My Ears Are Alight

Per Hiselius*
3M Personal Safety Division, 3M Svenska AB, Box 2341, Malmstensgatan 19, SE-331 02 Värnamo, Sweden

Résumé

Les capacités du système auditif humain sont phénoménales. La plage dynamique qu'il supporte, la plage de fréquences qu'il couvre, et également, sa capacité à détecter et identifier la parole en présence de bruits parasites sont étonnantes. Dans la vie quotidienne, on utilise ces capacités de différentes façons: la communication orale, les alertes et les alarmes, l'analyse des appareils et des machines comme les ordinateurs et les voitures. Par exemple, "Est-ce que mon appareil est en marche ?", "Est-ce qu'il semble fonctionner normalement ou est-ce que quelque chose ne va pas?". De plus, nous utilisons aussi notre système auditif pour diverses formes de divertissement. Le revers de la médaille de ses capacités magnifiques est qu'elles posent des difficultés par exemple lors de la conception des bâtiments, des machines et d'appareils électroniques comme les téléphones mobiles, les ordinateurs, les écouteurs, les microphones, etc. Et bien que nous ayons cette capacité phénoménale à comprendre la parole dans les situations difficiles, on a souvent du mal à entendre ou du mal à comprendre. De plus, le système auditif humain s'abîme facilement. Ce document présente d'anciens et de nouveaux résultats de recherches liés aux capacités du système auditif et quelques-uns des challenges posés par ce dernier. Le contenu de cet article a été présenté lors de la Semaine canadienne de l'acoustique 2014 (Acoustics Week in Canada 2014) lors de l'une des trois présentations plénières invitées.

Mots clefs: système auditif, plage dynamique, plage de fréquences, parole dans le bruit, perte auditive

Abstract

The capability of the human auditory system is phenomenal. The dynamic range it can handle, the frequency range it covers, and, not the least, its ability to detect and identify speech in the presence of interfering sounds is astonishing. In daily life we use this capability in many ways. We use it for speech communication as well as for alerts and alarms. We use it for analysis of devices and machines, e.g. our computers and cars. Is it on? Does it sound normal, or is something wrong? We also use it for various forms of entertainment. However, there's a flip side to the great capability. From an engineering point of view it poses challenges when designing buildings, machines, and devices such as phones, computers, headphones, microphones, etc. And although we have a phenomenal ability to understand speech in challenging situations, we often mishear or misunderstand. Human hearing is also quite easily damaged. This paper presents old and new results related to the capability of our hearing, and some of the challenges related to the same. The content of this article was presented at the Acoustics Week in Canada 2014 as one of three invited keynote presentations.

Keywords: auditory system, dynamic range, frequency range, speech in noise, misheard lyrics, hearing loss

1 Our Hearing is Remarkable

The international space station orbits the earth at an altitude of about 400 km [1]. Let's assume that it sends out a 1 W signal from an omnidirectional antenna. By the time the signal reaches our planet the 1 W signal is spread out across a sphere having an area of two million square kilometers. Assuming no losses or reflections along the way the intensity would at that point be approximately 0.5×10^{-12} W/m^2. A sound wave of that intensity is audible for many people if presented as a pure tone around 3.5 kHz in a perfectly quiet room.

The shape of the human ear canal is quite complicated and varies significantly between individuals. But let's assume an ear canal having a diameter of 7 mm and a length of 26 mm. Let's also assume it has a perfectly cylindrical shape, rigid walls, and a rigid eardrum. An intensity of 10^{-12} W/m^2 equates to a sound pressure of 2×10^{-5} Pa which is defined as0 dB, and can be perceived under perfect circumstances. An insert type headphone with a speaker diaphragm covering the entire cross-section of the ear canal, would only have to move about 10 pm, i.e. 10^{-11} m, peak-to-peak, to generate this sound pressure, at which point the movement of the eardrum is in the order of 1 pm, or 10^{-12} m [2]. To produce a sound pressure level of 120 dB the diaphragm would still only have to move 0.01 mm. The required displacement of the diaphragm, and the movement of the eardrum, may be compared to the radii of atoms, ranging between 30 and 300 pm [3].

