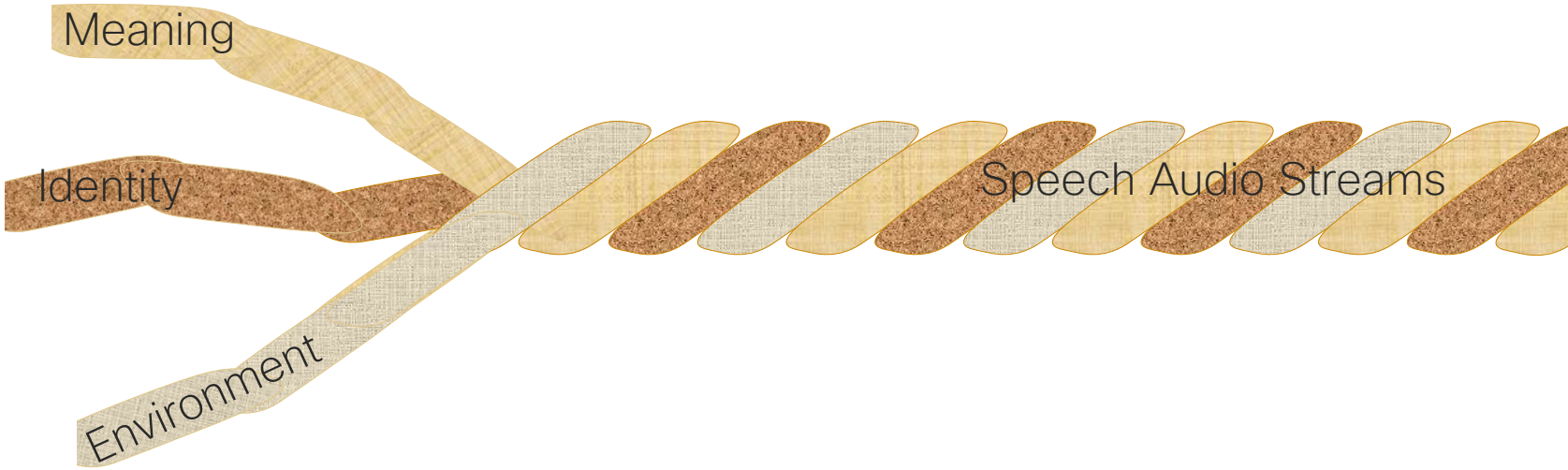
AI Wizardry in Webex

Advances in speech technology for differentiated user experiences

March 2021

Unraveling the richness of speech streams



Speech is complex, multi-faceted and ubiquitous – many applications in recognition, enhancement, generation, and analytics. Perfect for AI

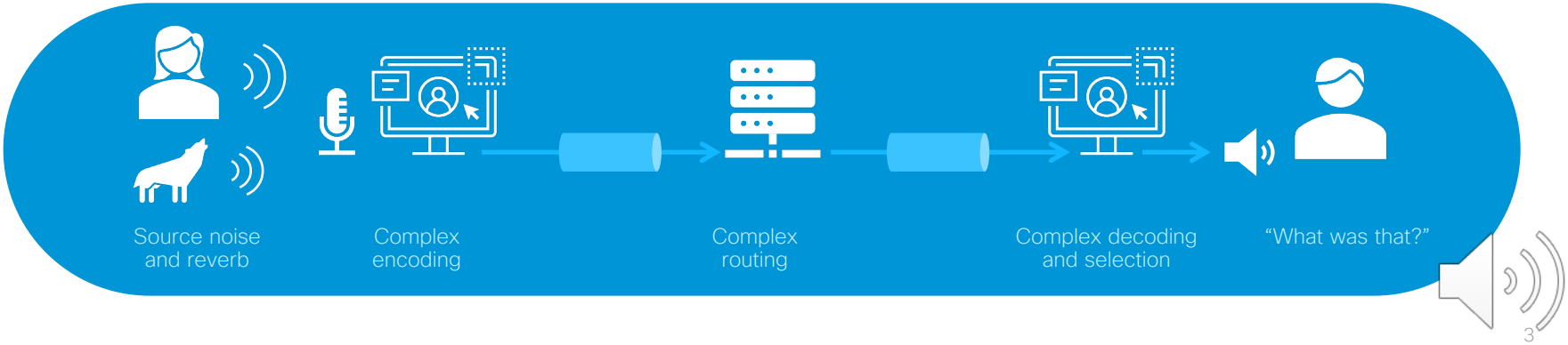
What matters in speech?

