Model Description

Dataset

This RAVE model is trained on a dataset of Jaap Blonk's vocal performances. The dataset was recorded by Blonk with Jonathan Reus specifically for training models as part of Reus's Dadasets project, in particular for text-to-voice synthesis in Reus and Victor Shepardson's Tungnaá instrument.

The dataset totals about 100 minutes of close-mic recordings of Blonk's solo voice performances, which are a mix of scored and improvised. Blonk performed in four moods: neutral, happy, worried, and aggressive. The scores were written in a reduced phonetic alphabet with 28 sounds.

Recordings were made across multiple sessions with AKG414 and KSM105 microphones. The dataset was edited to remove long silences, low-frequency noise, and occasional high-frequency interference, as well as stationary background noise using the FIR filter in Reaper.

Model Architecture

The model uses a causal variational RAVEv3 architecture with noise generator and spectral discriminator. Capacity was reduced to 64 from the default 96 to target lower resource hardware. The model sample rate is 48000 Hz and the minimum block size is the default 2048 samples.

Training was conducted using the `victor-shepardson/RAVE` fork, which reduces latency by one block. Data augmentation was applied, randomizing gain by +/- 1dB and speed by +/- 50c in addition to the defaults. The model was trained for 1M VAE-only steps and 2.5M adversarial steps.

The model was exported with 9 latent dimensions. Two additional features from the `victor-shepardson/RAVE` fork are noteworthy. First, a `noise_floor` attribute is enabled by default, which adds the same dequantization noise seen at training time, allowing hard zeros to correctly encode to silence in latent space. Second, sign normalization (https://nime.org/proc/nime2024_62/index.html) was applied, which causes each latent dimension to generally correlate the positive direction to loudness and/or brightness.

Artistic Intentions

This model was created as part of Dã̸͜ɗą͐sét̃s and in the process of ***BLA BLAVATAR vs JAAP BLONK*** – an experimental collaboration between Reus and Blonk, exploring voice dataset making as a laborious, improvisational and absurdly banal performance form.

Dã̸͜ɗą͐sét̃s (https://jonathanreus.com/portfolio/dadasets/) is an ongoing interdisciplinary music project, and a response to the cultural and economic ecosystem of voice AI and voice data that is rapidly terraforming the meaning, value and function of *voice*. The project involves researching the development of open digital music tools, bespoke voice datasets and artistic approaches to dataset creation that challenge the popular narratives and agendas around voice AI – which focus on the spectacle and fears around the technological results, such as digital clones that reproduce the voice of a famous narrator or singer, rather than the labor and values of vocal artists and laypersons who create these datasets.

BLA BLAVATAR vs JAAP BLONK (https://jonathanreus.com/portfolio/bla-blavatar-vs-jaap-blonk/) is an absurdist take on the cultural obsession with generative AI avatars and the goldrush to capitalise on creative automation. Rather than focusing on the uncanny realism of voice clones, the performance foregrounds the physical, creative and mental effort of performing sound poetry precisely according to an AI-friendly score, inspired by traditions of phonetically balanced reading scripts in speech research, and artistic traditions of typographical musical scores in sound poetry.

Reus performs as BLA BLAVATAR, using a custom real-time voice synthesis instrument called *Tungnaa*, which is trained on previous dataset poem performances by Blonk. The dynamic between Blonk and Reus sways from moments of precise and controlled material to improvised duets between Blonk and his vocal avatar. During the performance all of Blonk's vocalisations are recorded and added to a growing public dataset, and will be used as training data for making BLA BLAVATAR a better public speaker. Together, Blonk and Reus bring the vocal labor behind AI voice models to the stage, as well as use datasets and dataset making as a vocal art form that is fundamentally creative and exploratory, rather than a dehumanising means to an end.

From Jaap Blonk: “I would be very interested in future use of the AI model of my voice as developed by Jonathan Reus and Victor Riley Shepardson, especially because it includes not just semantic material but the whole range of possible voiced and unvoiced mouth sounds. Creative results from others working with the model would interest me in two ways: they could spurn me into further research of my own vocal possibilities, and they could give me material to work with in my electro-acoustic compositions, also along with my live voice.”

Audio Examples

The timbre transfer examples are included in two versions, `mix-` and `wet-`, where the `mix-` versions include the dry signal delayed by 85 milliseconds.

We processed the files using the RAVE VST in Reaper. Instead of the default pseudo-stereo, we set the RAVE VST to mono and used two instances for mid-side processing, so the two mono source files remain mono and the other three have a more similar stereo image to the originals.

For all files, we used a flat -6dB gain plus a -6dB low shelf filter at 50Hz to match the dynamic range of our dataset better.

The generation examples are made using the decoder only via NN.ar (https://github.com/elgiano/nn.ar) in SuperCollider, using a variety of simple waveforms and noise sources in latent space:

`pinknoise` uses pink noise which is slightly correlated across latent dimensions.

`sawwaves` uses falling saw waves with different relatively prime periods in each dimension.

`sines-medium` uses slower sine waves with relatively prime periods.

`3voice-composite` uses three correlated voices mixing gray noise and sine waves.

`3voice-drone` uses three voices with a static first dimension and slow sine waves for the other eight.

Acknowledgements

We acknowledge support by the European Union from call CNECT/2022/3482066 – Art and the digital: Unleashing creativity for European industry, regions, and society under grant agreement LC-01984767, as part of the S+T+ARTS programme; by the European Research Council (ERC) as part of the Intelligent

Instruments project (INTENT), under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 101001848); and the NVIDIA hardware grant.