Oticon Cue Preservation Technologies
Recorded November 19, 2019

Presenter: Don Schum, PhD
AudiologyOnline.com Course #33847
Partner: Oticon

- - [Donald] Hello, everybody, this is Don Schum with Oticon. And I wanna welcome you to this seminar on Audiology Online. The title of this seminar is Oticon Cue Preserving Technologies. And in this seminar, I wanna talk about the way we approach signal processing in the area of trying to make sure that we are providing the listener, providing the user of amplification as much of the natural information in the sound environment and especially in the speech signal as possible. So let's go ahead and get started. Let's start by taking a look at the learning objectives for this course. What I wanna do is spend a little bit of time talking at the beginning about how the cognitive system processes sound information, especially speech information, and how important it is to provide the cognitive system a full set of cues in order to let the cognitive system do as much as it possibly can. The second objective is to talk specifically about spatial information and about how one of the most important tasks that the cognitive system does for a listener, especially in a complex sound environment, is organize that sound environment then before the listener will hone in on specifically what that person wants to listen to. And spatial information is not the only cue you use to organize the sound environment, but it's a very important cue. And it can be, in some ways, disrupted by amplification, so I'll talk about that. And third, I wanna talk specifically about the way compression approaches can affect some of the natural cue sets in the speech signal.

We use nonlinear amplification, multichannel nonlinear compression, in hearing aids very widely, basically, almost all hearing aids that are sold and fit on patients use some form of wide dynamic range compression for very important reasons, very good reasons, but it's also important to understand how those systems work and how, in certain forms, they can potentially disrupt the information that the listener can use to do the job of understanding speech, especially in complex environments. If I start with the overall statement of why, why we feel we need to do a course. It's important to recognize that a lot of times, when manufacturers talk about their signal processing, they want to jump right into the details of the way we try to manipulate sound in order to make it easier to understand. Noise reduction, directionality, things like that, speech enhancement approaches, all the things that we do to try to somehow manipulate the signal to make it work better for the patient with sensorineural hearing loss. But it's also important to understand that, although conventional amplification can do a very good job of improving audibility and improving the signal to noise ratio for the patient, it's important to recognize that there are some downsides to the way we have historically approached creating amplification for patients. In other words, you can't manipulate things without changing them. Obviously, manipulation means changing something. But all the changing isn't necessarily good for the system. And sometimes, you might inadvertently be changing the signal in a way that is not necessarily good for the way the listener would perceive information, because you're trying to achieve another goal. So I'll talk about this in terms of spatial information, I'll talk about this in terms of bandwidth, and I'll talk about this in terms of compression systems to show that there's both an upside and a downside to these systems.

And of course, when we create amplification solutions for patients, we wanna make sure that we improve audibility, obviously, they have a threshold loss and that has to be overcome. And also, we know that the nature of sensorineural hearing loss is such that patients with sensorineural hearing loss have trouble in noisy environments. So anything we can do to directly improve the signal to noise ratio is a very good thing for the patient. But it's also important to recognize what we inadvertently may have changed about the signal and ways that perhaps somehow we can overcome that. At Oticon, we talk about BrainHearing. And what BrainHearing is is a term that captures our approach to creating solutions for patients with hearing loss. We understand that the very best signal processor that we get to work with is the human brain. The human brain can do a tremendous amount of very good culling through information and allowing a listener to really understand the different sounds in the environment and really hone in on the speech signal that they're interested in and still participate in that sound environment. In the presence of sensorineural hearing loss, a patient obviously is at a disadvantage because of the nature of the hearing loss. But we never want to lose focus on the fact that we always have the brain that we get to work with. And so when we create signal processing approaches, our goal is not necessarily to try to do all the signal processing at the level of the peripheral auditory system. What we wanna do is do the right kind of signal processing at the level of the peripheral auditory system so that, when we feed that signal into the central auditory system and it gets to the cortex, we allow the brain to do what the brain does very well. The cognitive system can do amazing things if it gets good information. So when we create signal processing approaches, our goal is to make sure that we're providing the brain with the very best information that the brain needs to get the job done. And when we think about things in terms of brain hearing, it's not just the manipulations and the improvements that we can do to the signal, but it's also to make sure that we avoid some of the downsides that conventional amplification could potentially do to the signal.

And that's why we talk about cue preserving technologies, is to make sure that we do the very best job of making a good signal available to the brain, but without stealing information that it might use in some way, shape, or form. It is the fundamental aspect about brain hearing is that speech understanding is a cognitive process. Speech understanding doesn't occur in the peripheral auditory system. Speech understanding occurs at the level of the cortex based on the information that can be fed into that cognitive system. And obviously, we are dealing with an impaired auditory system, impaired peripheral auditory system. And we have to somehow manipulate that signal before it's fed through that impaired system. But if we do things carefully and intelligently, we can hopefully, even though we have to work through that impaired system, still provide the brain some very important information that it can deal with. In order to understand the cognitive nature of speech understanding a little bit more, one of the things that we like to remind professionals about is the way the cognitive system handles complex environments. So when a person is in a complex environment with multiple sources of sound, especially if there's more than one source of speech. Maybe it's sitting around a dinner table with a large family, maybe it's at a restaurant with a great group of friends or colleagues, maybe it's at some sort of other gathering, maybe it's at a workplace in a noisy office. It's a variety of cognitive processes that happens. And if the assumption is that you're trying to listen to one particular person or a couple different people, one of the first things that happen is an orientation towards who you wanna listen to. So the idea that there's one particular source of sound or maybe a couple different sources of sound that you really wanna pay attention to. So it's important to orient and organize that sound environment so you know where the different sound sources are coming from, so you can choose which ones you wanna pay attention to. People with normal hearing do that all the time, and whether or not you're really aware of it, any time you walk in a situation that is not completely quiet and you're trying to listen to one person talk, there's other sounds that are competing, but your cognitive system does a very good job of organizing that environment and suppressing those other sounds.

