Whisper

Beyond Frequencies: Artificial Intelligence, Sound Patterns, and the Whisper Hearing System

by Andrew Song, Andreas Thelander Bertelsen, and Joseph Antognini

Abstract

Artificial intelligence (AI) can be used in hearing aids to process sound in a more targeted way. This is done by understanding the patterns within sound signals instead of only using frequency data. This shift in approach, combined with the ability to learn and improve, allows the Whisper Hearing System to process sound in a new and optimized way while delivering a learning hearing aid that gets better over time.

Introduction

Artificial intelligence (AI) is an emerging field that is already showing great promise in how we solve various problems across consumer convenience, medicine, and commerce. As the term AI is often used interchangeably with deep learning, deep neural networks, and machine learning, it is clear the technology itself is becoming ubiquitous. AI can now be found in Apple's personal assistant (Siri) that answers your pressing questions, Nest's learning thermostat that predicts when you might need a burst of air conditioning, and Google's Translate feature that makes foreign languages accessible to you.

This paper delves into the core artificial intelligence powering the Whisper Hearing System. It breaks down how the system uses AI to process sound in a new way, and why the Whisper Hearing System is a hearing aid that gets better over time. This can enable hearing care professionals to better support their patients' hearing needs.

The Learning Cycle Behind Artificial Intelligence

Artificial intelligence algorithms offer a powerful toolbox that can be used to solve complex computing problems. Products with AI use these algorithms to tackle real-world challenges. While solving the challenges themselves, these products are also refining their understanding of the problem through exposure to new data. That data is incorporated into the AI system and a new cycle of learning and improvement for the product begins. We call this the learning cycle.

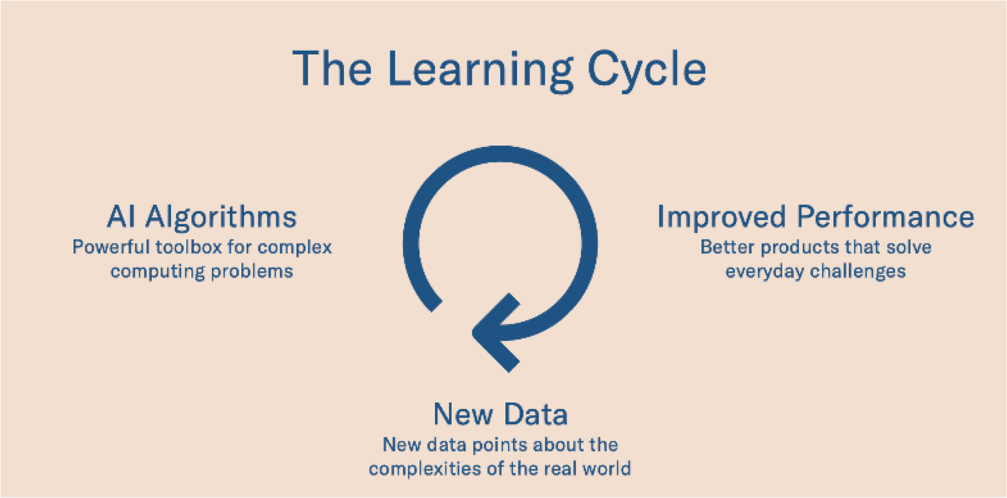

Figure 1: The Learning Cycle. Artificial intelligence algorithms power products which can solve everyday problems in a new way. With new data added, the AI algorithms improve and generate a product that gets better over time.

A real-world example is a mobile app that identifies a breed of dog just by taking its picture. The AI algorithms in the app use an existing set of dog images to extract patterns that then allow the app to predict a new dog's breed from its picture. As the AI in this dog breed-identifying app is used over time, it encounters new breeds not previously seen. It is at this point that the greatest benefit of AI can be leveraged: its ability to learn and improve.

By adding examples of the new breed, the AI algorithm is able to isolate this new breed's patterns to identify it in the future. It is important to note that by adding this new data, the overall data set that the AI is built upon becomes more reflective of the complexities of dog breeds in the world. All users of the app benefit from this improved product.

