

Review Article

Understanding Speech Processing: Mechanisms, Theories, and Applications

Abha Bhadauriya

Student, Bhabha Engineering Research Institute Research institution, Bhopal, Madhya Pradesh, India

I N F O

E-mail Id:

abhabhadauriya@gmail.com

Orcid Id:

<https://orcid.org/0009-0006-5738-967X>

How to cite this article:

Bhadauriya A. Understanding Speech Processing: Mechanisms, Theories, and Applications. *J Adv Res Comput Linguist : J Comput Sci Lang.* 2025; 1(1): 25-31.

Date of Submission: 2025-01-28

Date of Acceptance: 2025-03-08

A B S T R A C T

Speech processing is a complex cognitive function that involves the interpretation, recognition, and generation of speech signals. This review explores the key mechanisms behind speech processing, focusing on the neurological, psychological, and computational aspects of speech recognition and production. We discuss the major theories that aim to explain how humans process spoken language, including models of auditory perception, phonological representation, and neural processing. Additionally, the article examines advancements in speech processing technologies and their applications in fields such as speech recognition, synthesis, and assistive technologies. Finally, we look at the challenges faced in this field, such as accent variation, noise interference, and language ambiguity, and highlight the future directions for research and development.

Keywords: Speech Processing, Speech Recognition, Neural Mechanisms

Introduction

Speech processing refers to the cognitive and computational mechanisms that allow both humans and machines to perceive, understand, and generate spoken language. It encompasses various aspects, including auditory perception, linguistic interpretation, and speech production. This field is inherently interdisciplinary, integrating elements of neuroscience, linguistics, psychology, cognitive science, and computer science. The study of speech processing is pivotal not only in understanding how humans communicate, but also in developing technologies that facilitate communication, such as speech recognition systems, text-to-speech applications, and assistive technologies for individuals with speech or hearing impairments.

The mechanisms involved in speech processing are remarkably complex, involving both bottom-up (sensory-driven) and top-down (context-driven) processes. Humans must quickly and accurately decode auditory signals, often in noisy or ambiguous contexts, and match these signals

with phonological, syntactic, and semantic representations in the brain. In turn, the brain coordinates the motor actions necessary for speech production, utilizing neural networks distributed across several brain regions such as Broca's area, responsible for speech production, and Wernicke's area, which plays a crucial role in speech comprehension.¹

This review explores the current understanding of speech processing, with a focus on the cognitive and neural mechanisms involved in human speech perception and production. We will also examine the major theoretical frameworks that have been developed to explain these processes, including neural models and computational models of speech recognition. Additionally, we discuss the wide-ranging technological applications that rely on speech processing, such as automatic speech recognition (ASR), speech synthesis, voice-activated assistants, and assistive communication devices for individuals with disabilities.

Speech processing plays a crucial role in human activities, not only enabling communication in daily life but also

contributing to advancements in language learning and social interaction. In recent years, the field has seen significant technological progress, with innovations in natural language processing and deep learning revolutionizing the accuracy and efficiency of speech-related systems. Furthermore, emerging technologies in augmented reality (AR) and virtual reality (VR) are increasingly relying on speech processing for more immersive and interactive experiences. As speech processing technologies continue to evolve, it becomes imperative to understand both the mechanisms that underpin human speech and the technological applications that bring these mechanisms to life.²

Mechanisms of Speech Processing

Auditory Perception and Acoustic Analysis

The auditory system plays a crucial role in the initial stage of speech processing, where sound waves are received, converted into neural signals, and processed by the brain. This begins with the outer ear capturing sound waves, which travel through the ear canal and vibrate the eardrum. These vibrations are then transferred to the middle ear bones (ossicles) and amplified before reaching the cochlea in the inner ear. The cochlea, through its hair cells, transduces sound vibrations into electrical impulses, which are sent to the brain via the auditory nerve.

Once these auditory signals are relayed to the brain, the primary auditory cortex, located in the temporal lobe, is responsible for analyzing and interpreting the raw auditory input. The auditory cortex breaks down the sound signals into several acoustic features, including frequency (pitch), amplitude (loudness), and duration. These basic features help the brain detect speech sounds and other acoustic properties of the sound signal.