So this first step is orientation of the environment, then there's a separation of the different sounds that are out there into different sources. In other words, one of the things that the cognitive system loves to do is attach meaning to sensory stimulation. Sensory stimulation without meaning doesn't do the cognitive system any good, the cognitive system's job is basically to keep the person alive. And that's a notion that dates back to the early days of man when they had to fend off predators and things like that. As the world's become a little bit more easy to negotiate, there's this idea that the cognitive system wants you to be immediately aware of meaningful things in the environment. And one of the most meaningful things when it comes to sound, of course, is speech, we communicate in a great way with speech, and so we wanna make sure that the person immediately sees speech as a signal that carries information. So you wanna separate out these different sources of sound so that you have a better chance of focusing on what you want to and somehow tamping down or muting or somehow, the best you possibly can, ignoring those other sources of sound. Once you organize the sound scene, separate it into different tracks of sound, then that allows you to focus on one particular source of sound or maybe multiple sources of sound at the same time, depending on how you want to divide up your focus. Sometimes, you can monitor one talker but sort of monitor another talker, or sort of monitor other things happening in an environment, for example, let's say you're having a conversation while there's a good football game going on on the television. You can carry on that conversation, but with one small part of your awareness, you're also monitoring what's going on in the game, so if all of a sudden, something exciting starts to happen in game, you can switch your attention over to that. It happens all the time, and especially if you have normal hearing, you have that ability to separate into different sources and to focus on what you want to, but change your focus as time goes on.

And then obviously, once you've organized that environment, that allows you to recognize the speech signal that you're listening to and make use of that speech signal and to do with speech what listeners wanna do, which is to use speech as a communication, a vehicle to understand what somebody else is thinking, what they're trying to get across to you as a message, be able to formulate a response, be able to commit it to memory, all the things that you do when you have a conversation. So anyway, the reason I bring that up is just to remind you that the cognitive nature of speech understanding covers many different aspects. This is a nice little depiction of some of the things to think about of the way we pay attention to how the cognitive system works. So imagine over to the left that those different shapes are different sounds that are occurring in an environment, it could be, you can use the same description for visual situations or whatever, but assume that those are different sounds. What you want to get to is what's happening over on the right, which is figure ground. Meaning that you want to pay attention to those sources of sounds that are happening over on the left that are meaningful to you at that moment in time. And basically somehow put the rest of the sound sources in the background. Mute them down, put them on squelch mode so that they're there but you're really not paying attention to them. That's what you wanna get to. And what happens is that process of taking this very complex auditory input and creating figure and ground, that happens because we have a very good cognitive system that does a tremendous amount of organization for us mostly below our level of conscious awareness. So that middle part of this diagram where there's grouping and then you get different sources of sounds that tend to hang together, so that would be depicted here as PO one, PO two, et cetera, et cetera, et cetera, that organization of sound into different tracks or different streams of sound is what your cognitive system does, and like I said, most of that happens below your level of conscious awareness. Once it's organized, then you can consciously decide what parts you wanna pay attention to and which parts you wanna suppress. But you always have that ability to switch what you're paying attention to because you've already grouped it, but you placed it into some lower level of conscious awareness, maybe even a subconscious level.

So for example, if you are having a conversation with somebody at lunch or somewhere else and the conversation is lagging but you're still watching this person and you're politely nodding and smiling at the right moments, but there's a great conversation going on at the table next to you about something that you find far more interesting, you can actually, if you're aware, I'm not saying that you should do it, I know it's a little bit maybe perhaps impolite, but you can actually continue to look at this person and nod and smile every once in a while, but actually put more of your attention focused on this other conversation that you find more interesting. And if you have normal hearing and good normal cognitive function, it's something that you can do pretty much when you decide you wanna do that, it's just the way this system works. And the reason I bring this up is that most of what we're going to be talking about through the rest of this show is what happens there in the middle. What happens there, what allows your cognitive system below your level of conscious awareness to organize all that different sound and create different streams of sound, and then decide which ones you're gonna focus on and which ones you're gonna ignore. And it's all the subtle cues that the cognitive system can use to allow you to track certain streams of speech and suppress other streams of speech. For example, here is about one and a quarter second of speech of a single voice. So what you see is several phonemes being created during that time period. And one of the things that you should know, first of all, let me orient what you're looking at, so this is a spectrogram. And this is frequency going up the Y axis. And so that goes up to about, let me blow it up so I can see what it goes, it goes up to about, I think about 8000 hertz up on the Y axis. And this is time going across the bottom. The lighter the color, the more intense the sound. So the real dark dark blue areas are areas of silence. And then the lighter blue are areas of intensity of speech.

