Voctro Labs – TROMPA

Choir Singing Synthesis and Rehearsal Tools

Context

www.trompamusic.eu

Facilitate access to public-domain digital resources with state-of-the-art technologies

Voctro's contribution

- Build a rehearsal tool for choir singers
- Synthesize a large multi-lingual public-domain repertoire

Singing synthesis

Pitch / intonation

Timing / rhythm

Timbre, identity, quality

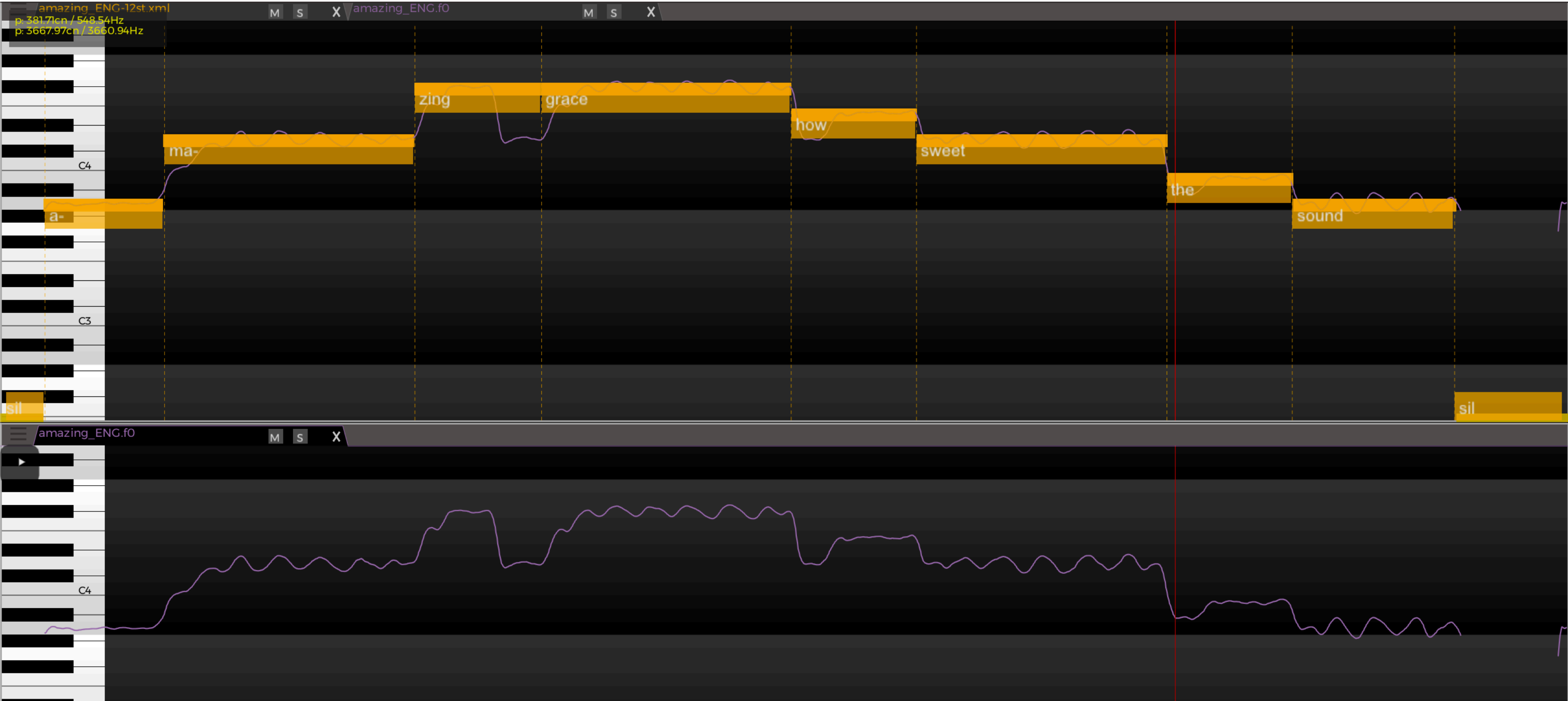
Phonetics / intelligibility

Dynamics / loudness

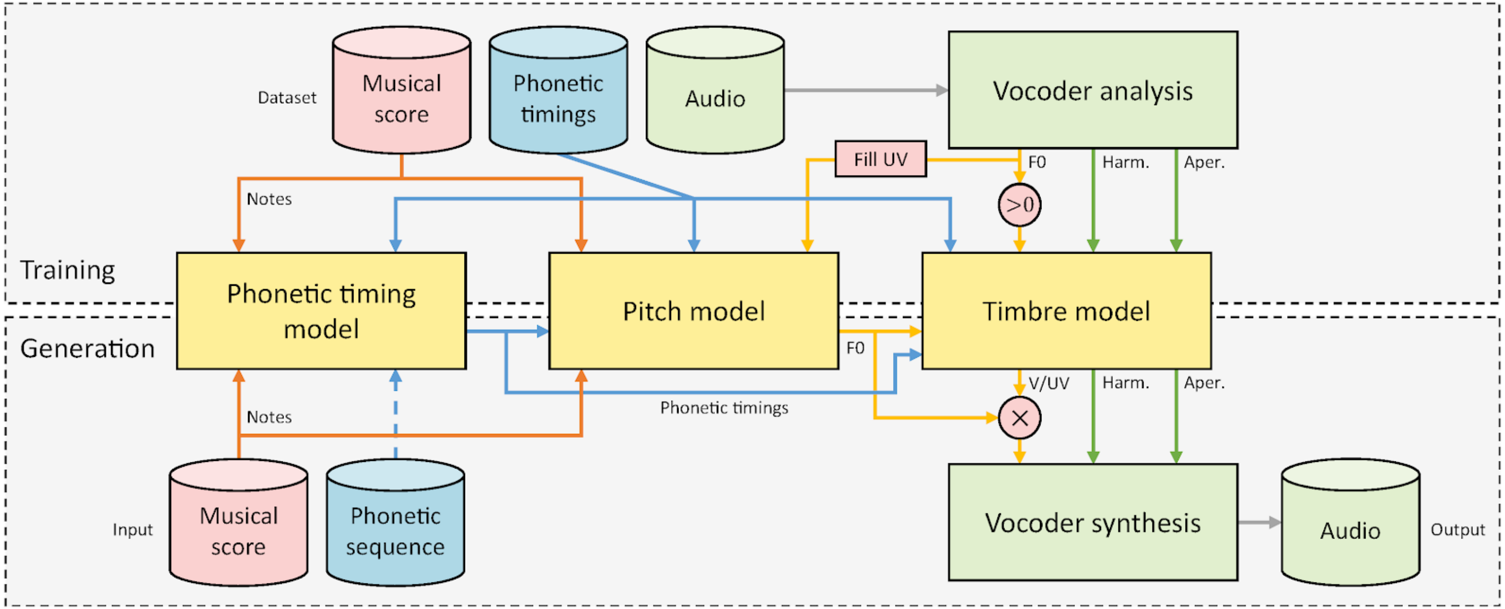
Style, vibrato, legato, …

Christian Mio Loclair
RayGan, GAN Sculptures

Modelling singing voice

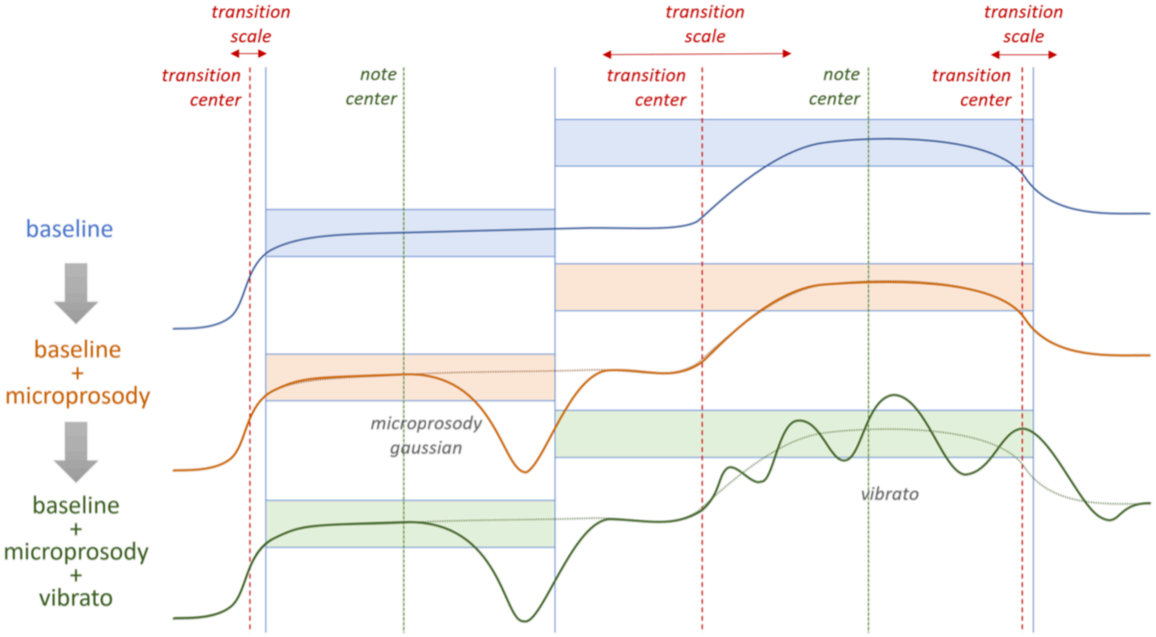


Singing synthesis



Baseline system: Blaauw, M. Bonada, J. (2017).
A Neural Parametric Singing Synthesizer.
In Interspeech 2017, Stockholm

Singing synthesis



Novel pitch model:

Bonada, J. , Blaauw, M. (2020).
HYBRID NEURAL-PARAMETRIC F0 MODEL FOR SINGING SYNTHESIS. In ICASSP 2020, Barcelona

Choir extensions

Specific datasets

16 professional singers (SATB)

Cor Francesc Valls, Barcelona

4 languages

Multiple voices

Voice cloning (DNN-speaker adaptation)

