A Deeper Look at Sound Environments
Recorded August 27, 2019

Presenter: Don Schum, PhD
AudiologyOnline.com Course #33536
Partner: Oticon

- - [Don] Hello everybody. This is Don Schum from Oticon and I wanna welcome you to this seminar called, A Deeper Look at Sound Environments. And in this seminar what I wanna do is a part about the typical signals that fill up the day of a person, especially a person with hearing loss, and talk about some of the ramifications of some of the maybe perhaps subtle aspects of both competition sounds that people face and speech sounds that people are looking to hear well with their amplification. So let's go ahead and get started. The learning objectives for this course is to first of all take care of some of the basic nuts and bolts of talking about noise and speech. So we'll talk about the range of signal-to-noise ratios in typical daily environments and talk about that in perspective of what a person with hearing loss probably needs in order to hear and understand speech well. Then we'll talk about a variety of other environmental factors that could compromise speech. Understanding things like distance and reverberation and those sort of physical elements that could add on top of the basic measurement of signal-to-noise ratio that we all know well as audiologists. And finally I'm going to spend a little bit of time talking about the role of sound in a person's life. In audiology we tend to minimize the discussion to pretty simple concepts and the reality is that there's reasons why sounds exist in the world and they bring value to people in different ways. And sometimes we lose sight of what the sounds around us actually mean as a human, and mean as a listener, and even though a person with a hearing impairment is looking for communication improvement we also can't lose track of what does sound really mean to the human and how does a person use sound to relate to the world that they live in?

So, let's get started. As I said a few minutes ago, a few moments ago, the basic reason why I wanted to do this course is that we attempt to have dumbed down the idea of speech and noise to kind of the Tarzan level of speech which is, noise bad, speech good. Meaning that when you talk about a person with hearing impairment you talk about technologies, hearing-aid technologies, or other technologies, that are designed to help the patient out. There tends to be somewhat of an oversimplification of what we're trying to do is always present good speech to the person and make all the noise go away. The reality is first of all we can't do that, you know, the technologies are not that good that they can totally get rid of all competition at every moment in time and only present target speech to the person. There's a lot of reasons why that doesn't actually exist. We've made improvements significantly in technology over the decades and we're gonna continue to see good improvements in this area, but it's not, there is no perfect solution out there for this problem. The second aspect about it, why this is an over simplification, is that not all noise is necessarily bad. Noise is in our environment. Sounds in our environment that aren't necessarily speech coming from one person that we wanna listen to, carries a lot of meaning, they carry a lot of value. And so when we talk about trying to put speech and noise in perspective of the person with hearing impairment and what they're looking for it's very important to understand the role of sounds in a person's life and make sure that we are not over-simplifying the issue too quickly.

So this is the agenda over the next hour. I'll talk a little bit about the nature of noise, what does it mean when we talk about, and use the word noise? I'll talk about what we know about sound environments, you know, what are some of the levels that we know, what are some of the characteristics of competing sounds that we should be paying attention to. I'll also spend a little bit of time digging into the speech signal itself because there's some aspects about the speech signal that will affect the way amplification works and I think that it's important to understand some of those realities about what the speech signal is all about. Spend some time talking about Soundscapes and the idea of creating a sense of place. Like I said, the sounds around us tie us to the environment and I wanna dig into that a little bit and make sure that we don't lose track of that, one of those fundamental realities of sounds in our environment and how they affect the person, even the person with hearing impairment. And then I'll make a few comments about the way we try to solve the speech and noise problem in technology. The idea of comparing to what the human brain can do versus what we can build into a machine, a hearing aid, to try to solve the speech-noise problems. I won't spend a lot of time specifically on that topic because we have a whole nother AO seminar specifically called, Brain Versus Machine, that really focuses on the contrast between the way the human cognitive systems works and the way traditionaLly hearing-aid technologies have worked to deal with speech and noise issues. And if you're looking to dig deeper into that I would certainly recommend that you go find the AO show called, Brain Versus Machine, on the Oticon portal on the AO site. Let's start and start talking about the nature of noise.

So the first question we need to ask ourselves is, what is noise? Because we, as audiologists, throw that term around all the time, but it's important that we are much more clear about what we are talking about. Because if we talk about noise in one way, and patient's talk about noise in a different way, then there's a mismatch between the care that the patient's looking for and the care that we are providing. When we test auditory skills using speech-shaped noise, or babble noise, clinical-babble noise, or other noises like that, we aren't necessarily capturing all the aspects of what noise is like. And so if we use an auditory test that is highly calibrated, and highly controlled as tests need to be, and believe that that totally captures the perceptual experience of the patient, we're fooling ourselves because noise is more complex then that. In fact what I would recommend typically is that we don't talk about noise, we talk about complex listening environments because that's what patient's actually are dealing with. They are talking about being in situations where there are multiple sources of sound. Some of those sound sources are valuable to the person at that moment in time and some of those sound sources are not, they're competition, but that can change all the time. The nature of the competition can change, the nature of the signal can change and they can be changing in opposite directions, there's a lot of things that can be going on. So really to understand the challenges faced by patients, and also to understand the role that hearing-aid technologies could potentially play in solving those problems, we have to be honest about what the patient's are talking about. They're not talking about one particular noise usually. They're talking about all the competing sources of sound that they have to deal with and they talk about their inability to really separate out those sources of sound. So if we ask our, the question of, what are complex environments really like, there's a lot of things that can be going on in complex environments. You could be talking about multiple talkers. Usually when patients are writing in the problems, the most challenging part of the problems they run into is multiple-talker environments. Being at a restaurant, being at a party, being around the dinner table with a large family, being at other events like that where there's just a lot of different people talking and the person wants to listen to one particular person at this moment in time and not hear everything else going on in the room.