So for example, right, let me pull up my pointer, which is right here. So for example, this area right up in here, that's a high frequency voiceless phoneme, that's probably an S or an SH probably. If you notice that there's no energy going down here in the low frequencies at that point in time, which means it's a voiceless sound, there's no harmonic structure to this, so this is a unvoiced consonant, like I said, either an S or an SH, that occurs up there, so that's a very clear phoneme. Here is the second formant, going from perhaps either one vowel to another, for example, in a diphthong, or maybe going from a voiced consonant to a vowel. But whatever you see that cue changing over time. Speech often changes over time, as you know, as you are moving your mouth to create different sounds. And that movement of your mouth, the movement of your tongue while you're talking, will be reflected in a variety of different cues, and one of them are the formant tracks. So you see that's the second formant moving across there. That's the third formant dropping down. And that's information that the cognitive system would use. But anyway, this is the track of a single voice. And one of the reasons I throw it up there is for you to understand just how complex acoustically speech is. You have phonemes lasting maybe 20-30 milliseconds sometimes at the most. You have them happening with information or at least energy in various different frequency regions at the same time. Sometime, you have them happening with information in only one particular frequency region. Sometime, you have information in multiple frequency regions at the same time. All that is information that the cognitive system uses to identify individual speech sounds and then create some sort of meaning out of those speech sounds. If we move to this graphic now, what you're talking about from here are three different voices happening at the same time. So this is about three seconds worth of speech. And there are three different voices intermingled in there. And what the cognitive system can do, and I'll prove it to you in just a second, what the cognitive system can do is take that very complex acoustic input where you have different sounds occurring in different places, for example, right here, you see a locus of energy right there.

Then the question is, is this energy that's occurring right there, is it somehow related to this energy up there, or are those two bursts of energy from two different talkers who are basically talking at the same time? That's always possible, you see the same complex interaction up here, you see a very heavily voiced section here 'cause you see a very strong harmonic section there. But is that somehow related to this formant change that you're seeing over here? And so the cognitive system's job is to take that very complex input and allow you to separate it into different streams. And then decide which stream you wanna follow and which streams you wanna suppress. To give you an example of how this works, I'm gonna give you a sound sample that you get to listen to. So this is a sound sample of three different talkers. There are two males and one female. And what I want you to do do to make your life a little bit easier is I want you to try to follow the female over time, okay. So like I said, there are gonna be three talkers, but there's only gonna be one female. And you don't know anything else about her, you just know that she's gonna be a female, and so that is something that you can use to help orient you. And I wanna see if you can track her voice over time. So let's see how well you do, give it a shot.

- [Female Speaker] My neighbor Tom is always on the go, he works long hours and gets home late. He's also taking a training course on computer networks. Once upon a time, there was a young woman looking for work. She lived in a tiny studio in a large city. One day on the train, she helped an old man.

- [Male Speaker] And it usually saves some time.

- [Donald] So how did you do with that? Okay. Wasn't easy, there's no, I won't deny that that was a very difficult task. But you knew that there is one particular talker that you were focusing it on. And you knew that she was a female, and so that helped give it to you and eventually, I'm sure you were able to find her voice in there. But can you tell what she was talking about? Sometimes, I've heard that sound, I've used that sound sample a lot, so I use it a lot. And so I hear her voice a lot, so I know what she was talking about, but she was talking a couple different people looking for jobs and how that was going. And you probably were able to track some of that if you're typical of what people can do. But not necessarily all the way through, because you don't always hear every single speech sound in a conversation, especially in a more complex environment. You hear parts, and hopefully you hear enough of the speech sounds that you're able to put the message together all the way across. For example, if you remember in probably what you learned somewhere along the line about the articulation index or the speech intelligibility index, that to follow sentences, you only need to hear somewhere between 60 to 70% of the speech information in order to put together the meaning of a sentence. And where does that extra 30 to 40% of understanding come from? Well, that comes from your cognitive system filling in the gaps. Because you're not gonna be able to hear everything. So when you were listening to this mix of the three different voices, you were using not just the acoustic signal and your knowledge of phonemes and speech sounds to be able to put her message together, but you were also using much deeper, more complex linguistic skillset to be able to put sentences together, 'cause she wasn't speaking nonsense words, she was speaking full sentences and telling you something that she wanted to get across. And so your linguistic system will fill in. In fact, if I asked you the question which cues did you use to follow her, it turns out that a lot of people say, well, I was using fundamental frequency, because she's a female and so her fundamental frequency was higher. Which may or may not be true, in that case, it probably was true. But the reality is is when you do that natural voice separation and tracking, your cognitive system will use whatever cue it can get its hands on. It doesn't necessarily have to be just fundamental frequency, like I said, it could be more acoustic related cues like supra-segmental cues or harmonic structure. But it could very much be linguistic use, for example, cues about sentence structure, about word frequency, about meaning, about, because the cognitive system is very predictive when we listen to speech. If I say a sentence to you like, for example, the watchdog gave a warning, and then I stopped at that point in time, you know, because of your knowledge of sentence structure in English, that there's gonna be a noun there, the watchdog gave a warning something.

And so you know it's a noun. And you know, based on the meaning of this being a watchdog and it's a warning, that I'm not gonna say a warning yip or I'm not gonna say a warning chair. It's gonna be something like growl or bark or snarl, something that fits into the meaning of the sentence. And your cognitive system is very predictive of the way you do that. So when you were listening to that woman speak, you were doing a lot of prediction, if you can get little snippets of what she was saying, about what she was probably gonna be next or how that works, and it's just the way the cognitive system works. Now, in that situation where you just listened to a mix of three different talkers at the same time, one of the things you did not have is you did not have visual cues, you did not have location cues, et cetera. So you didn't have the full set of cues that you may have had if you were actually in a situation where you had to do that listening. So even though you didn't have visual cues, you didn't have location cues, you still had enough information probably to do the task pretty well. It's important to remember, when we are dealing with patients with sensorineural hearing loss, the quality of the information that gets sent up from the periphery is quite reduced, right. So in the top panel here, what I'm trying to get at is that you have sound, you have it going through the cochlea. And that, in the normal situation, the neural code that's generated in the peripheral auditory system through the function of the inner ear and the eighth nerve and the ascending tracks in the peripheral auditory system creates a neural code that's very highly specific in terms of frequency, and so that's what I try to depict here with the different colors of the lines. And it is very highly coded in terms of temporal information, and that's what I try to do with the different dashes, meaning that you have low frequencies, mid frequencies, high frequencies, obviously much more resolve than that.

