國立中山大學 工學院 聯合專題競賽與展示

資訊工程學系
Department of Computer Science and Engineering

第 17 組：朱劭璿、陳居廷　　指導老師：陳嘉平教授

TSM-Net: 以對抗式時序壓縮自編碼器為基礎的音訊變速演算法
TSM-Net: Temporal Compressing Autoencoder with Adversarial Losses for Time-Scale Modification on Audio Signals

Introduction

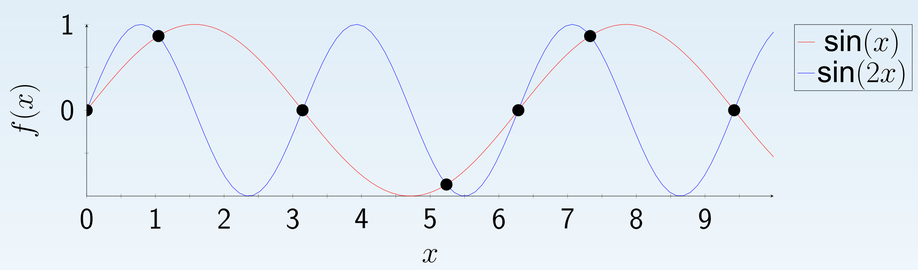
With the advance of technologies and digitalization, we can store and reproduce multimedia content nowadays. An ubiquitous application regarding audio signals called time-scaled modification (TSM) is used in our daily life. It's also known as playback speed control in the video streaming platforms such as YouTube. With the power of artificial intelligence (AI) and modern computation hardware, however, we haven't discovered any method using AI to refine TSM algorithm and leverage the quality of the synthetic audio to the next level.

We proposed a novel TSM approach. While traditional methods use framing technique and spectral approaches use short-time Fourier transform to get high-level units. TSM-Net, our neural-network model encodes the raw audio into a high-level latent representation called Neuralgram. Since the resulting Neuralgram is a two-dimensional image with real values, we apply some existing image resizing techniques on the Neuralgram and decode it using our neural decoder to obtain the time-scaled audio.

Related Works

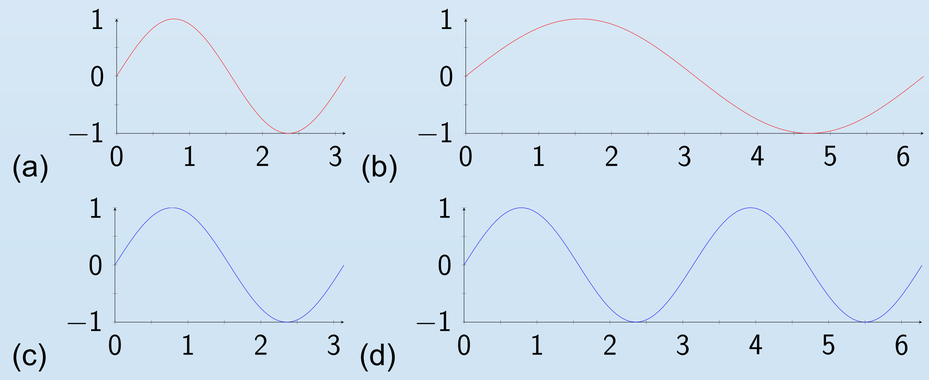
Modeling audio is not a trivial task for neural networks. Models that directly generate raw audio waveform are known as vocoder which can be conditioned on some high-level abstract features such as linguistic features or spectrograms. In applications like text-to-speech (TTS) pipeline, the network often predicts the speech spectrogram of given texts, then uses a vocoder to get the raw audio. Modern neural-enabled vocoders bring the synthetic quality to a new level.

Decreasing the sampling rate to simplify the dimensionality is another option. However, the Nyquist-Shannon sampling theorem suggests that the low sampling rate would lead to serious aliasing. For example, suppose the signals are sampled non-uniformly for a better illustration, two signals with different frequency components are the aliases for each other in the discrete domain, represented as black dots.



Methodology

Nulla malesuada porttitor diam. Donec felis erat, congue non, volutpat at, tincidunt tristique, libero. Vivamus viverra fermentum felis. Donec nonummy pellentesque ante. Phasellus adipiscing semper elit. Proin fermentum massa ac quam. Sed diam turpis, molestie vitae, placerat a, molestie nec, leo. Maecenas lacinia. Nam ipsum ligula, eleifend at, accumsan nec, suscipit a, ipsum. Morbi blandit ligula feugiat magna. Nunc eleifend consequat lorem. Sed lacinia nulla vitae enim. Pellentesque tincidunt purus vel magna. Integer non enim. Praesent euismod nunc eu purus. Donec bibendum quam in tellus. Nullam cursus pulvinar lectus. Donec et mi. Nam vulputate metus eu enim. Vestibulum pellentesque felis eu massa.



Experiment

Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

Conclusion

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetuer.