



I can't hear the TV!

The future of [intelligible] broadcast audio —
MIKE KAHSNITZ, head of product development, RTW



/ Mike Kahsnitz

As many manufacturers in the industry focus on developing equipment to create, record and edit sound for TV and film, we've been behind-the-scenes analysing the final products. What we've found at RTW is that many television viewers are having an increasingly difficult time understanding the spoken word of this art form. Immediate causes include actors not speaking clearly — be it articulation or use of slang — or inclusion of too much background noise.

But, the truth of the matter is, this inability to hear or understand the actors — speech Unintelligibility, if you will — actually relates to so much more than the viewer. So, together with Fraunhofer Institute of Digital Media Technology IDMT, which is conducting applied

research in the field of audio-visual media, we've set out to find all the causes, methods of analysis and solutions for correction.

What's the problem?

It's not so straightforward, and the more we've dug into it, the more complicated we've found it to be. Intelligibility is not contingent on each individual's hearing, each actor's speech, the quality of each recording, the process by which each project is edited, or the type of consumer-grade speaker being used. Audio pros and laypersons, such as journalists and producers, must learn to trust a measurement, even if they consider what they are hearing to be intelligible despite proof of the opposite. There are hundreds of factors that come into play, so we must look at every piece of the workflow.

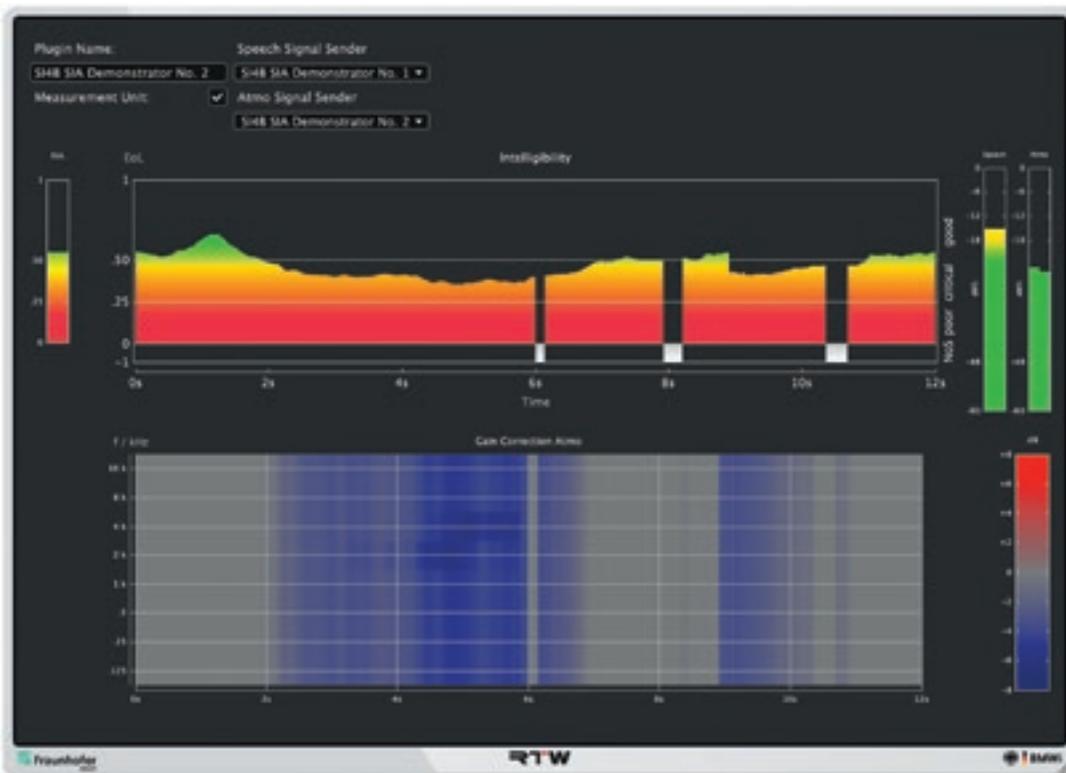
It goes without saying that budgets weigh heavily on all productions. Often, scenes must be captured in one or two takes, which means audio mixers must get the best sound possible

in those small windows. This is typically achieved using the least amount of equipment possible, unless production freelancers are supplying their own gear — but even then, they're working on a tight budget. The team might want to use additional microphones to ensure the best audio capture, but with a growing emphasis on 'artistic' filming styles, directors are opting for more cameras and less boom microphones to guarantee a clean shot. In field productions, audio crews are faced with the unavoidable issue of background noise.

What's more, production teams must turn out completed projects at faster rates and with significantly more compression due to decreased staff, high wireless mic counts and little to no time for sound checks. It must also be noted that a sound engineer can control and monitor only a limited number of channels on the set. In these cases, the safest way to record is to run mics with high gain to catch low-level audio, with a trade-off of dynamic compression so as not to overload the recordings with the high-level content — but high dynamic compression is not beneficial to intelligibility.