M. Blaauw, J. Bonada and R. Daido, “Data Efficient Voice Cloning for Neural Singing Synthesis," *ICASSP 2019*

https://mtg.github.io/singing-synthesis-demos/voice-cloning/

Large-scale repertoire

Choral Public Domain Library

Since 1998

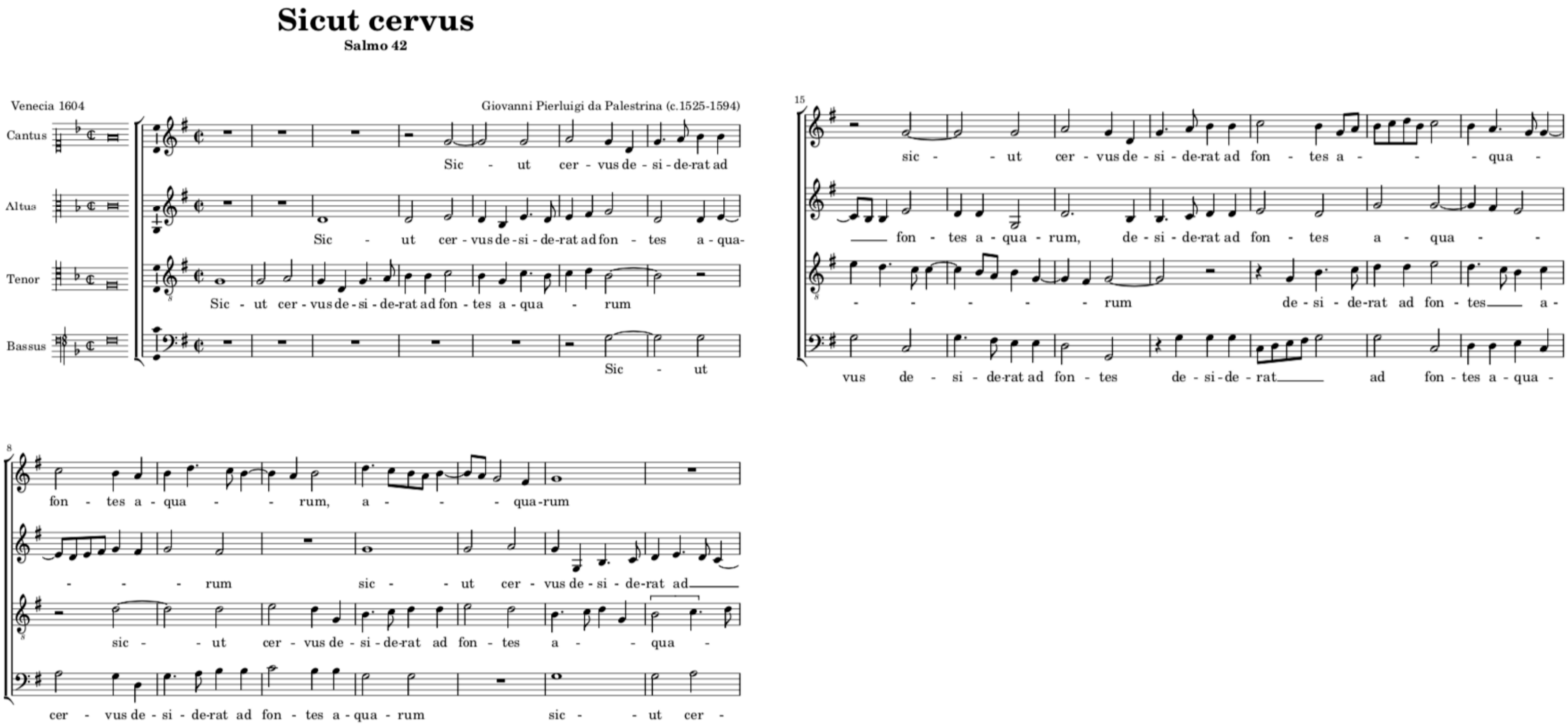
34,531 scores of free choral music

Prepared subset

4,107 scores in MusicXML format

In the 4 supported languages

Demos


Sicut cervus
Salmo 42
Venecia 1604
Giovanni Pierluigi da Palestrina (c.1525-1594)
Cantus
Altus
Tenor
Bassus
Sic - ut cer - vus de - si - de-rat ad
Sic - ut cer - vus de - si - de-rat ad fon - tes a - qua-
Sic - ut cer - vus de - si - de-rat ad fon - tes a - qua - rum
Sic - ut
8
fon - tes a - qua - - rum, a - - - qua-rum
- - - rum sic - - ut cer - vus de - si - de-rat ad
sic - - ut cer - vus de - si - de-rat ad fon - tes a - - qua -
cer - vus de - si - de-rat ad fon - tes a - qua - rum sic - - ut cer -
15
sic - - ut cer - vus de - si - de-rat ad fon - tes a - - - qua -
fon - tes a - qua - rum, de - si - de-rat ad fon - tes a - qua - -
- - - - - rum de - si - de-rat ad fon - tes a -
vus de - si - de-rat ad fon - tes de - si - de - rat ad fon - tes a - qua -