In short, the human ear is sensationally sensitive. At the same time even some of the more stringent safety regulations around the world, such as the Swedish Work Environment Act [4], allow workers to be exposed to levels up to 115 dB(A) - although the permitted daily noise exposure will be reached within half a minute.

* per.hiselius@mmm.com

A dynamic range of 115 dB equals a ratio of 0.56×10^6, i.e. the sound pressure at 115 dB is almost one million times higher than that of 0 dB. Expressed in terms of intensity, a sound wave at 115 dB carries almost one trillion times more energy per second than one at 0 dB. The dynamic range covered by our hearing is truly amazing.

The frequency range covered by the human hearing spans over 3 decades, or 10 octaves. As a comparison, a microwave antenna advertised as an "ultra-wideband microwave antenna" may cover significantly less than one octave. Some motorcycle engines may rev up to 12,000 rpm, at which point each piston completes 200 cycles every second - propelled by 100 explosions every second - typically generating 10 kilowatts per cylinder, or more. An even higher rpm would mean even more frequent explosions, and thus more power. But the inertia of the pistons, valves, fuel mixture, and exhaust fumes, makes it an enormous challenge to increase the revs without losing efficiency. To perceive 20,000 Hz our eardrums need to complete 20,000 cycles every second, propelled only by the sound pressure, having intensities in the order of magnitude of nanowatts[i] per square meter. This must be considered a remarkable achievement.

Another remarkable achievement of our hearing ability is signal and speech recognition. In a situation where the signal-to-noise-ratio (SNR) is only 10 dB a person with normal hearing will still be able to understand most of what is said [5], regardless of the type of speech material being used. In a study [6] using military call signs it was shown that, in white noise, test subjects were able to correctly perceive 65,88% of the call signs at an SNR of -18 dB. To compare the RMS levels of two very different types of signals can be problematic. An SNR of -18 dB implies that the amplitude of the noise is 8 times higher than that of the speech. However, speech is an irregular signal with pauses, and will have bursts that are significantly louder than the average RMS level. Nonetheless, as can be seen in Figure 1, at an SNR of -18 dB the speech signal is completely masked by the noise. Even so, a trained person still has the capacity to correctly perceive almost two out of three call signs. This would not be possible if the speech material consisted of totally random words selected from an infinitely long list. Further, white noise is primarily a high frequency type of noise[ii], and the auditory masking will therefore be less severe than if the noise had been a low frequency type of noise, with a spectrum more similar to that of speech. The test participants in [5] were also well motivated, and situated in a lab environment without distractions. But the achievement is still astonishing.

In a recently developed test [7] aimed to investigate the impact hearing protectors have on the perception of speech in noise, it was shown that in a low frequency type of noise test normal hearing test participants were able to correctly identify more than 50% of the words presented at an SNR of -17 dB.

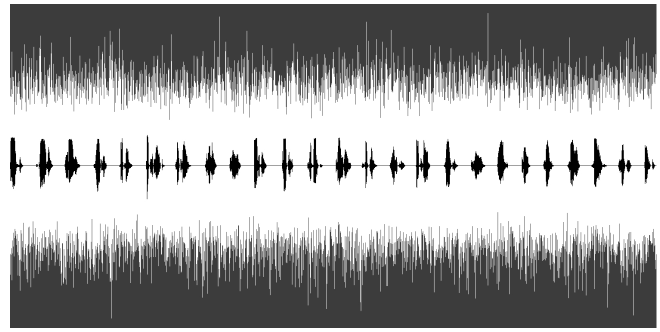

Figure 1: Time signals of words (black signal) and white noise (white signal) at an SNR of -18 dB.

This was again under perfect circumstances, but the result supports the Blue-Terry & Letowskistudy [6]. In fact, in the more recent study [7], which had a more challenging low frequency type of noise, test subjects quite frequently managed to correctly identify four consecutive words in an SNR of less than -20 dB. This could happen by chance, given the limited number of words used in the test, but it would (almost) literally be a one-in-a-million chance. It's safe to say that the capability of the human auditory system is remarkable.