Subjective misjudgement

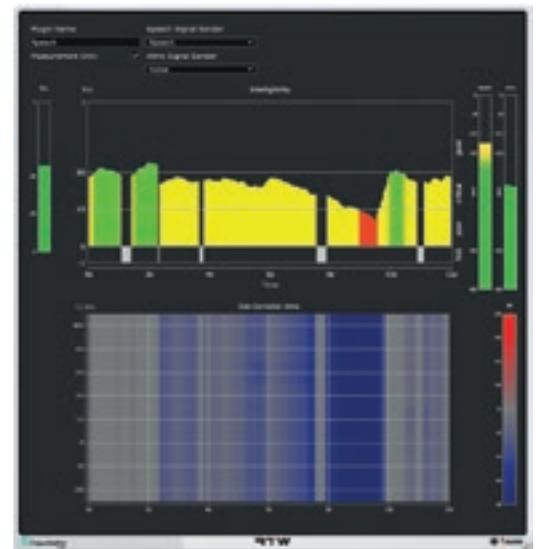
Regardless of the production and recording approaches, mixing adds another level of potential intelligibility problems. In modern filming styles, directors and editors often incorporate a significant amount of background



music or audio behind the speech, which typically causes those with sub-par hearing to misinterpret information. However, the more often someone listens to a piece of audio, the easier it becomes to understand what is being said. Even if you don't hear the exact wording, you still get a sense of the overall message after repeated listens, so the unintelligibility becomes less apparent.

The subjective misjudgement of intelligibility can be seen (and heard) with this very simple example. In a large venue, which usually has a lot of reverberation, place two people on opposite sides of the room. Have one speak a short sentence — that is unknown to the listener

— with normal sound volume and the listener will not understand what's being said. Cut the distance in half and try again — now they may decipher some words, at least you think so. Then, have the listener step right in front of the speaker and they'll fully understand the spoken words. Afterwards, have the participants move back to max distance and the listener should be able 'understand' the sentence. However, what's really happening is that the listener's brain already knows what's being said, so it interprets the words as being intelligible or understandable. This describes the situation with ambient influence on speech like reverb — but there's more.



/ Prototype software measuring Speech vs. Atmo

In Foley, editors create emotions by using sounds. The dynamics of background effects may kill intelligibility for impaired listeners and might even affect the intelligibility for a normal listener. Finding the perfect balance between background noise and speech may present a whole other set of problems. Let's focus for a moment on voiceovers: specifically one language over another, in the context of ADR. With two languages running simultaneously, a human who understands only one will likely be able to decide which of the voices he needs to follow, at least to a certain degree, but this isn't always as easy for multi-lingual people. However, an ADR system can't help alleviate this successfully in auto mode, so intelligibility measures will be limited on this front. Significant research must still be done for this avenue. Our hope is that, with the growth of artificial intelligence (AI) technology, we might find a way to measure this in an objective matter.



What about the listeners?

Even with a perfect recording, mixing and mastering environment, filmmakers still cannot control one very important aspect: the viewers. At home, most people have poor speakers. This is not due to a lack of want or understanding of this gear, but average consumers don't opt for high-end speaker systems. Instead, they rely on the built-in speakers on their televisions which, let's face it, are not the best. In fact, as the screen gets bigger and bigger, the speakers are exponentially becoming smaller and smaller. To



further complicate the listening environment, many built-in speakers are not even facing the listener, they are facing the wall. This creates reflections, and sometimes a muffling effect, both of which are bad for intelligibility.

The overall method of video consumption also affects intelligibility. A lot of people today, especially younger generations, use mobile devices with ear buds, which are generally poor in audio quality. The number of outlets from which someone can consume media is growing

at a rapid rate, which has led to a degradation of quality of the final product. Quantity over quality, if you will.

Part of our research and analysis has been focused on who's listening to what, and how? For instance, we looked at the average age of people watching traditional television versus that of people consuming media on mobile devices. What we found was that the average age of people watching broadcasts in their homes is age 55 or higher. We also found that older generations have a natural deprivation of their hearing ability, which makes it that much more difficult for these viewers to understand what is being said. This is only going to get worse as technology — and people's methods of consuming media — evolves.

With funding from the German government, our friends at the Fraunhofer Institute have done significant research into all the factors that contribute to speech intelligibility, which will aid in the development of technology to help rectify the problem.

It's safe to say we will potentially see an increase in the number of impaired listeners in the future because of the levels at which they listen to audio through ear buds. This technology, combined with the high-volumes and low-quality speakers, will have an effect on hearing abilities over time. With continued research into speech intelligibility, we can better alter our audio levels on the production and

**Not getting the newsletter?
 Subscribe now!**



Sign up at www.resolutionmag.com/newsletter for the all-new, fortnightly *Resolution* email newsletter, featuring exclusive *Resolution* features and content, news and competitions, and more!

resolution/