reflected back multiple times. "This is the 'holy grail' of acoustic signal processing, because it is extremely hard to separate the reflections from the clean acoustic signal," Doclo says. "We developed a new method which is able to filter out reverberation much more effectively than previously possible." This work attracted considerable interest among experts, and at the end of 2019 Doclo and his colleague Dr. Ina Kodrasi were awarded the annual publication prize of the Informationstechnische Gesellschaft (ITG).

In addition to statistical model-driven methods, the team works with data-driven methods. These methods are based on "machine learning", in which the scientists train so-called neural networks using large

amounts of data. "Basically, we feed a huge quantity of audio signals that we have either recorded under controlled conditions in the lab or simulated on the computer into the network," Doclo explains. The researchers can, for example, record speech and background noise separately and thus specify the clean speech signal as the desired result. The network then has to learn how to extract the clean signal from the noisy data.

But Doclo emphasizes that many data-driven methods are often "black boxes" which don't contribute to understanding the underlying processes. Statistical model-driven methods are generally more helpful in this respect, he believes. "This is one reason why we aim at combining the advantages of both approaches by merging statistical model-driven and data-driven methods," he says. In this way, the researcher hopes to achieve the best possible balance between an algorithm's performance and its robustness. By robustness, researchers mean the ability of an algorithm to generalize as much as possible, that is to be applied to many different situations. "After all, our algorithms should also function for unknown acoustic environments and unknown noises," Doclo says.

Another challenge for the team's research is that all their approaches have to work in real time, because speech communication devices such as hearing aids and mobile phones should not noticeably delay the signals for their users. "Therefore, we need to pro-

cess the signals as quickly as possible when they reach the device, extracting a clean signal within milliseconds," Doclo explains. This is another reason why the researchers need to take care that their algorithms don't become too computationally complex.

The team also uses their algorithms in so-called acoustic sensor networks, where information is used from several spatially distributed microphones, for example those in a hearing aid and a mobile phone. In the future, in a noisy restaurant setting for example, hearing-impaired people as well as people with normal hearing could simply place a mobile phone on the table and its microphones would work together with the hearing aid microphones. "All available microphones then analyse the surroundings acoustically," Doclo explains. "And even when the positions of these devices are not exactly known, our algorithms are becoming increasingly effective at extracting the desired speech signal in this type of situation."

But mathematical methods also have their limits: "In many applications, we first analyse the acoustic environment in order to determine all relevant sound sources and their location," Doclo says. Experts call this step computational auditory scene analysis. However, purely based on the acoustical signals it is not possible to decide which sound source the user wants to listen to. For this, an entirely different type of information is needed. By working together with the Oldenburg neu-

1 In their experiments, Doclo and his team use artificial heads designed to simulate the acoustic properties of a human head.

2 The researchers play back certain sounds over a loudspeaker. The microphones in the ears of the artificial head record these sounds. The result is a recording that authentically replicates what a person in the same position as the artificial head would hear.

3 Using sound cards, the scientists record different microphone signals and then use these signals in further experiments.

4 Microphones operating in tandem, known as "microphone arrays", are used in acoustic sensor networks. Scientists use them to analyse the acoustic environment of a room, for example.

ropsychologist Prof. Dr. Stefan Debener, one of the researchers' long-term objectives is to use EEG measurements to help determine which sound source a person focuses their attention on, in order to amplify this source.

"We want to make things better."

Another tricky problem that Doclo wants to solve mathematically is acoustic feedback. Anyone who has ever put a microphone too close to a loudspeaker to which it is connected will have heard the typical screech of a feedback loop. This acoustic phenomenon also occurs in so-called "hearables". These ear-worn devices go far beyond simple earphones. They use digital signal processing algorithms to provide individualized hearing support for people with normal hearing to, for example, better understand a conversation in a noisy situation. A number of research groups at the department and at the Division Hearing, Speech and Audio Technology of the Fraunhofer Institute for Digital Media Technology (IDMT) are currently working on optimising hearables.

"In the Collaborative Research Centre "Hearing Acoustics" we have developed a new hearable, which contains two or three tiny microphones in a small earpiece," Doclo explains. But because these microphones are located very close to one of the loudspeakers, the researchers have often grappled with feedback issues. "We have now developed a comparatively simple method to suppress this feedback," the electrical engineer explains. And it works much better than we could ever have expected." This came as a surprise and we were thrilled."

Doclo sees plenty of other interesting challenges for the future. The first item on the list is to further optimize the algorithms by combining statistical model-driven and data-driven methods. "We are engineers," he says, "we want to make things better." In the long-term he can imagine opening up his mathematical methods for further applications – in medicine, for example." This is another field where sensors are distributed across the body and deliver distorted signals." As in acoustics, the algorithms could help to filter out the desired information.

But his true passion is acoustics, Doclo says – from hearing aids to mobile phones to smart speakers. His dream is to develop algorithms that are truly robust and function seamlessly. "The Star Trek thing," he says, with a smile. "You can run around, talking to everyone, even over huge distances, without any annoying background noise." (cb)

Simon Doclo has been at the University of Oldenburg for more than a decade. Among other things he has developed a method for filtering acoustic signals to remove reverberations.