But the point is sometimes it's not just one particular person that is the target talker, it could be a group of people where a conversation is going on. And when you talk about being in a group with other people talking, and a group conversation, those are some of the most interesting, and sometimes some of the most fulfilling social situations, but they also can be a little bit challenging if there's other people in the room talking, or if this group of people are talking over each other, or if one person's a little louder, and if somebody else has a weaker voice and there can be a lot of complications just in that little scenario like that. You could have movement of the target signal, you could have movement of the noise sources, so the world doesn't stand still as you know so you can have a lot of things going on. You can have stable non-speech sources of sound, or noise, those are kind of the classic sort of targeted noises that we talk about. The air conditioner running, traffic, well, not necessarily the traffic noise, but other machine noises where there is a noise that has a certain spectral and temporal characteristic, it's in one location all the time and it's always there. Those tend to be the type of noises that are a little bit easier to get rid of via signal processing. But when you start talking about unstable non-speech sources, like traffic noise is a very good example where if you're outside, let's say you're outside at a cafe having lunch with someone nearby a street and you have traffic going by, well, depending on how fast the traffic is, whether the stoplight stop the traffic for a moment or not, whether it's a truck rolling by versus a little Prius rolling by, and I don't mean to say anything negative about Prius owners, but you know the idea of what I'm getting at. There's just a lot of different things going on in those environments and it's very unpredictable about what sort of competition the person's gonna be dealing with at that point in time. You can have distractions. You can be trying to listen to one person, but your attention gets pulled away because it's kind of like a self-inflicted distraction, meaning you keep looking down at your phone, you know, or something like that, or distractions that you're not necessarily in control of like your kids running in and asking you a question, or something else like that. But that can change the ability of the person to really focus and concentrate on the speech they wanna listen to.

You can shift your focus and a good scenario that I talk about all the time is imagine that you're at a group, a lunch, so you're sitting around a lunch table with, you know, three or four, five, six other people, and you're having a conversation with the person on your right side. Well, at that moment in time the person on your right side is a signal and all the other conversation going on at the table is noise. Until you decide that you wanna talk to the person on your left. Now that person, who just a moment ago was part of the noise, now is the signal you're interested in and the person on the right who now turns in the other direction and starts talking to somebody else, they just a moment ago were your signal of interest and now they're part of the noise source. And then the waiter comes down and he's talking to somebody down at the other end of the table. Well, he just added to the noise because you're really not paying attention to what he's saying until he starts reading the specials and then you're very interested in what he's saying and so you start paying attention to him and now he was a moment ago part of the noise and now he's part of the signal. So it's that idea that in real-life situations there can constantly be shifting of what you wanna be listening to and what you wanna be ignoring. The basic reality is in most complex situations, especially the stype of situations where people with hearing impairment have a lot of trouble, it's a little bit of everything. You know some of the most complex situations are also some of the most fulfilling situations where in their social situations, their family situations, there's a lot of liveliness going on, but liveliness tends to bring noise along with it, or at least competition along with it and that's the sort of thing that patient's with hearing impairments struggle with a little bit. One of the things to think about, and again, like I said I'm not gonna talk a whole lot about technology in this course, but one of the things to think about is the idea that in, in these situations where you're having all these sources of sounds and they're changing all the time, how can you tell a signal process or what is the thing that I'm interested, at this moment in time, and where are the parts that I want you to ignore?

Because since that can be changing on a moment-to-moment basis, depending on the way you're having a conversation, then the idea to think that a hearing aid somehow can track what you're interested in listening to, becomes much more tricky. And so one of the kind of future horizons in the whole area of kind of futuristic hearing-aid technologies to think about is can we build into a hearing aid some ability to understand what you wanna listen to? We do it to a little degree in those hearing aids with a very narrow-beam focus where, you know, the assumption being you're looking directly at what you wanna listen to and everything else is noise and then you shift who you look at if you want to point the beam in the right direction. But the reality is in a lot of naturalistic interactions patient's, or individuals, don't necessarily always totally orient at just who they wanna listen to, and if that person's at a distance, or if there's multiple people talking and the conversations bouncing back and forth, that's not necessarily a complete solution to the problem. So you see the complication of what's going on. I'm gonna play you a sound sample now and in this sound sample you're going to hear three different talkers at the same time. And what I want you to do is to notice whether or not you can focus on one talker and suppress the other two because if you have normal hearing it tends to be something that you are able to do even if you've never heard these three voices before you still will be very quickly able to lock onto an individual voice and follow that through. Now all three of these voices were mixed in at approximately the same overall level although in your ear one might sound more dominant than the other two. But what I want you to do is throughout the course of this sample just kind of listen to it, it's 21 seconds long, so it's not terribly long, but just try to follow the different streams at different points in time then shift between the streams and things like that. So this is just kind of a free, kind of a free-flowing opportunity for you to exercise your cognitive system. Let's give it a shot.

- [Woman] Always on the go, he works long hours and gets-- But at the last minute the firm ordered-- Once upon a time-- He always was at the very tip-- There was a young woman looking for work who lived in a tiny studio in a large city. I got tired of sitting -- One day on his train he helped an old man. 12 or 15,000.

- [Man] That usually saves some time.