Whisper leverages this same learning cycle to deliver a hearing system that gets better over time.

From Frequencies to Sound Patterns

Today, traditional hearing aids process sound by taking in the acoustic signal from its onboard microphones and using various processing methods to adjust and manage that sound signal for the specific hearing loss of the wearer. These different methods include compression and gain management, adjusting processing speed (e.g., attack and release timings), feedback management, spectral noise reduction, scene analysis and classification (e.g., noise, music, etc.), and utilizing directionality. The specifics of how these methods are set up, in addition to how the hearing aid fits a wearer's specific preferences and hearing loss profile, determines how the sound signal will eventually be modified.

For example, imagine a hearing aid wearer is in a noisy environment as someone speaks to them. Typically, a scene analyzer within their hearing aid will detect this noisy situation and start to change the processing as a result. First, the hearing aid will increase the level of spectral noise reduction applied to reduce stationary background noise. If the situation benefits from it, the hearing aid may also turn on a directionality feature to help the wearer focus on sounds coming from a particular direction. Finally, the hearing aid will try to perform gain management. It will increase gain at common speech frequencies (e.g., 500 Hz to 3.5 kHz) and lower gain outside of this frequency range.

> "AI lets a hearing aid process sound in a whole new way."

Gain management across different frequency bands is useful when there are not more sophisticated tools available, but it is limited in terms of its effect on a speech signal in noise. The spectrograms in Figure 2A and Figure 2B are visual representations of the speech in noise signal, with frequency on the y-axis, time on the x-axis, and the brightness of the region mapping to the sound intensity level at that point (brighter means more intense). Figure 2A shows the overall spectrogram of a man speaking in noise, with the areas of voice activity circled. In Figure 2B, the shaded blue area represents the frequency range a hearing aid amplifies more in noise to help bring out the speech signal — between 500 Hz and 3.5 kHz — while the red shaded areas are other frequencies outside this targeted range. Although there is definitely more speech signal in the blue area relative to the red areas, it is also very clear that there is significant noise in the blue area that becomes amplified as we increase gain. Conversely, in the red "non-speech" frequencies where a hearing aid will reduce gain, we know there is substantial and important speech signal content that will be diminished as well. This can hurt overall speech clarity as fricatives like the 's' and 'f' sound can become harder to distinguish.

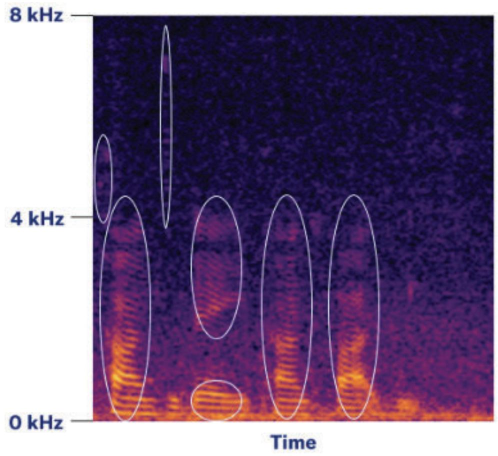

Figure 2A: A sound spectrogram of a speaker in background noise. Circled areas highlight regions of voice activity in the spectrogram.

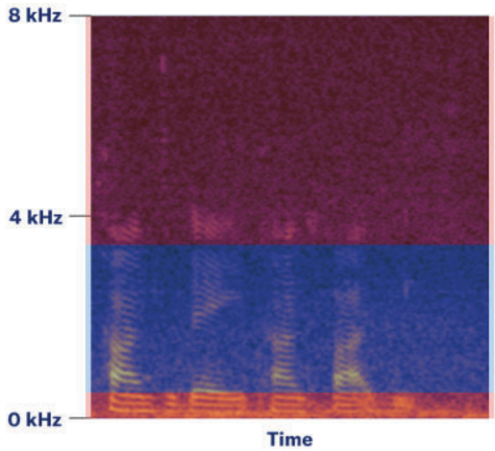

Figure 2B: Managing gain using only frequency information. Because of the limitations of using only frequency information, noise inside of the blue "speech frequencies" area will be amplified when gain is increased, and important voice information in the red regions will be diminished when gain is reduced.