Rehearsal tool

Live demo

https://trompa.netlify.app/

Rehearsal tool

Collaboration with choirs

9 choirs, mainly from Catalonia

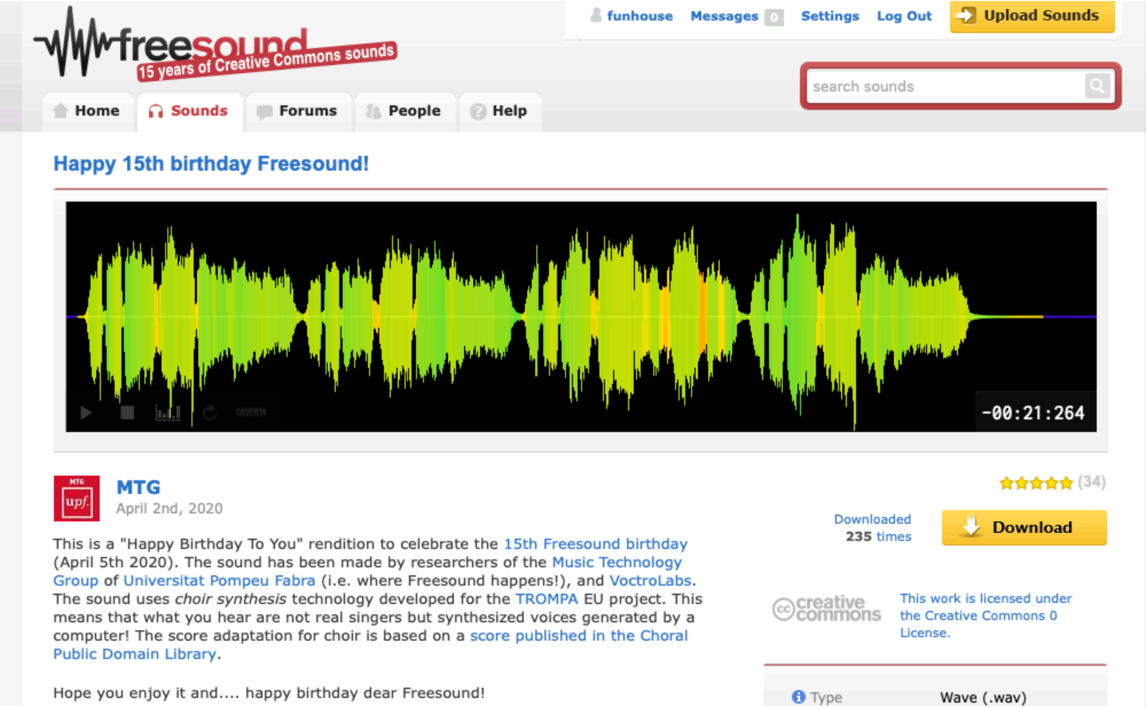
https://trompamusic.eu/node/105

Collaboration with Cantoría

Workshops for singers: access and practice Cantoría's repertoire
Perform a concert that will bring together all the participants in spring 2021

https://trompamusic.eu/node/123

Semi-supervised singing synthesis



Happy 15th birthday Freesound!

MTG

April 2nd, 2020

This is a "Happy Birthday To You" rendition to celebrate the 15th Freesound birthday (April 5th 2020). The sound has been made by researchers of the Music Technology Group of Universitat Pompeu Fabra (i.e. where Freesound happens!), and VoctroLabs. The sound uses *choir synthesis* technology developed for the TROMPA EU project. This means that what you hear are not real singers but synthesized voices generated by a computer! The score adaptation for choir is based on a score published in the Choral Public Domain Library.

Hope you enjoy it and.... happy birthday dear Freesound!

Downloaded 235 times

This work is licensed under the Creative Commons 0 License.

Type Wave (.wav)

https://freesound.org/people/MTG/sounds/511618/

Semi-supervised singing synthesis

- New system under development - submitted to ICASSP 2021
- Can learn new voices from audio data only, without annotations
- Encoder-decoder model:
  - two encoders – linguistic and acoustic
  - one (acoustic) decoder
- Training and inference:
  1. The entire system is trained in a supervised manner, using a labelled dataset.
  2. The system is adapted to a new target voice in an unsupervised manner, using the pretrained acoustic encoder
  3. At inference, the pretrained linguistic encoder is used together with the adapted decoder

https://mtg.github.io/singing-synthesis-demos/semisupervised/

Bonada, J., & Blaauw, M. (2020). Semi-supervised Learning for Singing Synthesis Timbre. *arXiv preprint arXiv:2011.02809*.