Criteria

- Comprehension rate
- Cognitive load
- Latency
- Computing overhead
- Privacy and security

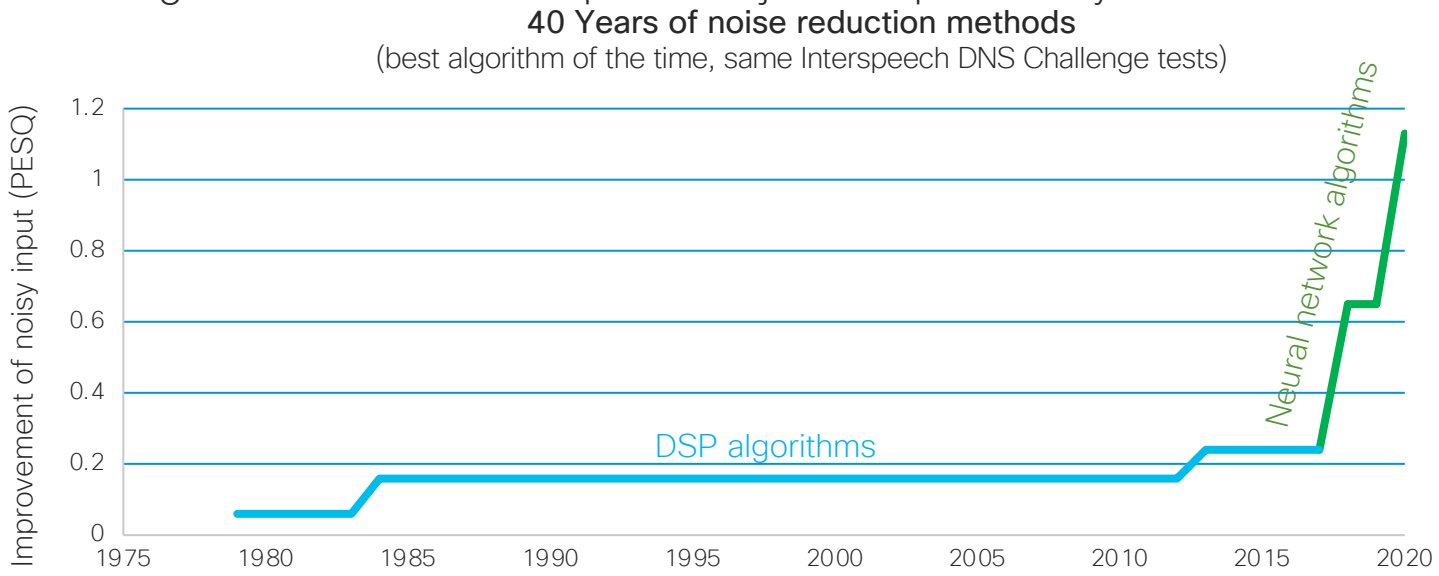
Audio Challenges

- Environmental noise
- Background talkers
- Reverberation
- Network latency and starvation
- Bandwidth compression
- Uneven talker volume



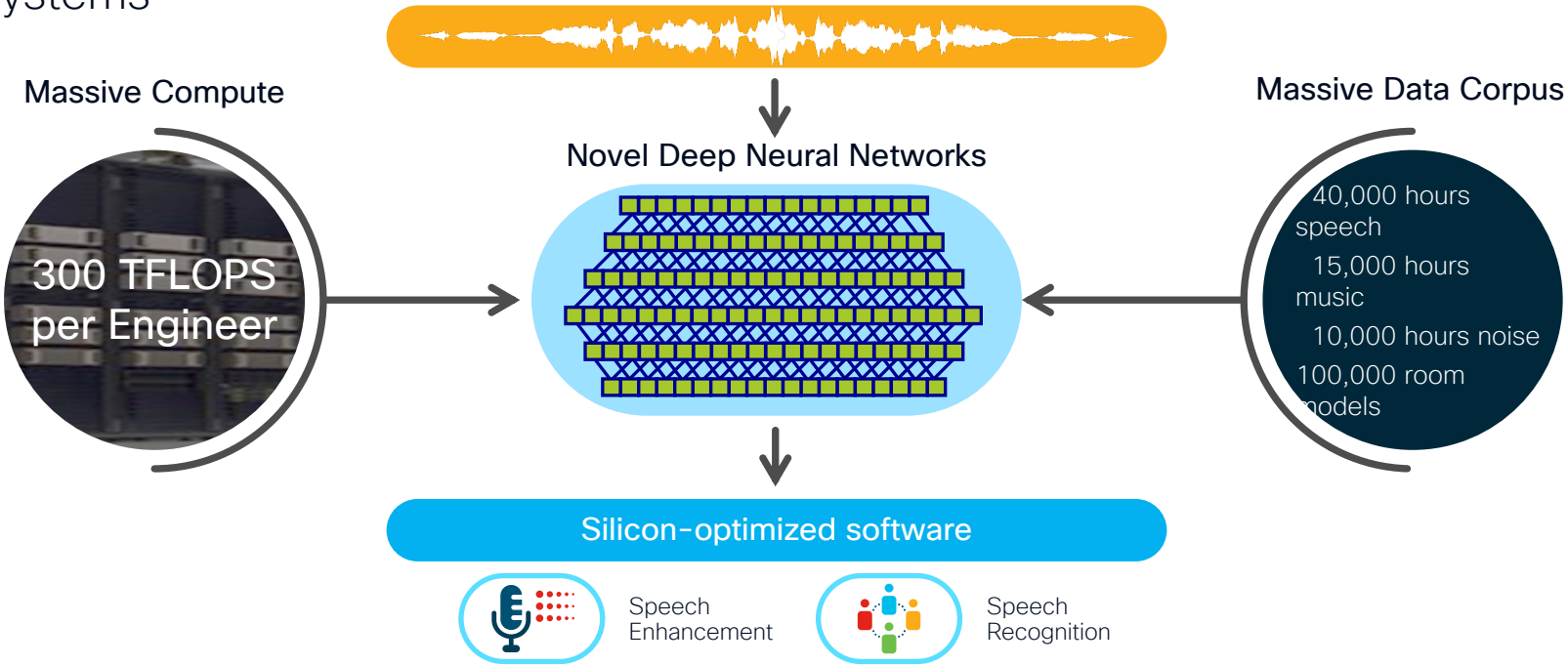
A brief history of speech enhancement

- Surprisingly, there has been little change in state-of-the-art algorithms over four decades
- Deep learning revolution reaches speech in just the past two years



BabbleLabs: AI Speech Wizardry

AI meets speech – deep experience in speech science, AI/ML, embedded systems



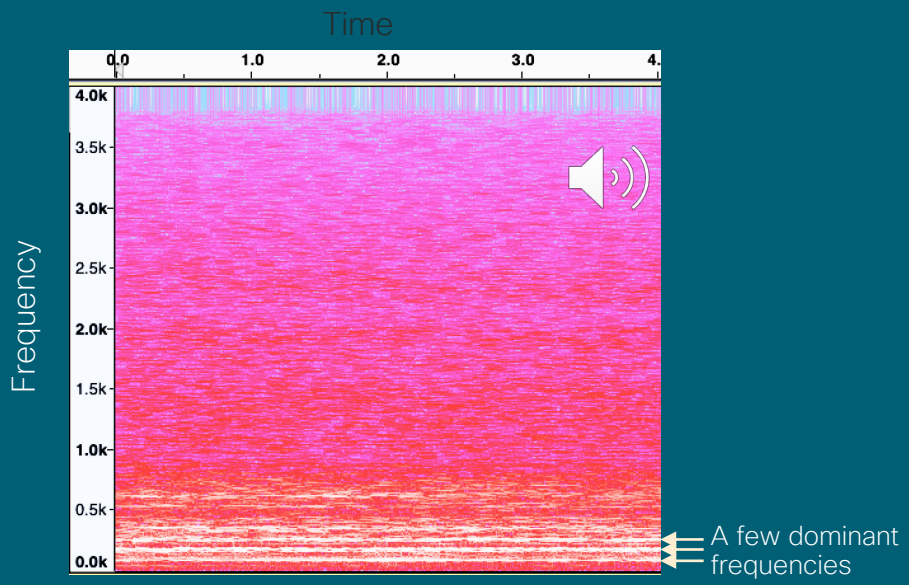
Separating human speech from stationary noise