Over time, hearing aid engineers have developed more sophisticated rules for how the red and blue boxes should be used, but an inherent challenge still exists when confined to this traditional approach. Hearing aids lack the specificity to pick out certain sound signals in specific situations because they are limited to only using coarse frequency bands to conduct gain management. Either hearing aid engineers will overselect and amplify unwanted sounds (as is the case when noise is amplified within the blue areas) or underselect and diminish important signals we are hoping to preserve (as is the case when speech sounds are diminished in the red areas).

This is where artificial intelligence can help us. AI (specifically, deep neural networks) provides a more powerful and more precise approach to processing sound. With its ability to learn and recognize patterns, artificial intelligence can target the patterns within the sound signal rather than only looking at the frequency content. As a result, AI lets a hearing aid process sound in a whole new way. It can look specifically at each area of the spectrogram to highlight the important sound signals. A hearing aid like the Whisper Hearing System, based on this type of artificial intelligence, can shift from analyzing frequencies to analyzing sound patterns. Figure 3 illustrates this novel approach.

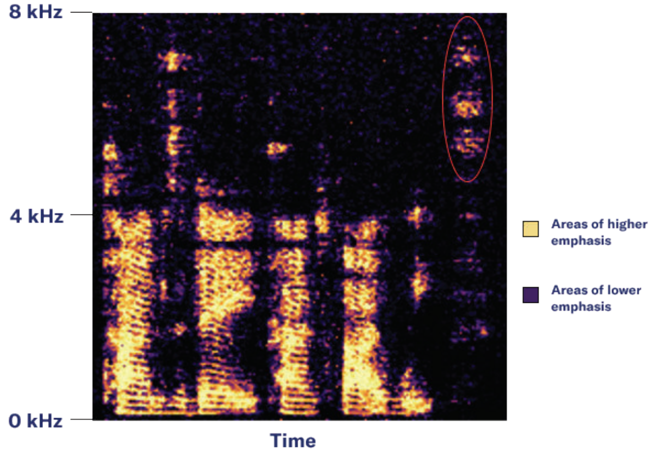

Figure 3: By understanding acoustic patterns, an artificial intelligence system can analyze sound signals more deeply to highlight specific areas of emphasis. The red circle shows a voice area which might have been missed using traditional processing methods but can be targeted by an AI using its knowledge of sound patterns.

Whisper's Sound Separation Engine

The Whisper Hearing System analyzes sound patterns using its proprietary Sound Separation Engine®. The Sound Separation Engine is based on an AI model designed to understand two key parts of sound: sources and environments. Sources are objects that create sound by emitting waves of pressure through a medium, such as air. Environments change a sound by affecting the movement of these sound waves. The unique properties of sources (e.g., human voices, TVs, and other everyday objects) and environments (e.g., living rooms and concert halls), listed in Table 1, form the predictable everyday patterns that the Sound Separation Engine is designed to understand.

Sources	Environments
- Pitch (frequency) - Volume or intensity (amplitude) - Harmonics (timbre) - Patterns and repetition - Variance (stationary vs. non-stationary)	- Room size and shape - Density (water vs. air) - Surface quality (smooth vs. spikes) - Materials (foam vs. metal) - Interfering objects

Table 1: Properties of Sources & Environments

The strength of Whisper's underlying model allows the Sound Separation Engine to process sound in a new way: by targeting patterns within a sound signal. The Sound Separation Engine is created with thousands of hours of unique sound sources in environments and learns to distinguish those portions of a sound signal that correspond to sources of high importance to the wearer from those sources that are less relevant. This is what makes it possible for the Sound Separation Engine to identify speech even in high frequency areas that would have otherwise been diminished by traditional processing methods, such as the circled area of Figure 3. Surrounding this new kind of AI processing are all of the traditional hearing aid technologies patients are already used to — methods like compression and gain management, feedback management, spectral noise reduction, scene analysis and classification, and directionality.