But you also have a neural code that codes the temporal characteristics of the information, how often the sound is occurring, things like that. And so you have both frequency resolution and temporal resolution that is information that the cortex will use to put together that signal, that auditory signal, and especially the speech signal. When you're talking about the case of sensorineural hearing loss, the code that's generated in the peripheral auditory system is far less specific. You don't have as much information frequency wise, either because of audibility issues or because of a lack or a loss of frequency resolution at the periphery. You have poor temporal resolution in many cases where you don't, you can't track the moment to moment changes in amplitude in the signal that carry a lot of information in the speech signal. And so the quality, the amount of information, the specificity of the information that gets up to the cortex is simply reduced in the presence of sensorineural hearing loss. So all those really helpful cognitive things that our system does for us oftentimes get lost because you're losing this level of specificity. The other thing that we have to recognize is a lot of patients who get fit with amplification are older, and that's just something that is part of what we deal with as professionals. But it's important to recognize that, when you take a look at normal aging, and I'm not talking about clinically significant things like Alzheimer's or dementia at this point in time, I'm just talking about normal age related changes in the body. There's a lot of skills that are affected as you get older. And this is a list, over on the left, is a list of some of the things that happen. Over on the right are the sort of things that typically don't get affected as you get older in terms of basic intelligence and basic linguistic skills and long-term memory. But it's more these sort of skills over here on the left that get affected as you get older. And one of the things to notice is a lot of these are based on the ability to very quickly and very accurately encode information. So for example, sensory-motor reaction time, processing and decision speed, selective attention. Those are all very cognitive based functions that can be somehow degraded as you get older because of just natural changes in the way the central nervous system can operate as you get older. And the problem is is that many of those are related to speech understanding.

And so, as you get older, there's just a generalized sense that you lose this efficient processing of information. And in general, the professionals in the area talk about it as neural slowing. Meaning that neurological events just take longer to happen, they don't happen as rapidly and in as much of a organized fashion as they do when a person is younger. And basically, you don't have that really strong, well timed robust response in the neurological system that's related to the reaction to stimulation. So those are all things that we have to keep in mind when we talk about what's going on in the cognitive system. At Oticon, we have grouped it down into the statement that one of the things that our patients deal with is the loss of the ability to organize sound. Meaning that one of the basic things that our cognitive system does for us is take this broad range of sensory input that we deal with and create a very structured interpretation of the environment, of the sounds that we are being bombarded with, it creates meaning that organizes that situation and allows us to pay attention to what we want to and suppress the rest. And that is all about organization of the sound environment. And one of the things that we deal with because of the peripheral hearing loss and perhaps because of maybe natural changes in the way the neurological system works as you get older, that patients oftentimes just can't organize sound in the way they used to. And so because of that, it's extremely important that, from a technology standpoint, we do the very best shot we can to at least help that organization, certainly not harm it. So when you talk about speech understanding, what it takes to allow a person to understand speech, especially in a complex environment. There's a lot of things that can go into it, and many of them are very cognitive based. When you take a look at that list, it's important to understand that there's a lot of different cues that the patient can use, and when we create solutions at Oticon, we're very conscious about trying to make sure that we don't disrupt any of the cues that the cognitive system can work with. Yes, it's very important that we improve audibility and it's very important that we enhance the signal to noise ratio as best as we possibly can. But it's equally as important that we don't disrupt naturally occurring cues that might be out there. So with that in mind, there are three basic principles of cue preservation that we pay attention to.

And again, in, I'm just talking about that part of our signal processing that's related to trying to protect information. We do a lot of other things to try to enhance information, and that's, you can get information on that from other courses we have on Audiology Online. But right now, I wanna talk specifically about the technologies that we have developed to try to make sure that we provide the best information we possibly can to the cognitive system. And the first principle of cue preservation that we follow is that where sound comes from matters. Like I said, when you walk into a complex environment, one of the things that your cognitive system is trying to do all the time is organize that sound space for you so that you know where the different sources of sound come from. So for example, if you imagine that the listener of interest in this situation is this woman right here, that woman in the beige blouse. That when she's there and she's trying to have a conversation with these three people here, she wants to be able to track these three people. But she also knows that there's a lot of different sound going on, you have a band, a little combo back here. You have traffic noise out here. You have a conversation going on right behind her, you have a conversation over here, you've got the cappuccino maker over here, et cetera, et cetera. That looks like there's a fan right there. All those are different sources of sound that could potentially disrupt her ability to focus on these three people that she's listening to. And so where all these sounds come from matters to her, because one of the things that the cognitive system will do is that it'll use localization and location information to allow this person to focus her attention spatially on these people and try her best to suppress all these other people. Now, there's other technologies that can be used, of course, in hearing aids, to help focus in certain areas of the room, directional technologies. But right now, I'm talking about just naturally supporting the cognitive system in doing this process. I wanna talk about the way hearing aids have worked for a number of years before we created something that we call spatial sound.