- [Don] So, how did you do? Well, you had three voices. You had one woman who was talking about her cousin, Tom, I think. You had one guy kind of complaining about the traffic, or the weather, and then you had another guy reading a passage out of, "The Great Gatsby". But the point being that as a signal at any moment in time who you were listening to was the signal and the other two voices were the noise. But a moment later if you decided to follow a different stream of speech, from a different talker, then at that moment in time that new talker was the signal and the other two were the competition. And the point being that that task is something that the cognitive system could do. It takes some effort, but it's something that the cognitive system can do. And the problem with a lot of people with hearing impairment is because of the breakdown in the peripheral-auditory system, and the poor coding of sound that's passing through the peripheral-auditory system, the brain doesn't have the same good signal to work with and so the ability of a person with hearing impairment to be able to follow those three talkers, and focus on one and suppress the other two and switch between, is very dependent on the brain getting really good information in order to do that job. And if the hearing impairment is causing some distortion in the coding that gets sent up to the cognitive system, then the ability to do that task becomes much more complicated. So I think I've kind of made my point hopefully by this point in time that, that complex situations are truly, truly complex. So let's talk more about environments and what we know about them is just some basic realities. Life can be a little noisy throughout the course of the day. This is data from one particular listener, in a project that the late Stuart Gatehouse did, where he was tracking the behavior of a lot of different hearing-impaired individuals throughout the course of the day. And this is data from one particular person and on the bottom of the screen, if you blow your screen up to a very high degree, you can read what was going on throughout the course of that day for that person in the little handwritten areas that the person wrote down.

But the point being is that throughout the course of a day of a typical adult person, or a typical child for that matter, the overall levels in environments can change dramatically throughout the course of the day. And so when we talk about noise noise doesn't come in at one particular level, speech doesn't come in at one particular level, it's all over the map throughout the course of the day if you're leading a real life. And because of that any technology, any solutions that we create, have to be able to handle real life in this way. It has to handle this range of sound inputs in such a way that you kind of get a, you kind of have the ability to count on your technology no matter what the sort of sound environments are like. Different people do lead different lives. This is a summary of the overall levels of eight different individuals from that study. So those are histograms of the overall levels and the amount of time that those levels exist. So the red numbers on there give you the range that were covered in that particular graph, and what you see, for example, some people just live more noisy lives than others. For example, if you look at the upper tier, all the way over to the right, you know that that person is leading a pretty quiet life and a pretty consistently quiet life that throughout the course of that person's day there's not a lot of variation in sound levels. But then you look at somebody like, let's say the person on the top tier, in the second spot, the second spot from the left, and you see that person's in a broad range of situations throughout the course of the day, or the person in the third spot on the upper tier you see that. That person in the first spot all the way over to the left on the upper tier, they have a broad range of situations, but you notice that the skew is placed, is kind of pushed towards the, to the louder end. That person seems to have more challenging environments that they're in throughout the course of the day. So the point is simply that some people just lead more active and more noisy lives than others. And that's something that you, that you just, you know, need to take into account when you think about the way patients are going to be able to need and use amplification. There's a classic study done back in the '70s by Pearson, Bennett, and Fidell. They were, they were contracted by the U.S. government to do a good-comprehensive study of speech and noise levels across the day of typical people. So here are some of the levels that they were able to see in some kind of classic situations, inside homes, outside homes, and a nursing station at a hospital, a department store, a commuter train. And the pink bars tell you what the average noise levels were and the blue bars tell you what the average speech levels were.

So what you notice right away is that signal-to-noise ratio is changing all the time. That, you know at that, for example, inside homes you see the best signal-to-noise ratio. It's kind of on average about six or seven dB positive, but you get into a commuter train and it's negative, you've got a negative six or seven dB, signal-to-noise ratio in that situation on average and you see them all over the place. One of the things that they point out in that study, and I don't have it necessarily on a graph, but in that study is that as the overall level gets higher on average the signal-to-noise ratio gets worse. Meaning that, that as the background noise level gets higher, talkers will adjust their level to try to match the background noise level increase, but they never typically increase their level to the same degree that the background noise level has gone up. So if you're in, let's say you're sitting in a restaurant, and you're sitting in the restaurant for over two hours and when you get there it's relatively quiet. So if you're a talker you're sort of monitoring the background noise level and you're putting in the amount of effort necessary to get a good signal-to-noise ratio. As the dinner hour starts to progress and more people come in the restaurant and it's gets busier and busier and busier, and the noise level starts to go up, a typical talker will raise their speech level to respond to that increase in background noise level, but they won't increase it at a one-to-one ratio, meaning they won't, for every dB that the background noise level, they won't increase their speech effort up correspondingly. So by the end of the evening, although at the beginning of the evening you may have been begin able to create a plus 10 dB signal-to-noise ratio, by the end of the evening if the background noise level has gone up 20 dB, as it got noisier, you might be only operating at a plus five, or a plus two dB, by the time the evening is kind of drawing to a close because you don't totally respond to the background noise level in a typical way and that's a pretty consistent observation about that. So the higher the background noise level typically, the worse the signal-to-noise ratio.