2 Challenges and Risks

The great capability of our hearing is valuable in our daily life, but also poses challenges. Being exposed to undesirable sounds affects our quality of life, and even our health. The sensitivity of our hearing, and the fact that it never sleeps, complicates the planning of cities, roads, railroads, airports, industries, etc. For many people one of the biggest financial investments of their lives will be buying a house or a car. Because of our sensitive hearing, the design of houses have to be more complex due to the need for soundproofing and they thus become more expensive to build. In a modern car, where reducing weight is critical for reducing the fuel consumption, soundproofing may account for up to 25% of the total weight [8a, 8b]. Even if it's just half of that it will still add more weight to a normal car than does the engine and transmission. There's no doubt that acoustics has a major impact on our society, and on our lives.

The high dynamic range adds further complications. We use decibels, which is a logarithmic scale. But most people cannot intuitively interpret a scale where 80+90=90, or 92, depending on whether we are adding intensities or sound pressures. And because the sensitivity of our hearing is level- as well as frequency dependent- we need different weighting schemes, such as dB(A) and dB(C). In spite of great effort there are still no perfect measures neither to predict perceived loudness, nor to accurately predict the risk of acquiring a hearing damage. There are several measures available, but they all have shortcomings. The capability and complexity of our hearing is simply too great to allow it to be explained by simple metrics. Consequently, one can often see the wrong unit of measure being used, and the right one being misinterpreted.

[i] One nanowatt per square meter equals 30 dB.

[ii] A source producing white noise will, on average, produce the same power at all frequencies. However, per frequency band the power will increase by 3dB/octave which is why it's perceived as a quite sharp, high pitched, type of noise.

Music is an important part of most societies. It's a major industry, and plays a major role in many people's lives. When it comes to loudspeakers and headphones the hunt for perfect fidelity or for the perfect sound - which may not necessarily be the same as perfect fidelity - is still on. Manufacturers of microphones are still trying to match the frequency range as well as the dynamic range of our hearing. It's still is a major challenge to match the capability of our hearing.

While the lower end of the dynamic range is challenging from an engineering point of view, the opposite end of the range is problematic from a quite different point of view. It's fairly easy to produce high sound levels. All you need is a hammer and a hard surface to produce peak levels that are hazardous. But the short duration of the peaks make us underestimate the level [10]. Most modern movie theaters are capable of producing sound levels that are damaging, not to mention the levels produced at many music events. An intense light is unpleasant, and we will automatically look aside or close our eyes. Our hearing, on the other hand, can tolerate harmful levels without necessarily causing us any pain or discomfort; when listening to music, it may even provide great pleasure. The sound of roaring engines may also be perceived as immensely enjoyable, while inflicting permanent damage to the hearing of the listener. Unfortunately the annoyance (or perceived pleasure) generated by loud sounds, does not correlate with their potential to cause hearing damage. And even if we would like to, we can neither close our ears nor 'look away' from sound.

There was a time when reproducing music at high sound levels required some effort. Powerful amplifiers were heavy and expensive, and had to be combined with big, bulky, loudspeakers. Today we just have to buy an mp3 player and a pair of headphones to produce extreme levels. As previously mentioned, it's easy to produce high sound levels in a small enclosure such as an ear canal. This has been recognized in Europe, where there's a directive requiring compliance to a standard stating the maximum permissible output voltage of music players, as well as the maximum sensitivity of headphones [11]. The combination of the two ensures that the average level does not exceed 100 dB(A). However, for workers there's a European directive [12] stating an exposure limit equivalent to 87 dB(A) for eight hours, and many countries have enforced a more strict limit equivalent to 85 dB(A) for eight hours. With a music player producing 100 dB(A) a dose equivalent to 85 dB(A) for eight hours will be reached within 15 minutes. To reach the 87 dB(A) limit will take an additional 10 minutes.

For some reason we seem to accept damage to our hearing when listening to music, or attending motor shows and other venues with high sound levels. But it's not acceptable at our workplace, and we would most likely not accept damage to our eyes when watching a movie or when attending concerts and other events.