What Hardware is Required to Enable Advanced Processing of AI?

The Whisper Hearing System carries the largest and most sophisticated deep neural network of any hearing aid to date. This is made possible by the earpieces and the Whisper Brain (see Figure 4). The earpieces perform the traditional hearing aid processing methods, while the Whisper Brain contains the Sound Separation Engine. In concert with the Whisper Brain, the earpieces use AI to find acoustic patterns and identify specific areas of importance in the sound scene. The Brain also gathers new data that improves the AI to drive future software upgrades.

Figure 4: The Whisper Hearing System, which includes the Whisper Brain (left), and BTE RIC-style earpieces (right).

In the Whisper Hearing System, patients benefit from the convenience of a BTE RIC form factor via the earpieces, which work well for simple, everyday situations. For more complicated sound environments, patients can leverage the Sound Separation Engine by having the Whisper Brain nearby. Unlike a remote microphone, the Whisper Brain can be hidden away in a pocket or nearby bag, or rest on a table while someone is walking around their home.

The Whisper Brain is able to go deeper to target the patterns inside of the sound signal because of the amount of processing capability it has. Like any smart device, the more computational capability that device has (as measured by operations per second), the more it can do. Modern hearing aids using traditional processors are capable of approximately 300 million to 1.2 billion operations per second. Newer hearing aids that build AI onto a RIC device can double that, reaching up to approximately 2.4 billion operations per second (Figure 5). However, this very modest increase in processing limits what is possible. In contrast, the Whisper Brain, and the processors inside of it, have a processing capability of 300 billions of operations per second, which is hundreds of times more than existing BTE RIC devices (Figure 6). This added processing power is what enables the

Whisper Hearing System to shift from working only with frequency information to working with full sound patterns.

Ultimately, when artificial intelligence is combined with significantly increased processing capability, it means better outcomes for hearing aid wearers. Consider what it is like to enter a challenging acoustic environment with many different sound sources [click to play]. Now imagine the same situation but with elevated speech sounds, a more comfortable auditory background, and differentiated acoustic signals [click to play]. The Whisper Hearing System is the first hearing aid that makes this level of processing capability available to patients.

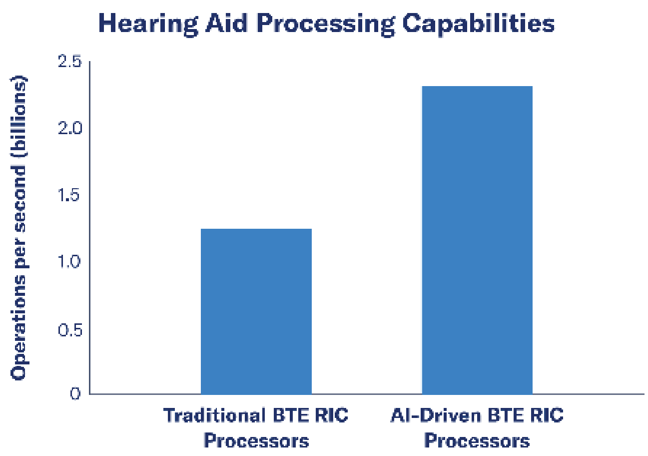

Figure 5: Processing capability of traditional BTE RIC hearing aids and AI-driven BTE RIC devices.

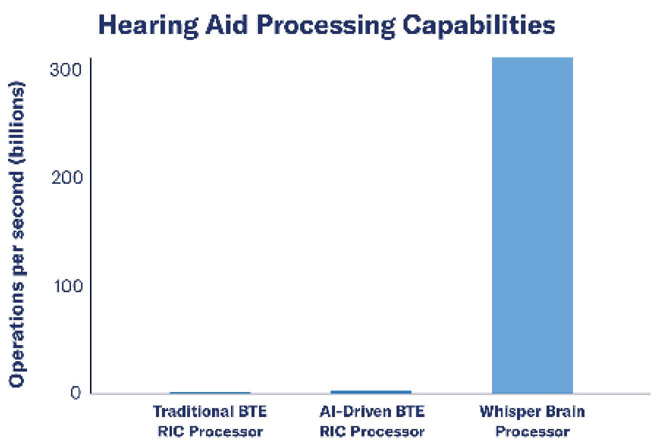

Figure 6: Processing capabilities of the Whisper Brain compared to traditional BTE RICs and AI-driven BTE RIC devices.