Next, speech undergoes phonetic analysis, a critical step where the brain further breaks down the incoming signal into phonetic units—phonemes, which are the smallest units of sound in language. The brain also processes prosodic features, such as intonation, rhythm, and stress patterns, which are essential for understanding the emotional tone, emphasis, and rhythm of spoken language.³

Through these processes, the brain uses the auditory cues to make sense of the incoming speech signal and match it with stored linguistic representations. Top-down processing plays a key role here, as the brain uses contextual information (e.g., prior knowledge, expectations) to help interpret sounds and resolve ambiguities in speech perception, such as when background noise or unclear articulation may distort the speech signal.

Phonological Processing

Phonological processing refers to the way the brain maps the acoustic speech signal onto its linguistic structure,

which is essential for both speech comprehension and production. This process involves two key stages: phoneme segmentation and phonological encoding.

- **Phoneme segmentation:** In this stage, the brain identifies and isolates individual phonemes in the speech signal. Phonemes are the fundamental units of sound that distinguish one word from another. The challenge here lies in the variability of speech, such as differences in pronunciation due to accents, coarticulation (where the articulation of one phoneme influences the next), or speech rate. These variations require the brain to be flexible and adaptable in segmenting speech into its constituent sounds. For instance, the pronunciation of the phoneme /t/ might vary in different linguistic environments, making segmentation a dynamic process.
- **Phonological encoding:** Once the phonemes are identified, the brain maps them to their lexical representations in the mental lexicon, linking each sound to a specific meaning. This process is essential for recognizing words and linking them to their meaning in context. Phonological encoding involves not only recognizing individual words but also understanding how words combine to form sentences and convey more complex meanings.

The integration of phonology and semantics (the study of meaning) is crucial at this stage. The brain must combine the sounds of words (phonology) with their meanings (semantics) to derive the overall meaning of the sentence. For example, the phonological representation of the word “bank” can refer to a financial institution or the side of a river, depending on the context in which it is used. Thus, contextual knowledge aids in determining the correct interpretation.

However, phonological processing faces challenges such as accent variability, where the same word might be pronounced differently by speakers of different dialects or regions. Additionally, coarticulation—the phenomenon where adjacent speech sounds influence each other—can make it difficult to isolate individual phonemes, requiring the brain to be highly adaptable in recognizing speech.^[4]

These mechanisms of speech processing illustrate the intricate and dynamic nature of how humans perceive, decode, and produce spoken language. From initial auditory perception to phonological processing, the brain must integrate a complex set of sensory, cognitive, and contextual cues to understand and produce speech. This flexibility allows individuals to adapt to various linguistic environments, accents, and speaking styles.

Neural Mechanisms of Speech Processing

Speech processing is not only a complex cognitive task but also one that relies heavily on the coordinated activity

of neural networks distributed across various regions of the brain. Several key brain areas play distinct roles in perceiving, understanding, and producing speech.

- **Broca's Area:** Located in the left frontal lobe, Broca's area is critically involved in speech production and the motor planning for articulation. This region is linked to the ability to construct grammatically correct sentences and produce fluent speech. Damage to Broca's area can lead to Broca's aphasia, a condition characterized by difficulties in producing speech, though comprehension remains relatively intact.
- **Wernicke's Area:** Situated in the left temporal lobe, Wernicke's area is primarily responsible for speech comprehension. It enables the brain to interpret the meaning of speech sounds and recognize words. Damage to Wernicke's area can result in Wernicke's aphasia, which leads to fluent but often nonsensical speech and impaired understanding of spoken language.
- **Auditory Cortex:** Located in the temporal lobe, the auditory cortex processes auditory signals, including speech sounds. It plays a crucial role in initial speech perception, where sound is transformed into neural representations that can be recognized as language. The auditory cortex analyzes acoustic features such as pitch, tone, and rhythm that form the foundation for phonetic analysis.
- **Motor Cortex:** The motor cortex, located in the frontal lobe, is involved in controlling the physical production of speech. It coordinates the fine motor movements required for articulating speech sounds, ensuring that the muscles involved in speech production (such as the lips, tongue, and vocal cords) function smoothly and precisely.⁵
- **Frontal Lobe:** Beyond its role in motor control, the frontal lobe is important for higher-level cognitive functions such as syntactic processing, working memory, and executive control, all of which are necessary for fluent and meaningful speech production and comprehension.
- **Temporal Lobe:** The temporal lobe, where both Broca's and Wernicke's areas are located, is essential for processing auditory information and understanding language. It also plays a role in semantic processing, helping to integrate speech sounds with their meanings.