So spatial sound is a system that we use in our hearing aids to try to preserve spatial information for patients in complex listening environments. So I want you to imagine a situation where you have a person wearing hearing aids in both ears, okay. And you have sound occurring over here to the person's right, so you have a sound originating over in this direction from the person wearing the hearing aids. What we know that, in the mid to high frequencies, that sound is naturally gonna occur at a higher level on the ear closer to the sound source than it is on the ear further from the sound source, that's just a naturally occurring effect in the mid to high frequencies of the fact that we have a head. It's the head shadow effect, it's the fact that the head is effectively blocking sound in the mid to high frequencies that tell the brain information about where that sound comes from. The brain notices the differences in level from sound occurring on this side of the head versus this side of the head, so the brain knows that the sound source must be over in this direction. Now imagine that the person has the same amount of hearing loss on each ear. And so they're wearing a hearing aid that's set basically identically on both sides of the head. And you have sound coming in that's naturally louder on the patient's right side than the left side, but it's a nonlinear hearing aid. And as you know, the job of a nonlinear hearing aid is to adjust the amount of gain it provides to the signal based on input level. So the louder the input or the more intense the input, the less gain is provided, the less intense the input, the more gain is provided to the listener. So what ends up happening, if you have a compression system working independently on side of the head versus the other, then the hearing aid over here on the off side or the left side is just gonna provide significantly more gain to the signal than the hearing aid over here on the right side. And so what you have done by having those hearing aids work independently is that you're taking away from the brain an important cognitive cue that the brain would use to help organize that listening space, which is the fact that sounds should be louder when they're closer to one side versus the other.

When we created spatial sound, our goal was to override the naturally occurring compression response of the hearing aid to allow it to maintain interaural level differences on one side of the head versus the other. And we do that by having the hearing aids communicate with each other in a binaural set. So it's looking for signals that are basically identical on one side of the head versus the other, just coming in at different levels. And if it sees that combination of signals, it knows that that signal is likely coming more from one side of the head versus the other. And so what it will do is, instead of allowing the off side to turn that gain up more so that you get the same level on both sides of the head, it overrides that natural compression approach so that you get a preservation to a greatest degree possible of the interaural differences from one side of the head versus the other. So basically, we see this as a cue preservation technology because it's preserving localization information that the cognitive system will use to organize the sound environment. And as most normal listeners know but don't really know necessarily consciously, organization of the sound environment is so important to be able to follow one listener versus the other. In order to create that communication link between the different sides of the head, we decided early on in the wireless era of hearing aids to use near field magnetic induction in order to do that. The reason is is that, at the time, when you basically had to choose between one type of wireless system in a hearing aid versus the other, near field magnetic induction was being compared to 2.4 gigahertz Bluetooth, classic Bluetooth. And classic Bluetooth made a lot of sense if you wanted to be able to transmit over a large distance. But near field magnetic induction had advantages in terms of how much energy it used, how power hungry it was, the size of the componentry, and very importantly, how quickly you were able to exchange information from one side of the head to the other. Bluetooth is just a lousy technology to use to try to get a lot of information from one side of the head to the other. And that's why we started using and now several different manufacturers use near field magnetic induction as a way of transferring information from one side of the head versus the other, because it's a much, it's a system that allows you to send a lot more information from one side of the head versus the other. So that binaural link is based on this near field magnetic induction. And to show you how it works, here's a recoding that we did with KEMAR.

So this is the open ear on KEMAR, and I place the sound off to KEMAR's right, 90 degrees off to the right, and measure the level in both the near ear and the far ear. And I notice in the mid to high frequencies, if you average across the mid to high frequencies, I'm seeing about a 10 and a half dB difference on side of the head versus the other. I then fit, I fit KEMAR with a set of hearing aids with this spatial sound feature, this binaural link between the ears that controls the compression system, and I put it on the patient, but I turn the spatial sound feature off. And then I measured on the near ear and far ear. And what you see is that the naturally occurring level differences drop from 10.5 dB down to 2.5 dB. So this relatively robust location cue that naturally occurs now got minimized to a very great degree because the compression systems on both sides of the hearing aid were operating independently over each other. I then activated spatial sound to link the two compression systems and exchange information. And what you see is most of the interaural level difference from one side of the head versus the other was then restored for the patient. Well, for KEMAR in that case, but that's the way it would work for a patient wearing spatial sound. So what you see is technically the systems designed to be able to preserve that source of information to allow the person to do a better job of organizing the sound environment. This is data that was presented by a group of our researchers back about 10 years ago, where they tested this system in 30 patients with sensorineural hearing loss to test localization errors for patients wearing spatial sound turned on or turned off. And what you saw was a statistically significant drop in the number of localization errors that occur when the spatial sound system was turned on versus turned off. So we know that it has the effect that we are expecting, which is to try to preserve information about location of information in the environment for patients. Again, because that's a naturally occurring cue that the cognitive system is gonna use. The next technology that, oh, and the next principle of cue preservation that I wanna talk about is the fact that the bandwidth of speech matters.

That trying to get a good full bandwidth of the speech signal is very important when understanding the speech signal, especially in complex listening environments. And so one of the movements that we've seen in the hearing aid field is this idea that you do want a fuller bandwidth of the product. Oftentimes, people talk about having a fuller bandwidth in terms of sound quality issues and things like that. But at a cognitive level, in terms of speech understanding in complex environments, it's very important to know that the bandwidth of the speech signal is very important for the cognitive system to be able to separate one speech signal from another, and let me show you what I'm talking about. Okay. So if you are in a situation where you have this audiogram here, so you have this typical mild to moderate sloping loss. And the question is, how much bandwidth would you like to try to be able to create for the moderate speech signal? So sometimes, if you only go up to about four kilohertz, you can get a lot of speech information, so your SII score or your AI score would be pretty good. But there's still maybe some extra information up in the higher frequencies that may be useful to the patient. And how do we know that that would be the case? Well, let's go back to the situation with the single voice versus the three talkers. So when I talked about this single voice situation, I said that there's information in the speech signal that can occur in many different frequency regions at any different time, for example, you have a high frequency consonant sound up here that, like I said, that's either an S or an SH. You see a lot of information that occurs up here in the very high frequencies that the listener might be able to use to track the speech signal. When you put the single talker then in a situation where you have multiple talkers, so now you're talking about the three voices, the reason why the cognitive system likes better bandwidth is that you're basically increasing the number of opportunities for the cognitive system to get information that would allow you to link one talker over time from one speech sound they produce to the next one to the next one to the next one. Because you never know at any point in time whether the phoneme that the person is producing is going to be masked by energy from another talker.

