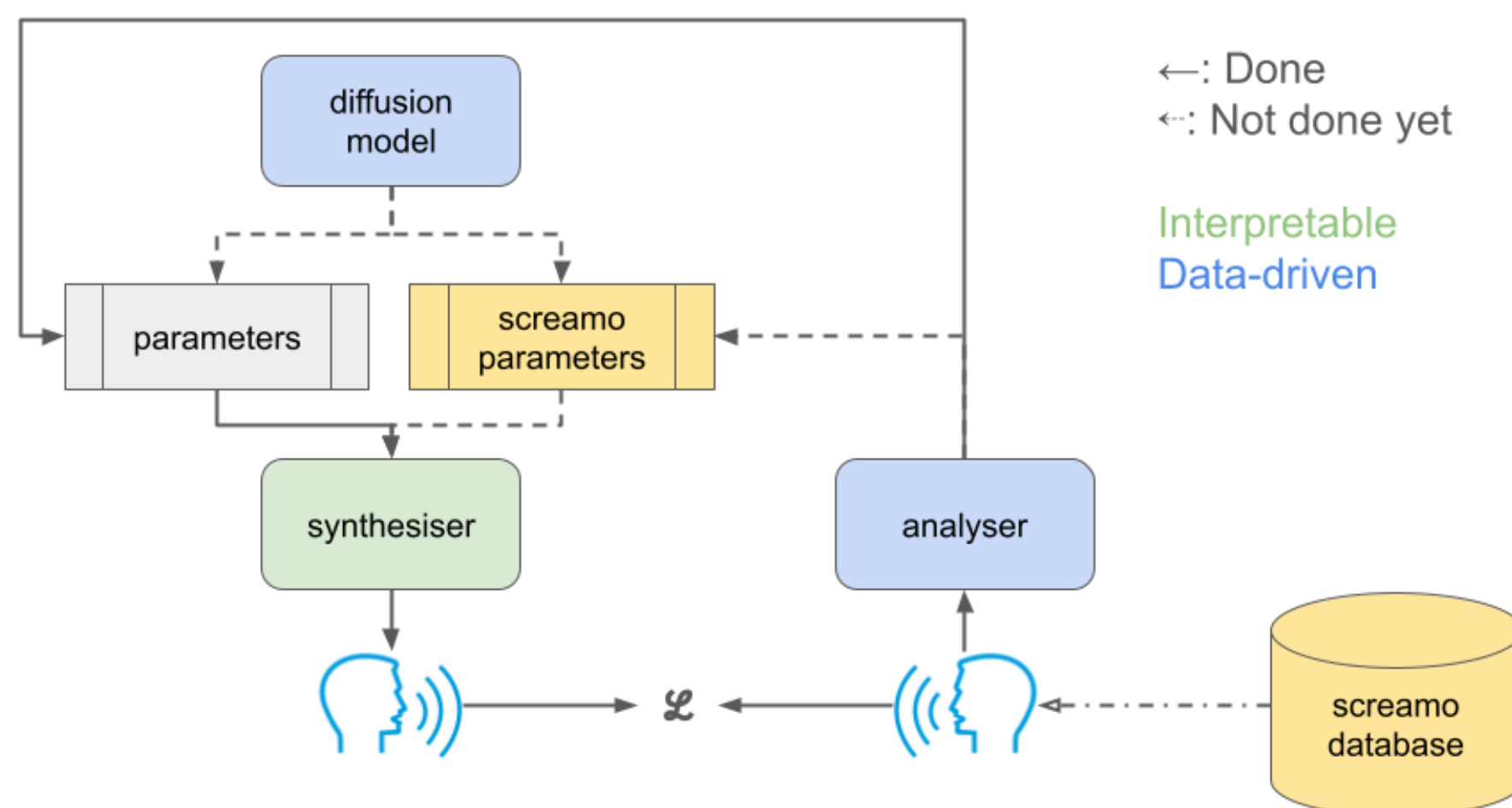


Analysing and controlling extreme vocal expression using differentiable DSP and generative models

Chin-Yun Yu

2022



2

Finding

1) Leveraging DSP components as prior knowledge reduces computational loads and training resources and increases controllability. 2) The available techniques for posterior sampling in diffusion models are still not capable of inverse problems that have obscure clues of the original signal (e.g., monotimbral sources).

Question

Can we find a set of parameters that 1) describe the screaming vocal in a way that correlates with human perception and 2) can control the synthesiser in an efficient way?