The interaction of these brain regions occurs through a complex network, with information constantly being exchanged across regions. This is facilitated by neural feedback loops, allowing for a dynamic exchange between bottom-up (sensory-driven) and top-down (contextually-driven) processes. Bottom-up processing refers to the sensory-driven decoding of auditory signals, such as recognizing phonemes and words. Top-down processing,

on the other hand, relies on contextual knowledge—such as expectations based on prior linguistic experience or the surrounding context of conversation—to help interpret and make sense of incomplete or ambiguous speech signals. This dynamic interaction between lower-level auditory processing and higher-level cognitive processing enables individuals to understand speech, even in challenging environments, such as noisy settings or when speech is distorted.

Together, these neural mechanisms ensure that speech processing is efficient, adaptive, and robust, supporting a wide range of speech-related tasks, from basic sound recognition to complex sentence comprehension and production.⁶

Theoretical Models of Speech Processing

Theoretical models of speech processing aim to explain how the brain efficiently decodes and processes the complex auditory signals that make up speech. Two of the most influential models in this area are the Trace model and the Cohort model, both of which propose distinct but complementary mechanisms for understanding speech recognition.

The Trace Model

The Trace model (McClelland & Elman, 1986) posits that speech recognition is a dynamic and interactive process, involving the parallel processing of multiple levels of linguistic information. According to this model, speech recognition occurs through the activation of phonetic, lexical, and semantic representations in the brain, with information flowing between these levels in a feedback loop.

- **Parallel Processing:** The Trace model suggests that, rather than processing speech in a strictly linear or serial fashion, the brain processes speech at multiple levels simultaneously. For example, the brain doesn't wait until a phoneme is fully processed before considering possible word candidates in the lexicon. Instead, feedback from higher levels of linguistic knowledge (such as knowledge of words and meanings) influences the processing of lower-level auditory signals (such as phonetic features).
- **Feedback Mechanism:** One of the key features of the Trace model is the importance of feedback loops between different levels of processing. This allows the brain to refine and adjust its interpretation of the incoming speech signal as more information becomes available. The feedback loop helps resolve ambiguities and predict upcoming speech sounds, improving the speed and accuracy of speech recognition.

This model is particularly useful in explaining how the brain can recognize speech in real-time, efficiently integrating

acoustic input with linguistic knowledge to form accurate representations of spoken words.

The Cohort Model

The Cohort model (Marslen-Wilson, 1987) takes a slightly different approach by emphasizing the sequential nature of speech recognition. It suggests that, as speech is heard, the brain activates a set of possible word candidates (known as the cohort) based on the initial speech sounds. This cohort is gradually narrowed down as more information becomes available, until the correct word is identified.

- **Early Selection:** According to the Cohort model, speech recognition begins with an initial set of possible words, based on the first few phonemes heard. As the listener processes more speech sounds, the set of candidates is refined, and incorrect candidates are eliminated. This early selection process allows for quick identification of words.
- **Contextual Influence:** The Cohort model also highlights the role of context in the selection process. The cohort of possible word candidates is evaluated not only on phonological features (such as the initial sounds of words) but also on contextual information. For example, if the speaker is discussing a particular topic, the likelihood of certain words being activated will be influenced by the surrounding conversation and expectations.

This model is effective at explaining how listeners are able to recognize speech so quickly and accurately, even in cases where the initial sounds are incomplete or ambiguous. By narrowing down the possibilities early on, the brain can efficiently identify words and understand their meanings within context.⁷

These two theoretical models—Trace and Cohort—offer complementary insights into the complex and dynamic processes involved in speech recognition. While the Trace model emphasizes feedback loops and parallel processing at multiple levels of representation, the Cohort model highlights the sequential nature of speech recognition and the importance of early word selection based on context. Both models contribute to our understanding of how the brain efficiently processes speech in real time.

The Dual-Stream Model

The Dual-Stream Model of speech processing, proposed by Hickok and Poeppel (2007), provides a framework that differentiates two primary neural pathways involved in speech: the ventral stream and the dorsal stream. This model reflects the complexity of speech processing, suggesting that speech comprehension and production are managed by separate yet interconnected neural systems.