So if you can increase the bandwidth, let's say from four kilohertz to eight kilohertz, for example, and so you have some usable speech information out there in the higher frequencies. It's not necessarily that you couldn't understand the speech signal if you only had it up to four kilohertz, 'cause you probably could. But because there's always the opportunity that the information in the lower frequencies could be masked at that moment in time by another talker, then by having the extra bandwidth out there, you increase the chances that you might catch a speech sound or some speech energy that will allow you to link that signal over time. Remember, when you're listening to a talker in the presence of other talkers, your job is to link that information over time. Speech doesn't take on meaning in isolation. Speech only takes on meaning when you can link it over time and follow the message. And so from a cognitive standpoint, the importance of having broader bandwidth is that you're increasing the number of opportunities to find information that'll allow that linking or that streaming to happen over time. And the data supports that, this is data from, out of Medical University of

South Carolina, Amy Horwitz and her colleagues that were taking a look at a group of patients again with this mild to moderate typical sloping loss. And looking at the ability to do constant recognition in the presence of a broadband noise. And they did low-pass cutoff for both the speech and the noise that was at two K and then three K and four K and five K and all the way up. And what you notice is that, as you increase the bandwidth that the patient's being provided with, that you are steadily increasing their ability to recognize those speech sounds. And remember, as they increase bandwidth, they're also increasing the bandwidth for the competition, the noise, so they're not just improving the bandwidth of the speech signal, they're improving the bandwidth for everything that the patient's listening to. But because you are increasing that bandwidth and you're increasing the number of opportunities the patient has to hear speech information, the cognitive system tolerates the extra noise. The cognitive system doesn't mind the extra noise coming in in the higher frequencies if it increases the opportunity of the cognitive system to get speech information up there. So again, it's a trade off that the cognitive system comes down very strongly on in favor of just give me the information, I'll figure it out. And so any hearing aids that tend to restrict bandwidth up in the higher frequencies, again, restrict potential information that the cognitive system could use.

And so trying to get good information up in the higher frequencies is important. One of the things that you may or may not have noticed about Oticon hearing aids though is that, when you take a look at the way we do bandwidth, we do one particular adjustment that I wanna point out to you. So here's the response that we would do with our Voice Aligned Compression plus approach, our VAC plus approach. And this is for someone with a flat, moderate hearing loss. And this is the insertion gain that we would prescribe for 45, 65, and 80 dB inputs. And one of the things that I want to point out to you is that you notice up here for the very soft inputs, we tend to roll off the high frequency response. The question would be, well, I thought you just told me how important it is to get as much information in the higher frequencies as possible. What I told you was that it's important to get as much information in the higher frequencies as possible in more difficult listening situations. And typically, in more difficult listening situations, you're talking at a minimum of moderate speech levels or maybe higher speech levels, noisier environments where you're trying to listen in the presence of other people talking. It's not terribly common to be listening to one talker in the presence of other talkers when everything is very very soft, I suppose it can happen, but it doesn't happen so often. So when we talk about the way we go about trying to create bandwidth, we tend to go more for the moderate to higher level inputs of preserving bandwidth. Then why, you might ask, do we roll off for the softer sounds? We roll off for the softer sounds for sound quality reasons. Patients, especially first time patients, but even experienced patients, tend to be a little bit, have a little bit of a negative reaction if the sound of their hearing aids is a little bit too hissy or too high frequency, especially for softer sounds. It creates a sound quality that they might not be necessarily all that interested in. So what we found over the years is that, if we roll off the higher frequencies just a little bit for very soft sounds, it becomes a much more acceptable response as long as, by the time we get the moderate and louder speech, that we preserve the high frequency bandwidth.

Because those are typically the speech levels where patients are gonna be working against other talkers in the environment. Okay, so those are the first two principles of cue preservation that we follow in our hearing aids, one is that where sound comes from matters, and the second one is that the bandwidth of speech matters. The third and very important principle that we've followed for a number of years is that the details of the speech sound matter. And this is one that we spent a lot of time talking about through the years, because this is one thing that hearing aids, especially compression, multichannel nonlinear compression hearing aids, have historically not treated the speech signal very well. And there's been more sensitivity to the way you do that in hearing aids over the last 10 years or so. But for a number of years, probably close to 20 years now, we've been very much talking about you have to be very careful about the way you do compression in hearing aids to make sure that you're not obscuring information in the speech signal. To give an example of what I'm talking about, let's take a look at a very simple situation, we're talking about the difference between SH and S, okay. And we know that there are two basic cues that separate SH from S. One is frequency cues. So the peak of the SH sound, the energy pattern in SH, tends to be a little bit lower frequency than the peak of the S sound. S is a slightly higher frequency sound just based on the way we produce it in our mouth. But the other thing that we know about the difference between SH and S is that naturally, SH tends to be more intense than S. It's based on where in the mouth we're producing it and how much energy you can produce by creating those two different sounds. So in the natural flow of speech, if you're producing SHes and Ses across time, you'll typically see that the overall level of the S is gonna be a little bit lower than the SH, that's just naturally occur. And it happens in the stop consonants, P, T, and K, there are natural differences in level in those consonants. And it happens a lot in different contrasts that can occur, that can happen in the speech signal.