This is the sort of situations that hearing-impaired individuals will find themselves in. This is data that I collected with a colleague of mine, Randi Pogash, a few years ago, where we took a group of 103 hearing-aid users, high-end hearing-aid users, meaning well, I can't tell if the users were high-end, but the devices were high-end, but I'm sure that the people were quality people too. Where it, in our hearing aids at the time that we were using, we categorized the type of environment that the person was in in order to drive the signal processing in the hearing aid. So we categorized it as, quiet only, speech only, speech plus noise, or noise only, and so you can see the graphic on the screen. And so for a typical hearing-impaired individuals you see the range of levels that they would have, but these were pretty consistent users of hearing aids 'cause they're using the hearing aids on average 11.2 hours per day, so they're pretty good full-time users of hearing aids. And what you notice is probably the single biggest block, of the type of signals that they're listening to, are speech plus noise signals in the range from about 50 to 80 dB, that kind of captures a lot of what they're listening to. So there, you know, one of the things to remember, is it's not like, well, you know, lower levels, or moderate-sound levels, are typically quiet and then noise occurs when you get above 70 dB, or something like that, and of course that's not true. There's noise throughout the course of the day. You can be in a relatively quieter, relatively moderate level situation, like a Starbucks at 2:30 in the afternoon where there's not a lot of activity going on, but there's still noise in the background, but it's not so high that a person can't speak over it at a relatively high level which is very different than that Starbucks if you had walked in that Starbucks at eight o'clock in the morning at rush hour, it's gonna be a very different sound environment.

So noise can be present in an environment, even in moderate environments, and most listening time, for hearing-impaired individuals, seems to be in speech plus noise situations. These are the longterm averages of a variety of different noises that you can be in. They're just a selection of noises that I picked out from a corpus of sounds that I have access to, the longterm averages of it. And one of the things to remember is first of all, speech is there in the green triangles, so you see the longterm average of speech, and you're kind of use to seeing that. Then what you notice when you look across there that many kind of typical daily noises have a very similar longterm average as the speech signal. There's a couple situations where that's not the case and so that is the subway which is, on the graph it is the black line with the open boxes. I know in the, in the legend it says it's a closed box, but that's the, no, no, I'm sorry, that's not the subway. I'm trying to remember what that is, that could have been a store. I'm not exactly sure what the black line with the open boxes are, no that is the subway, I'm sorry, that is the subway. But you see that in the subway you see a low-frequency dominance and those levels, the overall levels were corrected to equalize. So, you know, a subway you would expect that that'd be much higher. But most of those in other situations you see a very similar longterm average of the speech signal. The one other situation where it's a little bit different is the playground where you see that high frequency emphasis and that's probably because in that particular recording there's a bunch of little kids screaming at each other a lot, you know, that giggling, that happy scream that kids do all the time, and I think that that's why you're really seeing that mid to high frequency emphasis of that spectrum. It's not that I don't like kids, I like kids a lot, but sometimes they scream a lot when they're having a lot of fun at a playground. Anyway, the point of this graphic is that you can't necessarily count on static filtering to solve background noise. I think that's something that most audiologist understand at this point in time and the reason why you can't count on static filtering is the fact that so many longterm competition environments include speech, and so the competition is also a speech signal and it's gonna have a very similar longterm average of it.

So the cafeteria noise for example, or the store, or cafe, those the competition is also other people talking and so that's where you get some of that going on. One of the things that is important to recognize, because speech is a cognitive process, and what we mean by that is that if you're going to understand speech in a noisy environment, that's something, or speech at any time, that's something that your brain does for you, that doesn't happen in the periphery and the brain processes information. So the ability of a person to handle background noise is driven sometimes by the overall energy pattern, or the overall level, of the competition. So if you do, let's say speech and noise testing and using a speech-shaped noise, or clinical babble, and you set the level of the speech and noise, or set the signal to noise ratio, there's a real direct correspondence to the ability of the person to hear the speech based on the signal-to-noise ratio and it's a very kind of pretty clearly defined characteristic. But when you're in realistic situations information can also be a masker. In other words your ability to listen to one talker can be confused because you're getting information from another talker. So in that situation, several slides back, where I had you listen to one talker, versus another talker, versus another talker, sometimes what was making it difficult to follow one talker throughout the course, was the fact that individual phoneme or syllable, would have been masked specifically by a higher-level sound being produced by one of the other talkers, that would be energetic masking. But sometimes you got off track because you linguistically one, one track that you're trying to pay attention to gets confused with the message you're getting from another track. Now it wasn't so bad in that situation because the three talkers had pretty distinctly different voices. But if you're not in that situation you can run into the situation where you just can't separate out the voices because the message kind of overlaps each other. And to show you an idea of what we're talking about, where basic linguistic competition is the hardest to suppress, this is data from a little while back where they tested the ability of hearing-impaired individuals to understand speech in three different background noises. So they tested them with speech-shaped noise, they tested them with competing talkers and if I remember right from that study there are two competing talkers that would happen at the same time as the person you're listening to.

And those competing talkers could either be played forward in its natural condition, or backwards in an unnatural condition. And the thing about playing it backwards is that some of the linguistic information is taken out of there 'cause you can't follow, you can't build up sentences or phrases, in the competition and so it becomes a little less confusing. But you still hear words, you still hear speech sounds, you still hear syllables. Even though it's recorded backwards you still have a sense that there's some linguistic information, So those are the three different conditions. And so what you see here is the signal-to-noise ratio and the percent correct at the different signal-to-noise ratios. And what you notice is that the best performance, so the best performance at each of the different signal-to-noise ratios, is for the speech-shaped noise, so for noise without any linguistic information. The second best situation is for the backwards-competing message. So where you get some linguistic information, but not complete linguistic information out of the competition. And then the most challenging situation is where the competition was forward-competing masking. So it's another voice that you have to listen to, two other voices that you have to listen to, in it's natural forward competing situation and that's where at each of the signal-to-noise ratios, the performance is the poorest. And so the point being that when you have to listen to one person in the presence of other people talking, just what they're saying, just what the other people are saying, can sometimes interrupt your ability to follow the message of the person you're listening to. So that would be a case of informational masking which is different than physical or energetic masking, but they can occur at the same time. When you do a task in noise, for example, when you do a task like I had you do a few moments ago where you're listening to one voice in the presence of other voices, your brain will lock onto whatever information it can possibly find to differentiate the different voices. So it's not as if well one person had a higher pitched than another person, or one person spoke faster than the other person.

