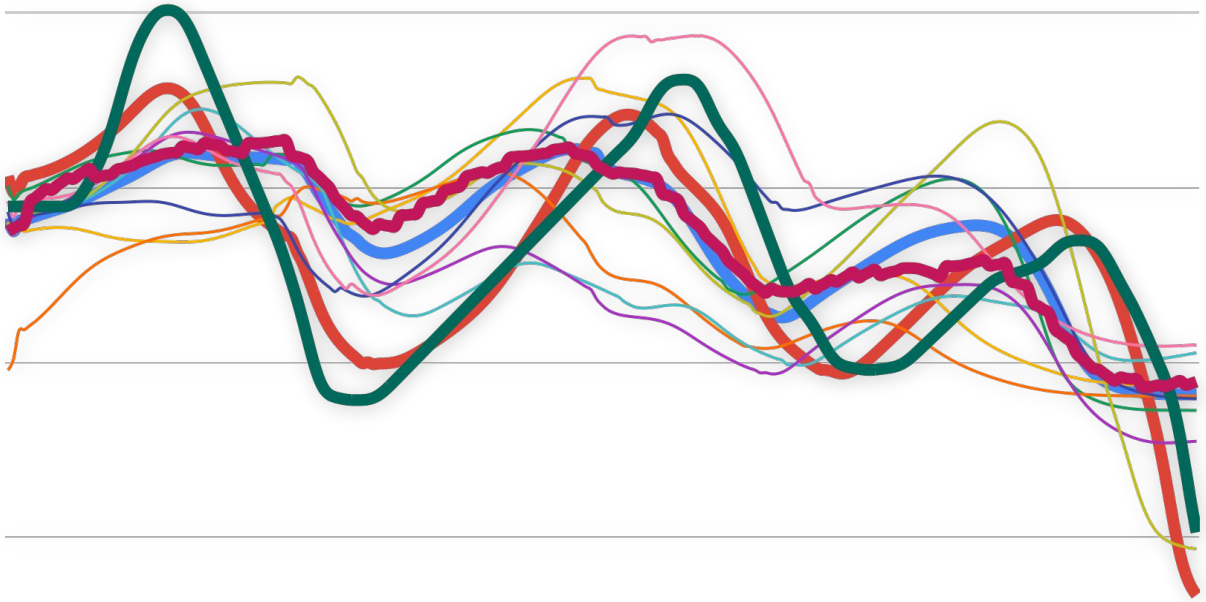
Google

CHiVE

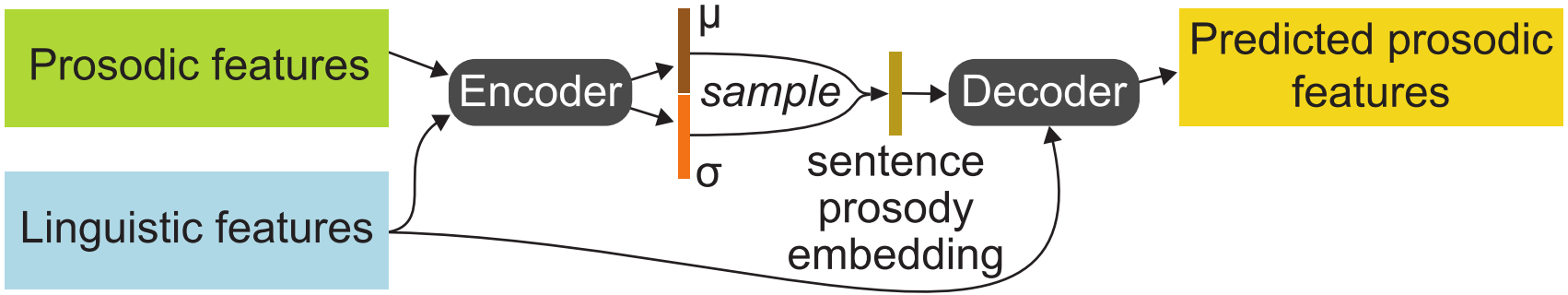
Varying Prosody in Speech Synthesis with a Linguistically Driven Dynamic Hierarchical Conditional Variational Network

