# Notes on references

Juan Corredor

October 21st, 2024

# 1 DDSP

[1] used as main reference for the research.

## 1.1 Important bullet points

- Interpretable and modular approach to generative modeling
- Classic signal processing elements with deep learning methods
- Models that rely on strided convolution or windowing (STFT) need to align waveshapes or suffer from spectral leakage.
- DDSP takes the approach of vocoders in using oscillators to synthesize signals.
- According to the DDSP paper [1], the DDSP library is capable of extrapolating timbres not seen during training, and independent control over pitch and loudness during synthesis.
- References to Neural Source Filter [2] imply this might be a good additional reference for research.

## 1.2 Newly learned concepts from this document

- Strided Convolution:

  Convolution that has a hop length, meaning, it skips some information to avoid analyzing redundancies.

- Teacher forcing:

  Feeding back the correct answers into training algorithms to reduce training times and lead the model in the right direction during training.

- Automatic differentiation:

  Also: 'algorithmic differentiation' or 'computational differentiation'

### 1.3 How it relates to my research

# References

[1] J. Engel, L. H. Hantrakul, C. Gu, and A. Roberts, "Ddsp: Differentiable digital signal processing," in *International Conference on Learning Representations*, 2020.

[2] X. Wang, S. Takaki, and J. Yamagishi, "Neural source-filter-based waveform model for statistical parametric speech synthesis," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5916–5920, IEEE, 2019.