Ventral Stream: Speech Perception

The ventral stream is responsible for speech perception, including the mapping of auditory signals to meaning. It involves higher-order brain regions that process speech sounds and associate them with semantic representations. The ventral stream allows the brain to map sound sequences onto linguistic meaning, enabling us to understand the content of spoken language. This pathway is involved in tasks such as recognizing words and comprehending sentences, utilizing areas such as the temporal lobe and Wernicke's area. The ventral stream is also integral to lexical access, where the brain retrieves stored meanings of words and processes speech in real-time, even when encountering noisy or distorted input.

Dorsal Stream: Speech Production

The dorsal stream, on the other hand, is primarily concerned with speech production. It connects auditory areas with motor areas of the brain, which are responsible for generating the physical movements necessary for articulation. The dorsal stream's role is to translate auditory representations of speech into motor commands that control the muscles involved in speech, including the tongue, lips, and vocal cords. This pathway is associated with areas such as the motor cortex and Broca's area, which coordinate the fine motor movements required for fluent speech production.

The distinction between these two streams highlights the tight interconnection between speech comprehension and production. The model suggests that information does not flow in only one direction but that there is continuous feedback between the ventral and dorsal streams. This dynamic interaction allows for real-time adjustments during both speech comprehension and production, improving the efficiency and accuracy of speech processing. For example, understanding speech sounds in a noisy environment (ventral stream) may help predict the motor actions required to articulate speech (dorsal stream), and vice versa.

This dual-stream model has provided valuable insights into how the brain processes the complexity of speech, explaining the dual roles of perception and production in speech processing and how these processes are intertwined in the brain.⁸

Applications of Speech Processing

Speech processing technologies have a wide range of practical applications in various fields, including human-computer interaction, healthcare, and education. Below are some of the most significant applications:

Automatic Speech Recognition (ASR)

Automatic Speech Recognition (ASR) technologies are designed to convert spoken language into written text or commands, and they are now integral to many everyday devices and applications. ASR systems analyze the acoustic properties of speech to identify words and transcribe them into written form.

- **Machine Learning and Deep Learning:** Modern ASR systems rely heavily on machine learning (ML) and deep learning techniques, particularly neural networks, to improve the accuracy and robustness of speech recognition. These methods enable systems to learn patterns from vast amounts of data, making them more effective at recognizing speech in diverse environments and from a variety of speakers.
- **Applications:** ASR technologies are used in virtual assistants such as Siri, Google Assistant, and Alexa, as well as transcription services, voice-activated devices, and in-car systems. They are also used for voice-to-text applications, enabling individuals to dictate messages or commands instead of typing them.
- **Challenges:** Despite the advancements, ASR faces several challenges, including:
- **Accent Variation:** Different accents and dialects can impact the system's ability to recognize words accurately.
- **Background Noise:** Noisy environments can interfere with speech recognition, particularly when multiple speakers or environmental sounds overlap.
- **Contextual Ambiguity:** ASR systems must resolve ambiguities in speech, such as homophones (e.g., "their" vs. "there") or words that can have different meanings based on context.

Speech Synthesis

Speech synthesis refers to the technology used to generate artificial human speech from written text. This technology is often used in text-to-speech (TTS) systems, which are integral to applications that need to "speak" to users, such as virtual assistants, navigation systems, and accessibility tools.

- **Neural TTS:** Recent advancements in neural network-based TTS systems have allowed for highly realistic synthetic voices. These systems generate speech that mimics natural human intonation, rhythm, and prosody, making the synthetic speech sound more lifelike and engaging. These systems can adapt to different speaking styles, including variations in emotion, speed, and pitch, which are important for enhancing user experience.

Applications:

- **Assistive Technology:** TTS is widely used in tools for visually impaired individuals, allowing them to access written content in a spoken format.
- **Voice Assistants:** TTS systems provide a natural-sounding interface in applications like Alexa or Google Home, allowing users to interact with devices through speech.
- **Audiobooks and News Readings:** TTS has been integrated into apps that convert books, articles, or news into spoken form, offering an alternative for users who prefer auditory content.⁹

Assistive Technologies

Speech processing plays a critical role in assistive technologies for individuals with speech, hearing, or cognitive impairments. These technologies leverage speech recognition and synthesis to improve communication and interaction with the environment.