When our hearing is damaged it will often result in a shift of our hearing threshold, which can be measured. But again, our hearing cannot be described by simple metrics. Hearing damage may result in many other types of complications, such as tinnitus, reduced dynamic range, reduced frequency resolution, reduced temporal resolution, and hyperacusis (abnormal sensitivity and pain caused by even relatively quiet sounds). As a result, the performance and capability of our hearing may be greatly reduced also in ways that are more difficult to quantify, and for which a hearing aid cannot compensate. A common complaint among people with a sensorineural hearing loss is that they have difficulties following a conversation in the presence of an interfering noise - especially if that noise is made up from other people's voices, i.e. babble or cocktail noise. When our amazing ability to understand speech in noise is reduced it may greatly affect our professional, as well as social, life. Hearing damage may also affect our ability to appreciate music and sounds of nature.

Communication is a vital part of our life. Today, phones are used more than ever for various forms of communication. When it comes to speech communication a mobile phone, or cell-phone, is quite an advanced piece of equipment. Again, our excellent hearing makes it a challenge to design the device. When comparing a modern phone to an old land-line phone the most striking difference is perhaps in the shape of the device. A traditional land-line telephone is typically designed so that the mouthpiece, i.e. the microphone, can be positioned close to the mouth. Further, the earpiece is typically designed so that it can cover most of the outer ear, thus creating a baffle for the loudspeaker (earpiece) and also allow ambient noise to be reduced by at least partially blocking the outer ear. The long distance between the loudspeaker and the microphone will also reduce the risk for acoustics feedback and echoes. A traditional land-line phone is thus well suited for its purpose, unlike a modern mobile phone.

Not only do size constraints on modern phones require small components, but the microphone is often positioned far away from the mouth of the talker, and relatively close to the earpiece. As a result, the microphone gain needs to be high, and the sound from the earpiece is easily picked up by the microphone. This sound will not only transmit via the surrounding air on the exterior side of the phone, but also via the interior of the phone, and it may be air-borne as well as structure-born, or a combination of the two. In fact, at the microphone, the far end voice may be considerably louder than that of the person talking into the microphone. This calls for several actions. Without active echo cancelation in the near end phone, the far end talker would hear her or his own voice as a clear, distinct, and loud echo. To cancel out the echo without affecting the desired signal is difficult - especially when using miniature speaker components pushed to the limit, providing a fair amount of distortion. A great distance between the mouth and the microphone makes the signal to noise ratio at the microphone poor if the phone is used in a noisy environment. To lessen this problem, phones use multi-microphone solutions for noise suppression. Further noise suppression may be applied by the near end phone, by the network, as well as by the far end phone. To reduce noise is not difficult. The challenge is to do so without affecting the quality of the speech signal. To actually improve intelligibility is even more difficult because of our amazing ability to perceive speech in noise. When applying noise reduction, inevitably the speech signal will be affected to some extent, and this may instead have a negative effect on our ability to correctly perceive what is being said - even if the SNR is improved.

Our phones do not only reduce noise, they also intentionally add noise. To reduce the bandwidth required for a phone call every cell-phone has a voice detection system and will only transmit a signal when it detects a speech signal. At the receiving side, the far end, it may therefore become quiet when there's a pause in the speech at the near end. However, we constantly use our hearing as a tool to analyze what's going on around us, and we are likely to assume that a call has been terminated if it becomes perfectly quiet. On the receiving end the phone will therefore generate a noise similar to that of the background noise from the far end, and in every pause it will apply this artificial noise, called comfort noise. There is, of course, additional signal processing going on during a voice call. To further improve the perceived sound quality there's compression and equalization applied, not to mention all the coding, recoding, and decoding going on in the phones and networks. However, since our hearing is so well adapted for recognizing natural speech in noise it's again difficult to actually improve intelligibility. On the contrary, signal processing can easily reduce both the intelligibility and the perceived quality of a voice call.