The Hearing Aid That Gets Better Over Time

In the earlier example of the mobile app that recognized dog breeds, the app was able to learn about new dog breeds when it received new examples of what these breeds looked like. Similarly, Whisper's AI algorithms identify sound patterns, and the Whisper Brain brings this new type of sound processing to hearing aids via the Sound Separation Engine. This is the first core benefit to patients. The second is that the Whisper Hearing System gets better over time by starting with patient and professional feedback and then incorporating new acoustic data into the Sound Separation Engine (see Figure 7).

Figure 7: The learning cycle as applied to dog breed identification and hearing aids. New data about patient and professional experience, sound sources, and environments helps the Whisper Hearing System learn and get better over time.

The Whisper team is continuously improving the underlying model for the Sound Separation Engine. The starting point of any improvement is the experience of the hearing aid wearer and their hearing care professional as they highlight the environments or situations that are most important to them. From there, new data about sound sources and environments is gathered to improve how the Sound Separation Engine finds patterns in sound. We view data as a critical piece of improving hearing and have partnered with leaders in sound processing AI, like Mitsubishi Electric Research Lab, to publish research that benchmarks new techniques in AI. We have also released real-world acoustic data to help researchers drive further innovation.

All of this new data is compiled into a global software upgrade so every patient and professional can benefit from the new learnings. These software upgrades give patients access to the latest in hearing health technology without having to buy a new device, whether it’s improved sound processing based on new data and hearing care professionals’ feedback or new features and capabilities.

Conclusion

Artificial intelligence is a new technology that helps us deliver better hearing outcomes for patients, and the Whisper Hearing System carries the largest and most sophisticated deep neural network of any hearing aid to date. This is enabled by the processing capability of the Whisper Brain. Using the Brain’s Sound Separation Engine, the Whisper Hearing System can deliver better care by targeting patterns within a sound signal instead of just analyzing coarse frequencies. All of this leads to patients getting better hearing technology over time via software upgrades without having to buy a new hearing aid.

Andrew Song is the co-founder and president of Whisper. Andrew was formerly head of product for Facebook Messenger Core, which is used by over 1B people worldwide. A mathematics & computer science graduate from the University of Waterloo, Andrew is an expert in the advanced applications of artificial intelligence. He is a member of Sequoia Capital's Scout program, which aids in the discovery and development of other high-potential companies.

Andreas Thelander Bertelsen is an industry leading architect within signal processing and machine learning for hearing devices. He has more than a decade of experience in researching, developing and designing noise suppression systems for hearing devices. Andreas led the design of the noise suppression system in the successful Oticon Opn S and Oticon More series of hearing devices. His work has led to the authorship of numerous key patents in the field.

Joseph Antognini is a machine learning engineer at Whisper, where he works on machine learning and digital signal processing. Prior to Whisper, he was a Google Brain Resident and conducted research on machine learning for audio and the foundations of deep learning. Joseph has also developed neural networks for EEG and EKG data at Persyst. He graduated from Caltech with a B.S. in Astrophysics and with a Ph.D. from The Ohio State University.

References

Ho, J. (2015, August 24). Qualcomm Details Hexagon 680 DSP in Snapdragon 820: Accelerated Imaging. AnandTech. https://www.anandtech.com/show/9552/qualcomm-details-hexagon-680-dsp-in-snapdragon-820-accelerated-imaging

Oticon More Technology. (2021). Oticon. https://www.oticon.com/professionals/brainhearing-technology/more-technology

Welle, J. N., & Bach, R. (2016). The Velox Platform. Oticon. https://www.oticon.com/-/media/oticon-us/main/download-center/white-papers/15555-9940-veloxwhitepaper.pdf