- **Speech-to-Text Systems:** These systems convert spoken language into written text, helping individuals with hearing impairments or those who are unable to speak. They are used in communication devices for deaf or hard-of-hearing individuals, facilitating real-time captioning or transcription of conversations.
- **Augmentative and Alternative Communication (AAC) Devices:** AAC devices help individuals with severe speech or language impairments communicate. These devices can use speech synthesis to generate spoken words based on user input, allowing for non-verbal communication through text or symbols.
- **Hearing Aids:** Advanced hearing aids use speech processing algorithms to enhance speech signals while suppressing background noise, making it easier for individuals with hearing impairments to understand speech in noisy environments.
- **Potential for Further Advancements:** Ongoing research in speech recognition, synthesis, and adaptive systems holds the potential to improve accessibility for people with disabilities. For example, more accurate recognition of non-standard speech patterns (such as dysarthria) could make assistive devices more inclusive.

The development and application of speech processing technologies have led to transformative changes in many fields, from virtual assistants to assistive devices. The integration of neural networks, machine learning, and deep learning has dramatically enhanced the accuracy and naturalness of speech recognition and synthesis. As research continues, we can expect further innovations that

will make speech processing more effective and accessible, improving communication for everyone, including those with disabilities.

Challenges and Future Directions

Despite the significant advancements in speech processing technologies, several challenges persist that impact their effectiveness and usability in real-world applications. These challenges highlight the complexities of processing speech in diverse contexts and environments. Addressing these issues will be crucial for further progress in the field.

Handling Noise

One of the most significant challenges for Automatic Speech Recognition (ASR) and other speech processing systems is background noise. In real-world environments, speech is often accompanied by various types of noise—such as traffic, crowds, or other ambient sounds—which can distort or obscure the speech signal.

- **Noise Robustness:** Developing robust noise-robust algorithms that can distinguish between speech and background noise is crucial. Advanced signal processing techniques, including noise filtering, speech enhancement, and source separation algorithms, are being researched to improve speech recognition performance in noisy environments.
- **Multi-microphone Systems:** The use of multiple microphones, coupled with beamforming techniques, can help improve speech recognition by isolating speech from noise.

Accent and Dialect Variation

Accents and dialects present another significant challenge for ASR and speech synthesis systems. Variations in pronunciation, intonation, and rhythm can lead to misrecognition or misinterpretation of speech by ASR systems. Similarly, speech synthesis systems may produce unnatural-sounding voices for users with accents or dialects that differ from the system's training data.

- **Accent Adaptation:** Research is focused on creating models that are accent-agnostic or that can adapt to different accents dynamically. Using large and diverse datasets to train speech recognition systems can improve the ability to understand a broader range of accents.
- **Personalized Speech Synthesis:** Personalized TTS systems could allow for more natural voice generation, adapting to specific accent and speech patterns of individual users, making synthetic voices sound more natural and representative of regional variations.

Multilingual Processing

The ability to process multiple languages or seamlessly switch between languages is another challenge in speech processing. Many existing ASR and TTS systems are optimized for single languages, and handling multilingual scenarios with high accuracy is a complex task.

- **Multilingual ASR and TTS:** Future research is likely to focus on developing systems that can handle code-switching (the switching between languages within a single sentence) and language switching without loss of performance. Such systems will need to identify language boundaries and adapt to various phonetic and syntactic features across languages.
- **Universal Speech Models:** The development of universal speech models, which can be trained on data from multiple languages simultaneously, may improve multilingual processing and enable systems to recognize and synthesize speech in a wide variety of languages.¹⁰

Real-Time Processing

Real-time speech processing is crucial for applications such as live transcription, virtual assistants, and real-time translation. However, achieving real-time performance without compromising accuracy remains a significant challenge, particularly as the complexity of tasks increases.

- **Latency Reduction:** Real-time speech systems need to process incoming speech data with minimal delay. Researchers are focusing on reducing latency in ASR and TTS systems to allow for more fluid and natural interaction.
- **Real-Time Translation:** Speech processing for real-time translation involves additional layers of complexity, as it requires not only accurate speech recognition but also the instantaneous translation of meaning and generation of speech in another language. This requires sophisticated systems that can process both the spoken input and generate an appropriate output almost instantaneously.

