Professor Notes

* The last time we applied for "mid-level research" was for an excellent research group. It seems that we are asking for a research topic that is a little more scalable and larger. However, our research topic was too specific. However, the "basic research" supported this time, I think that specific topics such as voice reinforcement are also competitive.
* Due to the limitation of 5 pages, it is necessary to write clearly what the research goal is and how to achieve the goal.
* In this paper, Voice data and noise data are introduced. If you adjust the intensity of these two and fuse them, you can create noisy speech data*. [Data Augmentation for DNN-based Speech Enhancement]*

note

Designing a new neural network

1) Neural network structure and 2) loss function must be newly defined.

This slide is about the architecture of a neural network for speech reinforcement.

The neural network structure of most speech enhancement papers is an Encoder-Decoder structure. A representative model of this structure is U-Net: Encoder-Decoder with Intermediate connection. The Encoder-Decoder structure can be trained on a given input data structure without labels.

Based on this structure, The intermediate result, laten variable, is given a transform (unique neural network phrase) to suit a particular application (eg, speech enhancement). We should pay attention to what unique structure the previous paper took in the middle of the encoder-decoder.

I am interested in

1) Instead of U-Net, which has adopted the Encoder-Decoder structure in many existing studies, I think it would make sense to replace it with Diffusion. (Of course, you will have to solve the problem of processing speed)

2) It would be nice to propose a new structure that transforms latent variables between encoders and decoders.

Another aspect of new neural network research not covered in the slides above is the loss function:

- I think we need to design/develop a metric loss specific to voice reinforcement.

How to speed up diffusion models by applying wavelets. -> [LINK](https://arxiv.org/abs/2211.16152)

The patent below describes the STFT disadvantages. To solve this problem, I would like to take the research of converting 1d voice+noise signal -> 2D time-frequncy as a topic.

By utilizing the Continuous Wavelet Transform. It seems that the patent below is a way to utilize CWT.

[LINK1](https://patentscope.wipo.int/search/ko/detail.jsf?docId=US354532117&_fid=KR308157823), [LINK2](https://patentscope.wipo.int/search/de/detail.jsf?docId=WO2020145509&_fid=KR300789094)

Final Note.md

Research subject:

Frequency extraction method that can increase time resolution and frequency resolution at the same time

Problems with STFT:

When frequency is measured using the short-time Fourier transform (STFT), there is a limit to simultaneously increasing time precision and frequency precision due to the Fourier uncertainty principle.

The Fourier Uncertainty Principle states that if a sound for a short time is converted into a frequency component, it has a frequency component with a low resolution, and if a sound for a long time is used to measure an accurate frequency, the time resolution at the time of occurrence of the measured frequency is lowered. will be.

Speech recognition and speaker discrimination In these applications, the use of STFT is sufficient. However, STFT or DWT (Discrete Watlet Transform) is not sufficient because voice signals must be locally processed in detail in voice enhancement.

If the former is at the level of object recognition, the latter is at the semantic segmentation task level.

Resolution:

I think you can use CWT (Continous Wavlet Transform) for some reason.

Related Skills:

https://youtu.be/8NeYnv0GP\_A?t=108 How did you do it? When extracting frequencies from sound, you can increase the resolution.

After a few seconds, it shows something like an existing/improvement spectrum.

Patents registered by this company:

[LINK1](https://patentscope.wipo.int/search/ko/detail.jsf?docId=US354532117&_fid=KR308157823), [LINK2](https://patentscope.wipo.int/search/de/detail.jsf?docId=WO2020145509&_fid=KR300789094)