Vincent Wan, Chun-an Chan, Tom Kenter, Jakub Vit, Rob Clark

Modelling intonation in prosody

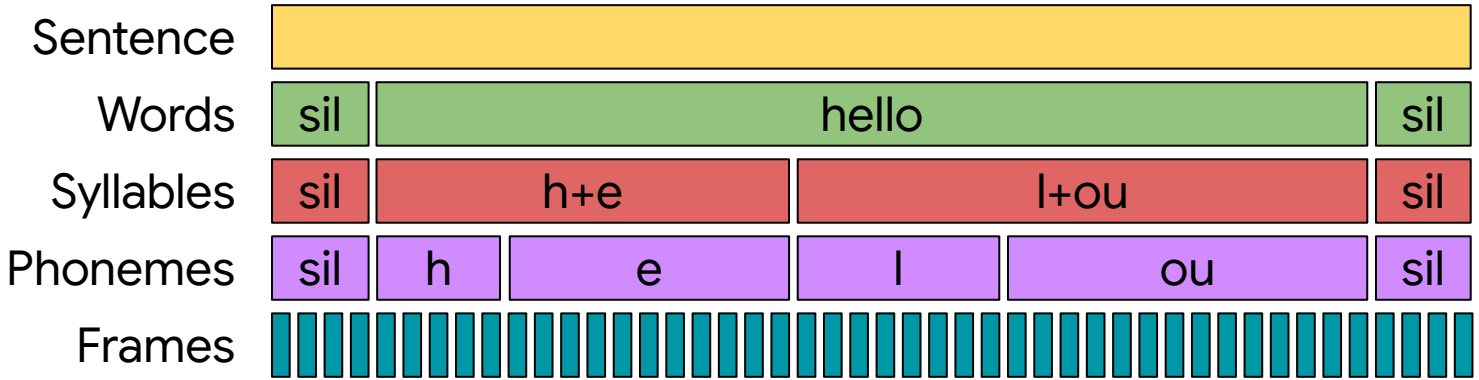


A conditional variational autoencoder captures the different intonations



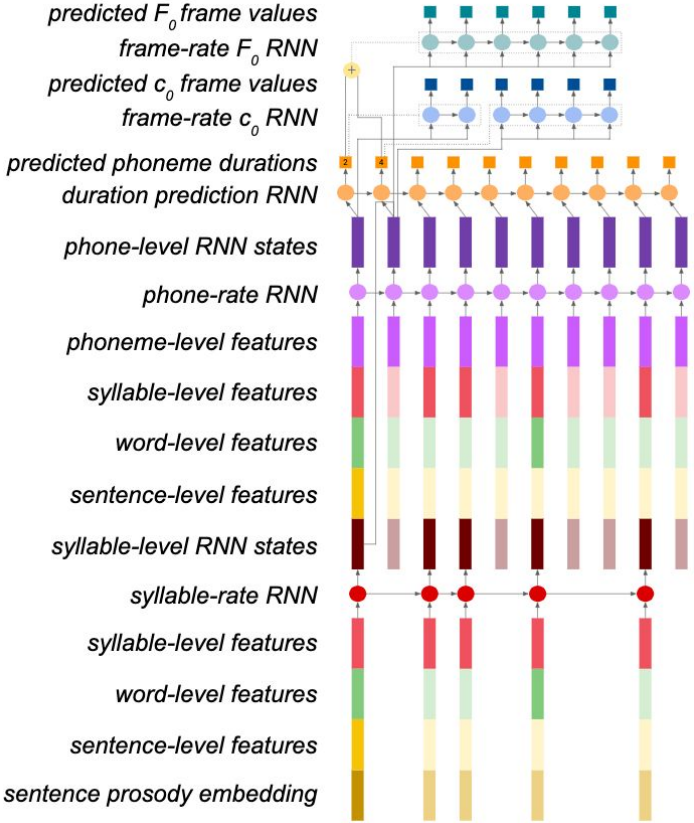
Google

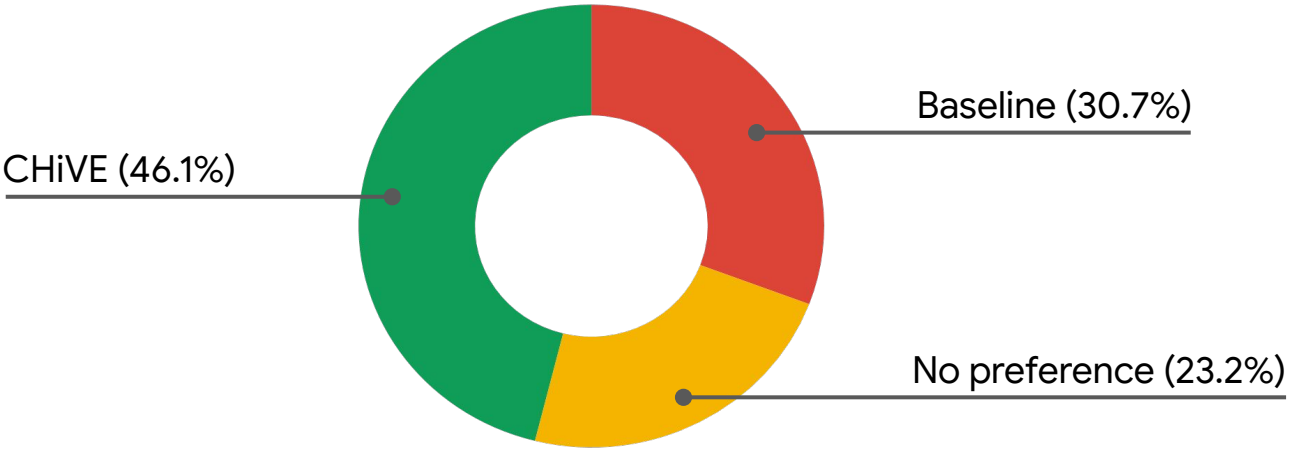
Language has a hierarchical linguistic structure



Google

Add linguistic knowledge to the network



Google

The structured model is better