Future Directions

As the field of speech processing continues to evolve, the following key areas are expected to see significant developments:

Integration of Multimodal Systems

Future speech processing systems will likely integrate multimodal data to improve overall performance. These systems will combine speech data with visual, contextual,

and environmental inputs to improve accuracy and adaptability. For example, integrating lip reading or gesture recognition could help resolve ambiguities in speech and make systems more effective in noisy or unclear conditions.

Personalized and Context-Aware Speech Systems

Advancements in machine learning and AI could enable personalized speech processing systems that adapt to individual users' speaking styles, preferences, and needs. Such systems could adjust for accents, speech impediments, or background noise specific to the user's environment. Furthermore, speech systems that are more context-aware could adapt their responses based on the situation, allowing for more natural, intuitive human-computer interaction.

Improving Speech Recognition Accuracy for Diverse Populations

One critical area of future research is to improve the accuracy of speech recognition systems for diverse populations, including individuals with speech impairments, those who speak non-standard dialects, and people with non-native accents. This research will involve gathering more diverse datasets and developing specialized models that account for these variations in speech patterns.

Enhanced Naturalness in Speech Synthesis

While TTS technology has made tremendous strides, making synthetic speech sound even more natural remains a key goal. Future systems may focus on generating emotionally expressive speech that mirrors human intonation and prosody, and can reflect varying emotions or intentions, improving interaction with virtual assistants or conversational agents.

Ethical Considerations and Privacy

As speech processing technologies become more widespread, there will be growing concerns about privacy and the ethical use of these systems. Speech data can be highly sensitive, and concerns about data security and user consent will become increasingly important. Research on privacy-preserving methods in speech processing will be essential, particularly for applications in healthcare or sensitive environments.

Conclusion

The field of speech processing has witnessed remarkable advancements, contributing significantly to human-computer interaction, accessibility, and the development of technologies that enhance communication. Despite the progress, challenges related to noise, accents, multilingualism, and real-time processing persist. Addressing these challenges requires continued research and innovation, particularly leveraging artificial intelligence, machine learning, and neural networks to create more

efficient, adaptable, and natural speech processing systems. As speech processing systems evolve, they hold great promise for advancing communication technologies and improving accessibility for people with diverse needs, making interaction with machines more intuitive and inclusive.

References

1. McClelland JL, Elman JL. The TRACE model of speech perception. *Cogn Psychol.* 1986;18(1):1-86. doi: 10.1016/0010-0285(86)90015-0.
2. Marslen-Wilson WD. Functional parallelism in spoken word recognition. *Cogn Psychol.* 1987;19(3):317-341. doi: 10.1016/0010-0285(87)90016-4.
3. Hickok G, Poeppel D. The cortical organization of speech processing. *Nat Rev Neurosci.* 2007;8(5):393-402. doi: 10.1038/nrn2113.
4. Papadopoulou F, Sato Y, Sato H, et al. Noise-robust automatic speech recognition for real-world applications. *IEEE Trans Audio Speech Lang Process.* 2020;28:870-882. doi: 10.1109/TASLP.2020.2977384.
5. Chen Z, Song X, Zhang J, et al. A comprehensive review of the speech synthesis systems: Approaches, techniques, and applications. *IEEE Trans Neural Netw Learn Syst.* 2021;32(10):4470-4487. doi: 10.1109/TNNLS.2020.3044471.
6. Binns R, McNeill D, Van Lieshout P, et al. Speech-to-text systems for people with speech disabilities: an overview of recent advances. *IEEE Trans Biomed Eng.* 2019;66(10):2798-2808. doi: 10.1109/TBME.2019.2899678.
7. Li H, Chen X, Xiao L, et al. Enhancing multilingual speech processing: A review of challenges and future directions. *IEEE Access.* 2022;10:19498-19515. doi: 10.1109/ACCESS.2022.3157674.
8. Wang X, Lee M, Zhao Z, et al. Real-time automatic speech recognition for human-robot interaction: Challenges and solutions. *IEEE Trans Robot.* 2020;36(1):95-108. doi: 10.1109/TRO.2019.2945092.
9. Bonin P, Boullier A. Understanding speech processing: From recognition to production. *Psychol Rev.* 2021;128(1):28-49. doi: 10.1037/rev0000167.
10. Chien S, Liu H, Hu W. Analyzing speech recognition and synthesis systems for diverse applications. *J Acoust Soc Am.* 2020;147(3):1632-1645. doi: 10.1121/10.0000335.