flowEQ: A SMARTER EQUALIZER PLUGIN USING THE β-VAE

Christian Steinmetz Xavier Serra

Music Technology Group, Universitat Pompeu Fabra, Barcelona (Spain)

`christian.steinmetz@upf.edu`

EXTENDED ABSTRACT

We present flowEQ, an intelligent interface for a five-band parametric equalizer plugin. This tool enables the audio engineer to control all thirteen parameters of the equalizer using either one, two, or three sliders. This is achieved by learning low dimensional mappings from one, two, and three dimensional latent spaces to the thirteen dimensional parameter space of the equalizer. We use a disentangled variational autoencoder (β-VAE) trained on a dataset of equalizer configurations collected from audio engineers. Different levels of latent space disentanglement are investigated, providing the user with the ability to adjust the latent representations directly within the plugin. By traversing the latent spaces of the trained models, the user can quickly search through relevant configurations of the equalizer. This enables both novice and experienced audio engineers to more efficiently apply timbral processing during post-production.
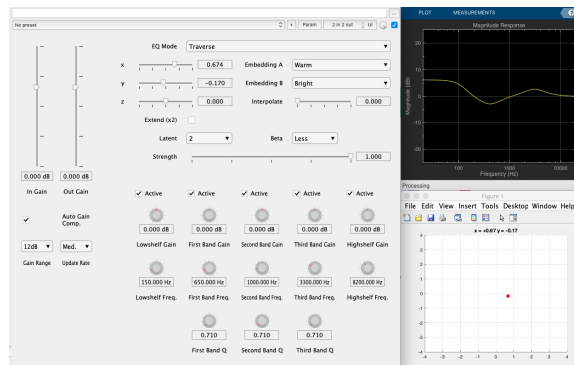
Figure 1. VST plugin running in REAPER with real-time MATLAB visualization tool.

The parametric equalizer is the de-facto tool used by audio engineers to shape the timbre of an audio signal. First employed with analog circuitry in the early 1970s [4], these processors have evolved with the advent of digital signal processing, and are readily available as plugins or standalone outboard processors. Providing direct control over the gain, center frequency, and bandwidth (known as Q factor) for each band, the parametric equalizer is one of the most powerful forms of the equalizer. Due to this increased complexity, experience is required on the part of the audio engineer in order to achieve the desired timbral effect. Our goal is to provide a simplified interface to efficiently search across relevant configurations of the equalizer without extensive experience.

Previous work has investigated the use of timbral semantics for the intelligent control of equalizers. In [1], the authors collect and analyze crowdsourced data to uncover universal, semantic descriptors of timbre, with the goal of designing an equalizer that can respond to descriptive language. The SAFE project [5] introduces compression, distortion, reverb, and parametric equalization plugins, which provide a means to collect semantic data directly from the normal workflow of audio engineers. The work presented in [6] is the most similar to ours. They use data from the SAFE project and employ a number of classical dimensionality reduction techniques to build a model that attempts to control the 'warmth' and 'brightness' of an equalizer.

While these semantic interfaces present potentially promising results in the control of an equalizer, they are inherently limited by the generalization and relative specificity of the semantic descriptors studied. flowEQ expands upon this work by employing a more modern and powerful model, the variational autoencoder, along with the investigation of variable latent space dimensionality and disentanglement.

We use the VAE formulation originally presented in [3], where we attempt to jointly learn the parameters, ϕ and θ, of a probabilistic encoder, $q_\phi(\mathbf{x}|\mathbf{z})$, and a generative decoder, $p_\theta(\mathbf{z}|\mathbf{x})$. During optimization we minimize the expected negative log-likelihood for the data generated by the decoder, after sampling a latent code from the encoder. This acts as a reconstruction loss, incentivising the model to accurately encode and decode the equalizer parameters. We assume the latent prior is gaussian, such that $p(\mathbf{z}) = \mathcal{N}(0, 1)$. We then utilize the KL divergence between the encoder, $q_\phi(\mathbf{x}|\mathbf{z})$, and the prior, as a regularizer, which incentivizes the encoder to learn well structured latent mappings.

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}, \mathbf{x}^{(i)}) = -\mathbb{E}_{\mathbf{z} \sim q_\theta(\mathbf{z}|\mathbf{x}^{(i)})}[\log p_\phi(\mathbf{x}^{(i)}|\mathbf{z})] + \beta \, D_{KL}(q_\theta(\mathbf{z}|\mathbf{x}^{(i)})||p(\mathbf{z})). \tag{1}$$

Equation 1 shows the loss function, with the addition of a hyperparameter, β, known as the disentanglement factor [2]. Adjusting β, the relative weight of the reconstruction and KL loss, allows us to investigate different levels of latent space disentanglement. Greater disentanglement means the latent space is structured so that each dimension more closely models a single characteristic of the data. We train a total of twelve different models with three different latent space bottlenecks (1D, 2D, and 3D), each with one of four values of β.

The plugin is implemented with the MATLAB Audio Toolbox[1], which provides an API for compiling MATLAB code directly to VST/AU plugins. Compiled plugins for macOS and Windows are made available on the project website[2], as well as the source code[3]. Three modes of operation are featured in the plugin. *Traverse* mode allows the user to directly sample from the latent space of the selected model. The user can select among the twelve models to investigate different levels of latent space dimensionality and disentanglement. In *semantic* mode, semantic descriptor labels are used to build a linear model in each latent space space to classify 'warmth' and 'brightness'. Pre-computed semantic embeddings are provided and the user has the ability to interpolate between two different selected embeddings. Finally, *Manual* mode allows the user to directly control each of the parameters in the equalizer.

flowEQ provides an intelligent modality for interfacing with the parameter space of a five-band parametric equalizer. By employing a β-VAE, we build a number of latent variable models. A real-time audio plugin provides audio engineers with the ability to directly interface with these models. By selecting among the models, the user can adjust the latent space representation, with control over dimensionality and disentanglement. This tool provides a simplified interface to efficiently search across relevant configurations of the equalizer, allowing for greater efficiency in the search for the desired timbral transformation.

REFERENCES

[1] M. Cartwright and Bryan A. Pardo. Socialeq: Crowdsourcing an equalization descriptor map. 2013.

[2] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. β-VAE: Learning basic visual concepts with a constrained variational framework. *ICLR*, 2(5):6, 2017.

[3] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv e-prints*, page arXiv:1312.6114, Dec 2013.

[4] George Massenburg. Parametric equalization. May 1972.

[5] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. D. Reiss. SAFE: A system for the extraction and retrieval of semantic audio descriptors. 2014.

[6] Spyridon Stasis, Ryan Stables, and Jason Hockman. Semantically controlled adaptive equalisation in reduced dimensionality parameter space. *Applied Sciences*, 6(4), 2016.

---

[1] https://www.mathworks.com/products/audio.html
[2] https://flowEQ.ml/
[3] https://github.com/csteinmetz1/flowEQ