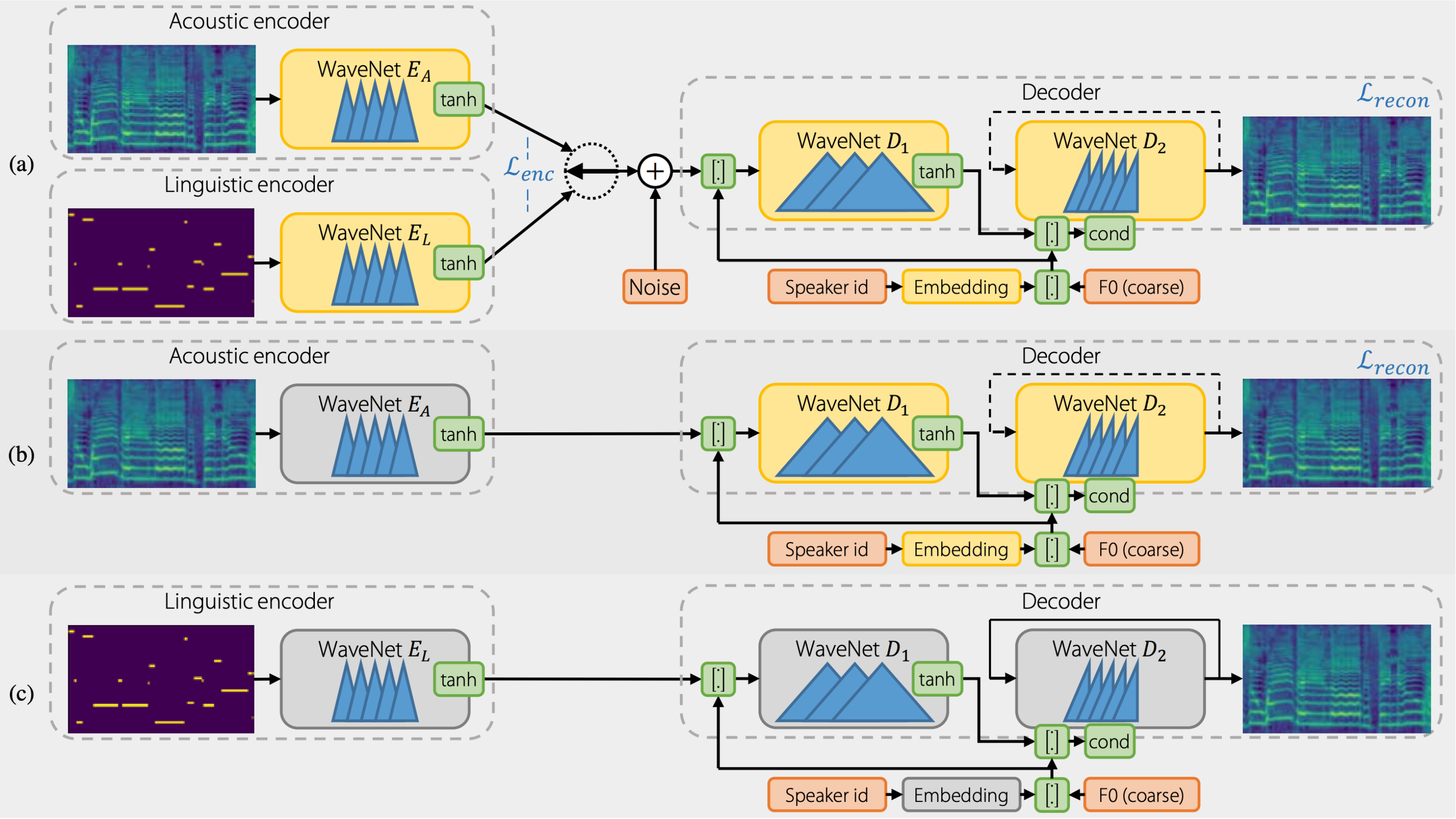


Fig. 1. A diagram of the model architecture in three different phases: (a) Training the encoder-decoder from annotated audio (supervised). (b) Training the decoder from audio (unsupervised). (c) Inference from linguistic features. Gray colored modules indicate their weights are kept fixed. The shape of the triangles in the WaveNet blocks represents the size of the receptive field and whether it is causal. A dashed autoregressive connection in the D_2 WaveNet block indicates teacher forced training with additive noise to avoid overfitting, while a solid connection indicates true autoregressive inference.

Thank you!

www.voiceful.io