But if you pass an SH and an S through a fast acting compression system, and remember, what does a fast acting compression system wanna do? It wants to provide more gain for softer inputs and less gain for louder inputs. And so if you present two sounds that are naturally different in terms of level, an SH and an S, through a fast acting compression system, you run the risk of the intensity differences between those two phonemes being reduced. It's very similar to the spatial sound issue that I talked about, that compression systems, when they're operating independently, are going to try to minimize differences between two different sounds. A fast acting compression system that is taking different speech sounds is going to try to minimize differences between, naturally occurring differences between different speech sounds. And that's what happens. So if you present speech through a fast acting compression system, basically, what you're doing is you're taking a situation where there are two naturally occurring cues for SH versus S, one is frequency, here versus here, but one is also level, up here versus down here. And you're taking away one of those two cues. So now you still have this frequency difference, but remember, frequency resolution tends to be one of the first things that go in the presence of sensorineural hearing loss. But you're losing the naturally occurring level differences between those two different sounds. So it's one of those side effects of amplification, especially fast acting amplification, that is important to pay attention to. One of the other effects of fast acting amplification, this is probably more important than anything, is the way it's gonna handle noise that happens in the low points of an input signal. So in this case, what you see is the input signal is speech presented in the background of a lower level noise. And I recorded that at plus 15 dB signal to noise ratio. That's a very good signal to noise ratio, that's a signal to noise ratio that allows most people with sensorineural hearing loss, certainly in the mild to moderate range, to understand speech at a relatively good level. And what you see in this case is that you see the peaks of the speech signal over time and you see some low level noise down here, but not so much that it's going to be a big issue for most patients with sensorineural hearing loss.

The problem is is that, when you pass that through a fast acting compression, what does fast acting compression wanna do? It wants to take these low points in the signal and boost them up. So what you notice is here, where you normally have a pretty good drop between this speech peak and this noise section and this peak, all of a sudden, you lose a lot of that difference over here in the fast acting compression system. Because basically, the noise in between those peaks tends to be ballooned up because it is a fast acting compression system. In other words, it's timed so that every time the signal drops even for a few milliseconds, that you boost it back up. And so what happens with fast acting compression systems is that they tend to sound noisy when you're in noise. More noisy than naturally occurs in the room. Because even if you're at a good signal to noise ratio as an input signal, what's gonna end up happening is you're gonna get noise that balloons up during the low points of that signal. Now, you're not changing the signal to noise ratio at any given moment in time, because the peak is going to drive how much gain is applied to the signal. But as that peak goes away and you drop to a lower intensity section for the input signal, then that lower intensity section is going to be ballooned up a little bit. So fast acting compression systems tend to sound noisier because of this sort of activity that it does when it balloons up softer sections. What you see over here in the right panel is our system called Speech Guard, and I'll describe how Speech Guard works in just a moment. But our goal with Speech Guard is to have a nonlinear compression system, but without using fast acting compression as the primary way of trying to preserve audibility. And what you notice when you compare the speech guard waveform with the fast acting waveform with the input is that Speech Guard does a better job of preserving the naturally occurring peak to valley differences in the output signal that the patient gets to listen to. Basically, speech arches sounds quieter to a patient. Not meaning that the speech sounds too quiet or too soft. You can provide the full gain you want to the speech signal. But that noisiness that tends to occur in fast acting compression systems doesn't occur anywhere near as much with Speech Guard, because it is set up so that it doesn't react in that very fast manner in between sounds. And so it ends up giving the patient a much cleaner sounding signal to listen to.

The way Speech Guard works is that, compared to traditional compression systems where the input comes in and is measured by one system, in Speech Guard, as the input comes in, it's measured in terms of level in terms of two systems. One is a very long-term look at the signal. So what is the overall long-term level of the speech signal? So it's monitoring that and it's keeping track of that long-term level of the speech signal. Another system that's operating in parallel is watching the very fast changes in the signal. So it's looking, for example, if all of a sudden, I drop my voice to a softer level. All of a sudden, I start to get very loud with the way I talk. It's looking for those faster acting changes in the signal. If somebody slams a door, if a dog barks, those sort of things that would cause a very fast change in the signal that the patient would have to listen to. And these two different systems can drive how much gain is applied to the signal. But in general, we try to make sure that it's only the long-term system that's driving the gain applied to the signal. And by having only a long-term look at the signal, driving the gain on the signal, what ends up happening is you don't respond to those gaps between the speech signal that fast acting compression would do that creates that noisiness out of the fast acting compression. It starts, very importantly, to disrupt some of the normal cue sets in the speech signal. The only time we allow the fast acting system to control the gain of the hearing aid is when that fast occurring sound happens outside the normal range of the speech signal. As you know, the speech signal operates within a certain range naturally, as I pointed out earlier in this talk, different phonemes occur at different natural levels. And so you expect some variation in speech, that's the way speech is structured, and I talked about there's a lot of cues in that varying levels over time. But you have a window where you expect it to occur in. And as long as the signal is occurring within that window, it's not going to be adjusting gain. It only starts to adjust gain very rapidly when, for some reason, the signal jumps out of that window.

