

이교구 교수님: 말하고 노래하는 인공지능(AI Voice)

The seminar titled "AI Voice: Speaking and Singing Artificial Intelligence" given by Professor Lee Gyogu introduces the natural principles of voice generation, utilizing artificial intelligence to create synthesized voices, and explores various applications in this field. Language intelligence, which has significantly contributed to human development and civilization, is discussed as a crucial means of communication that requires collaboration. The process of producing speech involves the air expelled from the lungs passing through the vocal folds, generating vocal source signals that are further shaped by articulatory organs to produce speech sounds. The complexity and abstractness of information conveyed through speech enable survival and the blossoming of vibrant civilizations, in conjunction with the invention of written language. The seminar traces the history of speech synthesis attempts, from early experiments with resonating tubes and speaking machines to mechanical, electrical, and computer-based methods. The development of computer-based speech synthesis led to the advent of text-to-speech (TTS) systems, which convert written text into human-like speech. Different techniques such as concatenative synthesis and statistical models have been employed, and the emergence of deep learning techniques enabled accurate prediction of complex speech patterns by leveraging vast amounts of speech data. The applications of voice synthesis range from TTS systems to automated response systems, guide broadcasting, AI speakers, and audiobooks. Voice conversion, which allows transforming one voice into another while controlling various characteristics such as timbre, pitch, stress, and semantic content, is also discussed. Singing, a specialized form of voice expression aimed at conveying emotions rather than information, can also be synthesized using trained models. The seminar envisions a future where voice-based communication with human-like agents, such as digital companions and multilingual real-time translation systems, becomes essential, inspired by the movie "Her" and the concept of voicebots.