Another, related, area is speech recognition and text-to-speech. Every smartphone has some form of text-to-speech engine installed or available for download. They can do an excellent job, and it's not always easy to tell the difference between a genuine human voice and an artificial one, at least not if there's a background noise adding some degree of masking. However, many words are spelled the same but pronounced differently, and have different meanings, i.e. heteronyms such as address, bow, row, wind, etc. These words will generally be used in some kind of context making it obvious to a human how to interpret, and pronounce, the words. This task is much more difficult for a phone or a computer. We use the very same ability when we interpret speech where homophones may be an issue, i.e. words pronounced the same way having different meanings.

If a speech signal contains homophones, is too soft to be clearly perceived, if there's an interfering sound masking a major part of the speech, or when a word is mispronounced, we use additional cues. The subject and context of the conversation, the situation and general circumstance, and, not the least, specific visual cues, may dramatically improve our ability to understand what's being said. This is why the speech recognition of our smartphones cannot compare to that of a real person. Our ability to recognize patterns, and to 'fill in the gaps' is quite remarkable. This is partly why we, under perfect circumstances, can correctly recognize words even when the signal to noise ratio is less than -20 dB. However, this ability may backfire. When we fill in the gaps we sometimes get it wrong. As a result we cannot trust our hearing. An obvious proof of this is all the internet sites posting examples of misheard lyrics[iii]. On a more serious note, the fact that we quite frequently mishear can also cause or worsen disputes and conflicts.

3 Conclusion

In conclusion, the capability of human hearing is truly phenomenal. It has a great impact on our lives and our society, and still poses challenges for engineers. Unfortunately it's also very susceptible to damage, and we can't trust our intuition when assessing the risk related to a noise exposure. We need to recognize this fact, and make sure we protect our hearing if we don't want to lose our ability to enjoy music and sounds of nature, or lose our amazing ability to understand speech also in challenging environments. And finally, as much as we would like to, we cannot always trust our hearing.

References

[1] ISSracker, August 2014, *http://www.isstracker.com/*

[2] Hallett, F. R., Speight, P. A., Stinson, R. H. 1977, Introductory Biophysics, New York: *John Wiley & Sons*.

[3] Wikipedia 2014: h*ttp://en.wikipedia.org/wiki/Atomic_radius*

[4] AFS 2005:16, Arbetsmiljöverkets föreskrifter om buller samt allmänna råd om tillämpningen av föreskrifterna, *ISBN 91-7930-455-9, ISSN 1650-3163*.

[5] Steeneken, H.J.M.1992. On measuring and predictingspeechintelligibility.Doctoratethesis*ISBN:90–6743–209–1*.

[6] Blue-Terry, M., Letowski, T. 2011. Effects of white noise on Callsign Acquisition Test and Modified Rhyme Test scores. *Ergonomics*, 54, 139-145.

[7] Hiselius, P., Edvall, N., Reimers, D. 2014, To measure the impact of hearing protectors on the perception of speech in noise. *Int J Audiol*. 2014 Dec 30:1-6(Epub ahead of print).

[8b] SEMCON 2014a: *http://www.semcon.com/en/Industry-solutions/Automotive-en/Technology-roadmap/Light/*

[8a] SEMCON 2014b: *http://www.semcon.com/Global/Docs/AUTOMOTIVE/Semcon_CFRP_Acoustics.pdf?epslanguage=en*

[9] Peus, S. 2004, Modern Acoustic and Electronic Design of Studio Condenser Microphones, *AES E-Library*, AES Convention:116 (May 2004)

[10] Brüel. P. V. 1976. Do We Measure Damaging Noise Correctly?

[11] *EN 60950-1:2006*: Information technology equipment - Safety -Part 1: General requirements.

[12] *Directive 2003/10/EC* of the European Parliament and of the Council, on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (noise).

***Editorial Note:* This manuscript is published as an invited tutorial paper and meets peer-review criteria that are different than for regular research papers.**

[iii] Recommended phrases to use when searching for misheard lyrics: "me ears are alight", "misheard lyrics blinkbox music survey", and "commonly misheard lyrics".