And because of that, normally, what ends up happening is you have a very stable, almost linear amplification applied to the speech signal and it only goes into a non-linear mode when it has to make a very rapid reaction to something that has occurred in the speech signal or in the environment. And that allows the system to operate in much more of a linear fashion so you can preserve the natural shape of the input signal at the output of the patient's ear and you don't get all this noisy fill in there. Very importantly, what also you preserve is some of those naturally occurring level differences from phoneme to phoneme, like I said, in the example with the S versus the SH, you just get a better representation of the naturally occurring level differences that would distinguish one phoneme to another. And so because of the electronic structure that we used in setting up Speech Guard, we have created a situation that allows a much more, a higher level of fidelity if you compare the outputs that the patient gets to listen to with the input signal. To give you an idea what this looks like, we ran an analysis of the correlation between the details of the waveform, comparing the input signal to the output signal using Speech Guard versus using a fast acting compression approach. So basically, if you go back to those blue on black figures that I showed you a few moments ago, we're basically doing a point by point comparison of the input signal versus the output signal. And we're seeing how close the correlations go. So for example, if you take a look at two voices and you compare the input to fast acting compression, you see a correlation around .7. If you take speech with some impulse noise in it, you see a correlation around .64. In our original version of Speech Guard that we first produced about, oh, about 10 years ago or so, you saw that the correlation was somewhere, significantly higher, in the range of .87 to .86 if you compare the input to the output using Speech Guard. After Speech Guard was available for a couple years, we were able to make some adjustments to it to try to make it even a better representation of the input signal. And when we were able to produce what we called Speech Guard E, we were able to improve the correlation between input versus output by another nice little step.

And so the point being is that there is details in that speech waveform that the listener will use, and those naturally occurring level differences from phoneme to phoneme or from speech versus non-speech moments are important for the listener to be able to lock into, to be able to hear and understand the speech signal effectively. And with Speech Guard, we've been able to do a better job of that. To give you an example of what this looks like, here is the phrase Doctor Simpson. And processed through a fast acting system versus Speech Guard. And again, you're looking at a spectrogram, which is frequency going up the Y axis, time across the X axis, and the brightness of the color means the intensity, so the closer to yellow or almost white that the color gets, the more intense it is. And what I want to do is point out a few examples of where the Speech Guard was able to better preserve some of those clearly defined phonemes so that they would stand out to be more distinct for the listener. So if we focus in on here, for example, here's the burst of the D in doctor. And what you notice here is that you just see it just light up more brightly in the Speech Guard situation than in the fast acting compression system. And again, this is because it's trying to preserve that, those naturally occurring peaks in the intensity of the signal, those are just better preserved. This is, I think this is the first S in Simpson. And what you notice is that just the brightness of that S, the location of the information, just brights up, is just more, it just shines better when you process it through Speech Guard than you do with a fast acting system. And here's the second S in Simpson. Again, you see more of a lighting up of the signal out there in the higher frequencies compared to, in Speech Guard than when passing through fast acting compression. So anyway, the idea being you see the same sort of situation there. Right, I'm sorry, I wasn't moving the arrow properly, so you see another example there where you just see a brighter representation of the phoneme in Speech Guard than you do with fast acting compression. You see it, right here is another good example where you just see a difference in how well preserved that peak energy of the consonant occurs.

So again, what we're trying to do is preserve those details of the waveform so that the listener has the best opportunity to hear the speech sound. To give you an idea what the data on this looks like, to show you how effective this is, about five years ago, Dr. Andrea Pittman known at Arizona State University created a nice complex testing paradigm where she was able to show that in a complex, more realistic listening situation, having better preservation of the details of the speech waveform matters. So she did a situation where she was having patients identify the category that test words fall into. For example, owl would be an animal. That would be the patient's task, is to classify the keyword into different categories, and you'll see what those categories are in a second. At the same time, she was providing some extraneous sounds that would occur in the environment, either overlapping or even before the test word that she was looking at for the listeners. To make the test a little bit more difficult, she was presenting it in some background noise. So that your, basically, the patient's task was to hear this word and to categorize it in the presence of some background playground noise at the same time where there might be extraneous noises that pop up from time to time in the listening task. At the same time, the patients were given a visual task where they were looking at some object and trying to pick what the next object in the sequence was. And the reason she did this, it was trying to load the cognitive system up a little bit more. In other words, she was trying to replicate a situation where you're doing multiple tasks at the same time, for example, you're listening but trying to remember, you're listening but trying to formulate a response. So she wanted just to load the cognitive system a little bit more. And like I said, the job wasn't just to identify the word and repeat back the test word, but it was to identify the word and then categorize it into a person, food, animal, or something else. And so the idea here was you wanted to do a little bit deeper level processing, it's not just repeating the world owl, but recognizing that the owl is an animal. And again, the whole job was to give the person a little bit more complex cognitive tasks to do so that the details of the waveform would be more salient to the person and really jump out to allow them to be more automatic in their recognition of the sounds.

And what the data shows is that, when she compares slow acting compression to fast acting compression to Speech Guard in what she called adaptive compression, especially when there is an overlap between the test sound and the extraneous sound that can come in, that you saw a strong statistical difference between Speech Guard compared to two of the more traditional compression approaches, fast acting compression and slow acting compression. And so what she was able to demonstrate is that preserving that information in the signal becomes important, especially in a complex listening task like listening to speech in a background of noise. So to sum up the talk, I'll get back to the why. And the question is, why do we talk about cue preservation technologies? And it comes back to the fact that it's very important to recognize that, over the years, we tried to do some very good things in hearing aids, especially multichannel nonlinear compression. But when we do something like multichannel nonlinear compression, there can potentially be some downsides, it could disrupt spatial information, it could disrupt some of the naturally occurring cues in the speech signal. And so to be consistent with our BrainHearing approach to creating sounds, we wanna make sure that not only do we do the basic jobs that a hearing aid has to do, which is improve audibility and improve signal to noise ratio, but we also do it in a way that we preserve so much of the naturally occurring information in the speech signal to really allow the patient to have the very best opportunity to hear and understand speech, especially in these complex situations. If you have any questions or comments about any of this information, you're always welcome to contact me, my email is on the final screen. And I wanna thank you for your time today.

.