Your brain really doesn't know that ahead of time, but the cognitive system is really, really good at immediately identifying what, what characterizes a certain voice and using that cue to track that voice over time and it could be any one of these things, that are the screen, that it latched on to, or a combination of those, or it could use these different cues at different point while it's trying to track one person over time. But the point I wanna make with this, is first of all, this is a very good reflection of how effective the cognitive system is in dealing with background noise. But it's also a reflection of what's the biggest, one of the biggest challenges in creating signal processing out of a hearing aid to do that. Because typically in signal processing you have to tell the signal processor what to look for, there's some changes in the future that, you know, where the system can find the differences itself, but that's a futuristic sort of technology, but typically you would have to tell a signal processor what sort of cue to be looking for and the point is the brain doesn't have to know that ahead of time it just figures it out on the fly. And so that's one of the reasons why creative-signal processing that mimics how good the human brain is, is so difficult, because a human brain is so good at that task. Okay, so let's talk about speech a little bit and just kind of clear up a few details about the speech signal. And one of the biggest details to clear up is that speech doesn't happen at one level, you know, it's, often as we talk about, well, the typical speech level comes in at 58 dB SPL. The point is that 58 dB SPL is a good average overall level and that's the number that came out of this study by Pearson, Bennett, and Fidell. Other studies have come up with like 63 dBA, or some other things, but it's about in that range. But that's just a longterm average of moderate-level speech. It's not capturing the full range of speech in everyday life. So be very careful of thinking that speech always occurs at one particular level. Even, excuse me, conversational-level speech, or moderate-level speech, can be at a lot of different levels depending on the background level going on. It also depends on the talker who is someone who's a louder talker, or a softer talker, or whatever. What you also see on the screen are the range of the speech signals.

So you see typically 12 dB over and 18 dB under the longterm average and that accounts for the peaks and valleys of the speech signal. We often talk about the speech banana being 30 dB high, meaning that the range within an overall level of speech, you're going to see energy fluctuations on the order of about 30 dB. So that's what those bars. above and below the overall average, are intended to indicate. If we move into different situations, for example, now we go back to the different environments that came out of that Pearson overall study, you see the overall level, you see the overall average over here that came out of the study, but you also see this range of levels in the different environments that they tested. And the point being that when you take a look at these different levels with different environments and you add the 12 dB on top, and the 18 dB on the bottom, when you go from the softest parts of the quieter environments on average, to the louder parts of the loudest environments on average, you're talking about pretty much filling up the entire dynamic range of the human person, of the human, with normal hearing. Meaning that throughout the course of the day if you're leading a normal life, and think back to that Gatehouse study, that Gatehouse data that I showed you awhile ago, throughout the course of the day you're using, a person with normal hearing, is pretty much using their entire dynamic range to capture all the speech energy that can occur in the environment. And so it's just important to kind of keep that in mind because it really shows how important it is to use a multi-channel nonlinear as a technology and of course that's what all hearing aids basically do these days, but the reason we need to do it is because speech at happens at levels all over the place throughout the course of the day. If we dig into it in a little bit more detail, this again, the conversational level speeches range from about 55 to 65 dB and so, you know, even for that level different people are gonna be speaking at different levels, again, the 30 dB range and it's gonna vary by speaker and situations like I said. Finally it will raise with the noise level, but as I said earlier, we typically, the typical talker will not raise consistently enough to match increases in the background noise level. For soft speech you're talking about 40 to 50 dB SPL overall. There's a big drop in the region of the second formant and I'll talk about formants in a few more minutes, but the second formant of voiced speech, whether they're vowels or voice consonants, carries a tremendous amount of information. It's probably the one thing in the speech signal that carries the most information.

When I was working on my dissertation, the late great da-an-o-laf was my department chair, and he also was a speech scientist, and he said basically, you know, the second formant of speech is basically a tracker of what the tongue is doing. The better you can track the second formant the better you understand the sort of articulation that a person is using to create speech sounds and so it carries a lot of information. And what happens with soft speech when you drop your voice to a softer level, because you're changing the amount of effort you're putting in to it, you're not filling out the second formant with as much energy as you would need to. So one of the reasons that soft speech is a little bit harder to hear is not necessarily that the overall level is dropping, but in addition to that the important region around F2 is dropping even more relative to the overall level and so you're losing one of the biggest sources of information when you start to drop your voice to a lower and a softer level. When we move to loud speech it can vary up to 80 to 90 dB, now you get that boost in the second formant region back. So loud speech, the nice thing about loud speech, other than it usually is carrying a lot of emotion with it, but I won't get into those details, loud speech will carry a lot of information because, again, the second formant carries so much information and so when somebody raises their speaking level to get over the background noise one of the things that sort of mitigates the fact that they don't increase your overall level at the same rate that the background noise goes up, is the fact that their increasing the information region around F2 more than their overall level on a relative basis. So you actually are getting more information, bang for your buck, when you get into loud speech. If you summarize these three graphics, well, and here's an indication of the difference between normal speech and loud speech, and one of the things is that for normal speech the longterm level of females, males, and children, are not all that much different. You know we sometimes think that's the case, but it's really whether or not a person is just a soft-spoken person or not.

