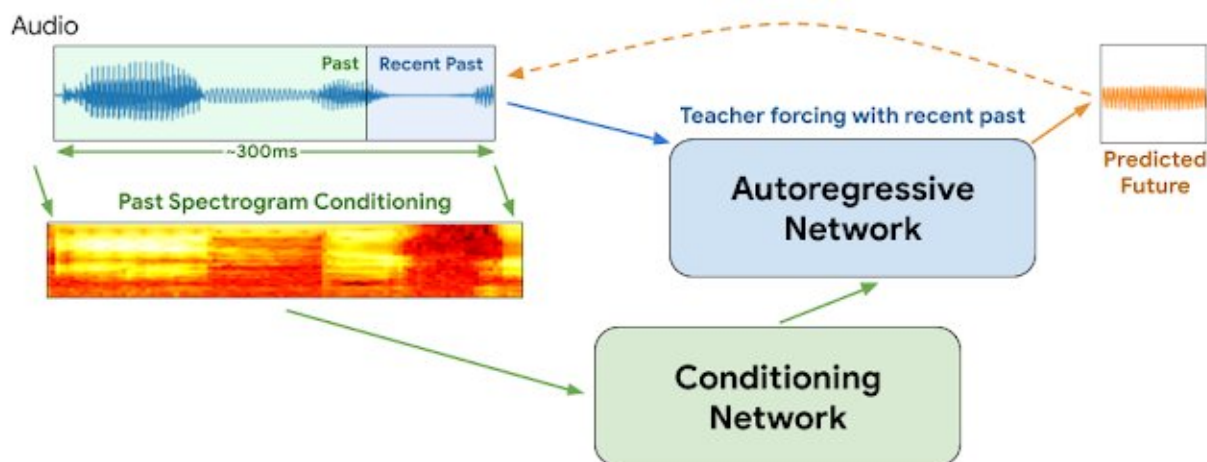# Google Duo audio boost won't leave you hanging on the phone

April 3 2020, by Peter Grad

WaveNetEQ architecture. During inference, we "warm up" the autoregressive network by teacher forcing with the most recent audio. Afterwards, the model is supplied with its own output as input for the next step. A MEL spectrogram from a longer audio part is used as input for the conditioning network. Credit: Google

"It's good to hear your voice, you know it's been so long
If I don't get your calls, then everything goes wrong…
Your voice across the line gives me a strange sensation"
— Blondie, "Hanging on the Telephone"

In 1978, Debbie Harry propelled her new wave band Blondie to the top of the charts with a plaintive tale of yearning to hear her boyfriend's

[voice](#) from afar and insisting he not leave her "hanging on the telephone."

But the questions arises: What if it were 2020 and she was speaking over VOIP with intermittent packet losses, audio jitter, network delays and out-of-sequence packet transmissions?

We'll never know.

But Google this week announced details of a new technology for its popular Duo voice and video app that will help ensure smoother voice transmissions and reduce momentary gaps that sometimes mar internet-based connections. We'd like to think Debbie would approve.

We've all experienced Internet audio jitter. It occurs when one or more packets of instructions comprising a stream of audio instructions are delayed or shuffled out of order between caller and listener. Methods employing voice packet buffers and artificial intelligence generally can smooth over jitter of 20 milliseconds or less. But the interruptions become more noticeable when the missing packets add up to 60 milliseconds and greater.

Google says virtually all calls experience some data packet loss: one-fifth of all calls lose 3 percent of their audio and one-tenth lose 8 percent.

This week, Google researchers at the DeepMind division reported that they have begun using a program called WaveNetEQ to address these issues. The algorithm excels at filling in momentary sound gaps with synthesized but natural-sounding speech elements. Relying on a voluminous library of speech data, WaveNetEQ fills in sound gaps up to 120 milliseconds. Such sound bit swaps are called packet loss concealments (PLC).

"WaveNetEQ is a generative model based on DeepMind's WaveRNN technology," Google's AI Blog reported April 1, "that is trained using a large corpus of speech data to realistically continue short speech segments enabling it to fully synthesize the raw waveform of missing speech."

The program analyzed sounds from 100 speakers in 48 languages, zeroing in on "the characteristics of human speech in general, instead of the properties of a specific language," the report explained.

In addition, sound analysis was tested in environments offering a wide variety of background noise to help ensure accurate recognition by speakers on busy city sidewalks, train stations or cafeterias.

All WaveNetEQ processing must run on the receiver's phone so that encryption services are not compromised. But the extra demand on processing speed is minimal, Google asserts. WaveNetEQ is "fast enough to run on a phone, while still providing state-of-the-art audio quality and more natural sounding PLC than other systems currently in use."

Sounds samples illustrating audio jitter and improvement with WabeNetEQ are posted on the Google Blog report.

  More information: ai.googleblog.com/2020/04/impr … ity-in-duo-with.html

© 2020 Science X Network