But when you get into loud speech now you see, notice that the change in the overall shape of the longterm spectra, and that hump out there around, you know, between a thousand and 2,000 Hertz, that's where the second formant is really showing that energy level that I talked about. So when somebody raises their voice to be loud they're also increasing information in that area. It's because when you put in effort like that you fill out those voice segments better so the harmonic structure of the signal shows a boost as you move higher out in the frequency relative to a softer level of production and because of that those second formants take on a greater shape, and a greater form, and a greater emphasis in the signal. If you take a look at, in terms of the audiogram, and speech on the audiogram, that's a common way of taking a look at things these days, we take a loud speech, moderate speech, and soft speech, notice the roll-off of soft speech a little bit more in the higher frequencies as opposed to a little bit overall flatter spectrum with loud speech. So there's some advantages to using loud speech other than the fact that you don't want to be expressing emotion all the time like that. There's a lot that we can understand about just looking at the acoustic speech, the acoustics of speech, understand perception. For example if you take a look at phonemes on the audiogram, you know, and that's one way to look at it. There's another way to talk, to take a look at the aided spectrum if you're doing speech mapping or something like that, that all accounts, audibility accounts for a lot that's going on. One thing to remember is that speech has different value at different frequency regions. So this is a speech importance function from the Articulation Index, or the SII, the Speech Intelligibility Index, and speech is weighted in terms of information content more in the mid to high frequencies. So from about, oh from about one kilohertz which is right about here, out to about four kilohertz, that's where you see the peak in the information content in the speed signal. If you wanna see it in more detail what I've done in this slide is to put the frequency positions of the different sources of information in the speed signal.

For example, the difference between the place of the different affricate bursts occur somewhere from about 1500 out to about four K or so. The place differences for the fricatives is somewhere from about two kilohertz out to maybe about five or six kilohertz. Again, the second formant, which I said is so important for carrying information, that starts somewhere a little below a thousand Hertz, whether you're talking about F2 transitions of the vowels, the second formant of vowels, that starts somewhere around, you know, a little bit below a thousand Hertz and goes out to near 4,000 Hertz or three to 4,000 Hertz, and again, that carries a lot of information. The point being is that when you kind of stack all these up what you notice is that the reason why speech is so important to hear well from about one kilohertz out to four kilohertz, is because there is so many classes of information that are differentiated in that frequency region and so it makes a lot of sense. And so there's a lot of acoustic, there's a lot you can understand about the speech signal by simply looking at the acoustics. But you also have to understand that it's more than just acoustics because when you take a look at the percent correct on a speech task, versus the Articulation Index, what you notice is that you can do pretty well in terms of speech understanding, this happens to be data for connected discourse, you can start, you know, you can do pretty well. You can be up there in the 80% understanding of connected discourse, with only having about half the information in the speech signal available to you. And that just shows you how redundant the speech signal is meaning that there's enough extra cues in the speech signal that the cognitive system can fill in the missing gaps really well. So the acoustic explanation of the speech signal doesn't capture exactly everything that's going on in the speech signal, there is some things that the brain fills in for us which is a very good thing, that's why we have brains, it does a lot of good, the brains do a lot of good things for us. Another thing to remember about the speech signal, another one of these subtleties, is the fact that sometimes the lack of energy is information. We normally think of the speech signal as being the energy level, or the spectrum level, in the speech signal. But remember that we have something called, Stops, and we have Affricates, and we have other sounds that the lack of energy is information.

So for example, this is the spectrum of the sentence, I'll be traveling on business for the next few days. And if you don't know how to read a spectrum this is time going, let me get my arrow up, hang on one second. This is time going across the X-axis here and so this is about, about 2 1/2 seconds of speech material. This is frequency going up the Y-axis. So this is 5,000 Hertz about there, 4,000 Hertz about there. And the lightness of the color indicates the energy level. So the brightest colors are the most intense energy levels. And you see a very clear harmonic structure to the speech signal here, the formants, the second formant, let me kind of track the second formant through here as you see it moving up and down in level, that band of energy through, right through there is what the second formant is all about and that carries a lot of information. But you notice here at the beginning of the, T, in traveling, T is a Stop consonant meaning you stop doing something, you stop the energy flow, you stop the air flow for a moment in time and that carries information. So it is that lack of information for a moment that's important at that point in time and then followed very quickly by the burst energy out of the Stop. Same thing in the K sound in next, remember what the X is, it's a K and an S combination, and so the K has to have a Stop element to it and you see it right there. You see that energy being stopped when you create that K sound and then the S sound that follows it, you know, that create the, the X sound in next, you see that energy level starting there. So sometimes it's the contrast between no energy and a burst of energy that allows the perceptual system to recognize phonemes. So phonemes often times are not necessarily just what the energy pattern is, but is the lack of the energy pattern. Another thing that we know about the speech signal is sometimes you can miss things and your brain will fill it in. For example, if you need to tell the difference between pool and tool, there's a burst of energy called, a plosive burst. So for P it's a little lower in frequency than it is for T, that's just the nature of the acoustics of those two signals.

Then in those two words, pool and tool, you're talking about an unvoiced Stop, the P and the T, which is a closure phase and then a burst of energy at the plosive burst followed by a voiced sound which is the oo vowel, okay? But the, to go from the P, or the T, into the second formant of the oo, and the second formant of oo oo tends to have a relatively low second formant, there has to be a transition phase. There has to go from the position where the placed position of the T has to then drop down in that second formant energy down to where the natural second formant occurs for oo. And so this transition from the consonant to the vowel, or in this case, or in this case the consonant to vowel, but you could also do it vowel to consonant, but this consonant to vowel transition also carries information. Because if you see where the track is having to move from to get down to the lower second formant of the oo, you have a pretty good idea of where it must have started. So one of the things we know about people with hearing impairment, is not only do they have frequency-resolution issues, but often times they'll have temporal-resolution issues. So something like a plosive burst happens pretty quickly and it might be a little bit hard to perceive it, but vowels are longer and they're more stable and so you can, it's easier for a hearing-impaired individual to perceive a vowel because of the length and the stability of it. So in this particular project what the authors did was simply get rid of the plosive burst information that would've occurred for pool versus tool, and then see whether or not people could fill in and tell whether or not the word, the original word was pool and tool and the point is they can. Even though you don't have the plosive burst information, it was taken out of the wave form, the track of the formant also gives information. So the reason I pull that out is I just think it's a really cool indication of how well the cognitive system can fill in missing information. One little point I wanna make about the speech signal that I think sometimes people forget is that vowels are not low frequency and consonants are not high frequency. Meaning that often times, you know, people will say that, well, vowels are low frequency, consonants are high frequency. Vowels have a lot of energy in the low frequency and unvoiced consonants have no energy in the low frequency.

Like voiced consonants often times will have some energy in the low frequency, and very important, voiced or vowels, which are always voiced, because of this F2 and F3 that they have, the second formant and the third formant, they also have energy in the high frequencies. So this is, this a measure, or a wave form, of the phrase, ages 60 to 65, and this is filtered at 1500 Hertz. And so what you notice is below 1500 Hertz, in the area that is considered to be low frequency, you see the big burst of energy that goes along with the vowels in ages 60 to 65, okay? But what you notice up in the high frequency is you also see those energy bursts up in the higher frequencies in addition to other energy patterns that go along with those unvoiced consonants up there in the high frequencies. So the point being is that you'll also get vowel energy up in the high frequencies so it's a little bit of a dumbing down to say that speech doesn't have energy, or vowels don't have energy in the high frequencies, that they're only low frequencies. They dominate it by low frequencies, but it's not missing out of the low frequencies. Anyway, so, all I want to do with that discussion was just to talk about how potentially complicated the speech signal can be and it's a little bit more subtle and complicated than sometimes we refer to. Okay, at this point I want to change topic slightly and talk about Soundscapes and the sense of Place. And what I mean by that is that sound in our environment has meaning. It gives a sense of where you are and what you can expect out of a situation. And like I said early, early on in this talk, often times we sort of dumb-down the idea that all noise is bad and all speech is good. But reality is sometimes noise carries good information with it. So I wanna take a look at it in a couple different ways, okay? The first thing is to recognize when hearing loss is relevant, and what I mean by that is that you can be in a lot of different situations where the person doesn't necessarily feel that they're at a disadvantage because they're not trying to communicate. So often times we think of sound as only being, or background noise as being disruptive to the speech signal, which of course it is, but it's only disruptive if the person is trying to communicate.

And this is a project that a group of colleagues of mine at our Eriksholm Research Facility in Denmark did, where they took a look at the diaries that, that individuals filled out about their listening difficulties that they ran in to and they tried to compare that to Soundbar readings of the overall levels in their environments. And they got a little bit confused because there's a lot of situations where there was a lot of noise in the background, but the people weren't necessarily complaining about the noise. Well, the reason they weren't complaining is that weren't trying to communicate in those situations. So again, background noise from this perspective, is only relevant if it gets in the way of communication. So that's one, I think, very interesting level to keep in mind, so, just so you know. The next way that I wanna look at it is in terms of what sound means to a human and I hearken back to a book chapter that was written 50 years ago at this point in time, or probably more if I remember. It was written by a guy named, Donald Ramsdell, and he was a, he was a psychologist, or a psychiatrist, that was brought in after World War II, to try to better understand why returning Veterans who lost their hearing in World War II, were having so much trouble with depression and even suicide. And in a very classic chapter that he wrote back at the day that was repeated over-and-over again in a classic textbook from the day called, "Hearing and deafness", by Davis and Silverman, he talked about the psychology of sound and how when you have hearing impairment you miss certain aspects of sound. So one of the things that he did is he separated sounds out into three levels. One level is the symbolic, or the social level, and that's the level at which, which individuals use sound to carry information, and the most important, to carry symbolic-based information. And the most common use of that of course, in our field would be speech, that we have created speech sound as symbols that carry meaning, and that's something that hearing-impaired individuals of course miss. The second level that he talked about was the signal or warning level and that's the way we use sound to keep us alive basically. So we hear the truck coming down the street, we hear the smoke alarm going off, and to this day we even hear our phones beep when somebody sends us a text message, you know, that keeps us alive too.

But so the idea being that, that individuals use sound to gain information about things that they should know about in the environment, the buzzing of a bee or something like that. But the third level in what Ramsdell really felt was missing for a lot of these patients and that was what he refers to as the primitive level and that's that connection to our environment. And I think the best way to describe it is just to hear his words. So I'm gonna read you a little bit that he wrote. Finally, and most basically, sound serves neither as symbol, nor as warning, but simply as the auditory background of all daily living. At this level react to such sounds as the tick of a clock, the distant roar of traffic, vague echos of people moving in other rooms in the house without being aware that we hear them. These incidental noises maintain our feeling of being a part of a living world and contribute to our own sense of being alive. We are not conscious of the important role these background sounds play in our comfortable merging of ourselves with the life around us because we are not aware that we hear them. And I think he does such a good job of capturing this idea that sound connects us to the world, that sound is supposed to be there. That this idea that all noise is bad and needs to be removed from the situation that's really not the case because it connects us to sounds and such and to the world around us in such an important way. And there's a whole area of science called, Soundscapes. And in Soundscapes it is the idea of understanding what a place should sound like. That every environment should have a certain signature sound. That Starbucks should sound like Starbucks, and Starbucks should sound like Starbucks differently at eight o'clock in the morning than it does at 2:30 in the afternoon. That a restaurant should sound a certain way, a quiet home should sound a certain way. And part of our experience of the world around us is this multi-sensory experience of knowing that the environment that we're in should sound a certain way. And that if you lose that sound of the environment, if you lose that sense of what the sound of that environment is all about, then you lose a certain connection to the world around you. And it becomes very important to think of it that way in relation to what we're trying to do for hearing-impaired individuals.

You know, this Tarzan level of dumbing it down to say, you know, noise bad, speech good, really doesn't capture what sound is supposed to do for the humans. Now, there's no doubt that hearing-impaired individuals need the background noise to be controlled effectively. Meaning that if the sounds in the background, the competition in a background, are making it hard to perceive certain things, then somehow that has to be controlled for the person there's no doubt about that. But that doesn't mean that all sound has to be eliminated for the person. The sound of a space is important and often times a goal when we talk about fitting amplification, is to try to reconnect the person with the world of sound, but that doesn't necessarily mean just the speech that they're listening to at that moment in time. It also means the sounds of the world around them. And so hearing-aid technology has to run a fine line of trying to eliminate enough of the background competition to allow the person to hear and understand the speech that they're interested in, but without isolating the person over too much. And at Oticon this becomes a very important part of the way we do technology and I'm not going to go into the details of the way we approach it from a technology standpoint, but we have a very firm opinion about the way we wanna create technologies for patients in such a way that we are trying to make sure that they stay connected to the world of sound and, because sound carries value. And this idea of looking at it from this standpoint of Ramsdell of knowing that sound is an important emotional part of the way we deal with our environment becomes very important. The final topic that I wanna talk about is a topic that we refer to as brain versus machine. And, as I said, there's a whole AO show on this and so I'm only going to touch on it briefly. But the idea of brain versus machine is that I've brought a few examples up during the course of this seminar about how important the cognitive system is in being able to deal with background noise and allow a person to hear and understand speech and the way it sort of links information and it finds missing information and all those sort of things.

And that's very different in dealing with speech understanding in a complex environment then the way most signal processing in hearing aids these days deal with the situation. So I wanna give you just an example to show you where the difference occurs and that is, you know, in general we're talking about the difference between subtraction versus completion. That signal processing in hearing aids these days pretty much works for the most part on subtraction principles, it's trying to make the bad stuff go away, noise bad, speech good, so it's trying to make the noise go away. But the reality is we can only do that to a certain degree and the other reality is, especially from the Soundscape way of thinking about it, there's some value in the sounds in our environment. So another way of thinking about it is to create technologies that work on the concept of completion, meaning you're trying to give more information to the signal, to the brain. Like I said one of the problems that the brain runs into is that the information that the brain gets from speech in a complex environment is often times is lacking, because of the peripheral hearing loss, you're not getting a good set of information up to the cognitive system. So one of the approaches you could do in technology, and one of the approaches we use at Oticon, is the idea trying to fill in as much information as possible. But anyway, there's a lot that goes into that discussion, but I just wanna bring you one really quick example of how that works and that's the importance of bandwidth. And this is data and there's other studies that show the same sort of thing, but this is data from a study done at Medical University of South Carolina, Amy Horwitz and her colleagues, where they took a look at the bandwidth of amplification and the ability to understand speech and noise. And remember what happens when you increase bandwidth in understanding speech in typical competition. When you increase the bandwidth of amplification the idea is to bring more speech available, more speech cues available to the patient, which is important, but it also is gonna bring more noise into the situation. So increasing bandwidth also carries along with it increasing noise for the patient. And so you would think, well, those two should sort of cancel each other out, but consistently what you typically see is improvements in speech understanding as you increase bandwidth for individuals, even though you're bringing more noise into the situation and that's an indication that the brain likes the extra information. The brain is basically saying, I'll deal with the extra noise you give me, just give me more speech information to be able to latch on to. 'Cause like I said earlier, the brain will find whatever information it can to understand the speech signal in a noisy environment. So if you extend the bandwidth in order to bring in more speech information, even if it's bringing in more noise, the brain typically is okay with that because as long as it's good information it'll use it, and it'll just deal with the noise the best it can. So again, the idea of bandwidth it's just a very good example of kind of taking a look at the way the brain likes to get information and a lot of times the signal processing, that we use in hearing aids, is becoming more and more focused on this completion aspect and not necessarily on the subtraction aspect.

So again, this idea that noise bad, speech good, the Tarzan approach to thinking about speech and noise, it actually doesn't really capture what we really wanna do. Yes, sometimes noise is bad, yes, it needs to go away, yes, patient's need an improved signal-to-noise ratio, but there's a lot of subtleties of understanding what speech signals are really like, what noise signals are really like and then understanding what we wanna do in technology. I wanna thank you for your time today. Hopefully you got a few kind of interesting ways of thinking about this very important topic in our area. Obviously, you know, looking at sound environments is something that we do all the time when we talk about fitting hearing aids and helping patients, but understanding at a deeper level the nature of speech, and the nature of noise, becomes a very important part of where we go about doing clinical practice. So if you have any questions or comments you can always contact me. My email is D.Schum@Oticon.com and I wanna thank you for your time

.