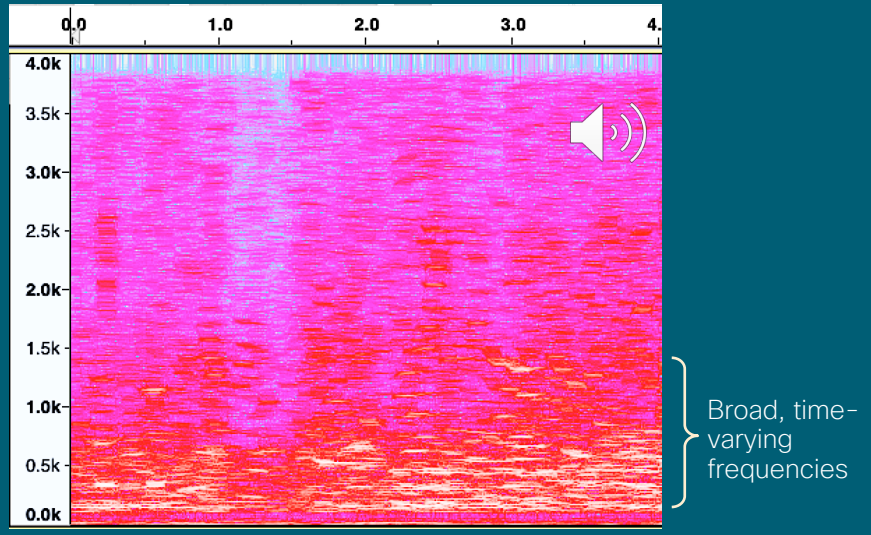
Stationary noise: Fan

- Steady frequency over time
- Most energy below speech bands (most critical: 250-2500Hz)

Non-stationary noise: Human babble noise

- Time variation makes conventional adaptive filters ineffective
- Squarely in speech bands



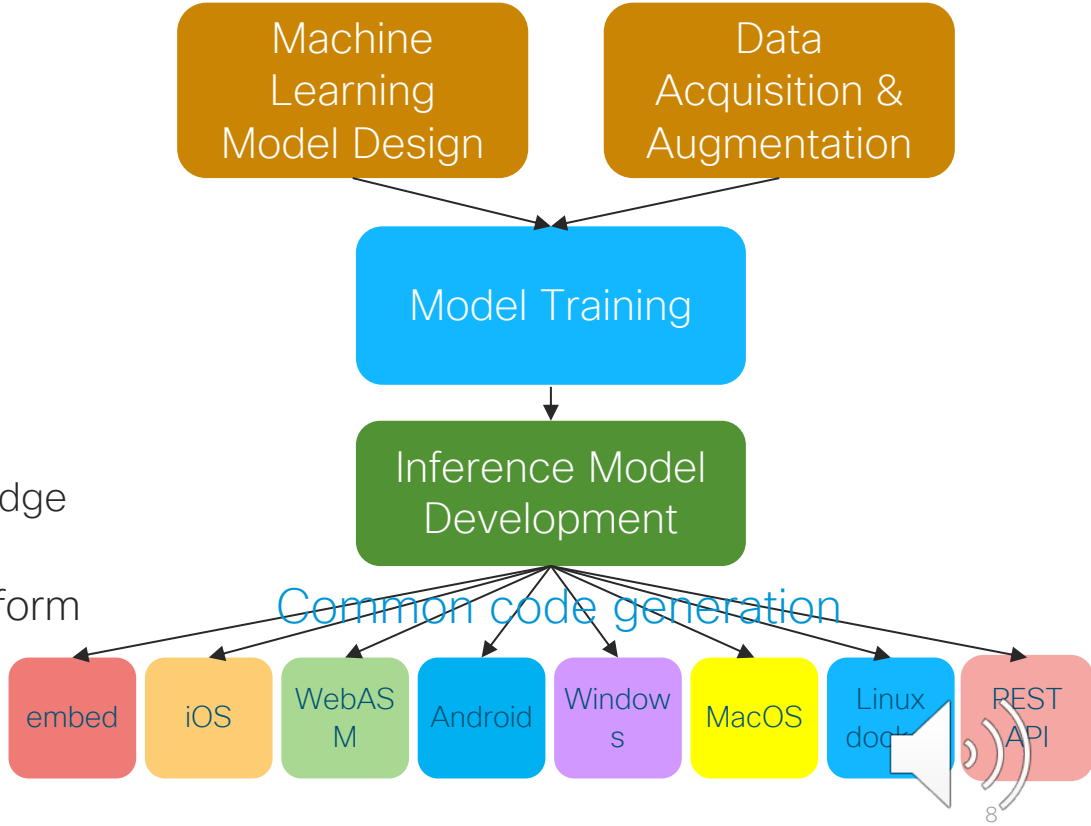


Speech enhancement demos

Strong coverage across devices, languages and noises

Implementing AI based speech enhancement

- Smart:
  - Separates of speech from unwanted sounds: Typing, dog barking, traffic, appliances
  - Pulls intelligible speech from human babble
  - Reduces noise by >20db (100x)
- Efficient:
  - Optimized code generation across edge and cloud deployments
  - Scalable compute to fit target platform



What does it mean for the Webex audio experience?

Remove keyboard and noisy disruptions

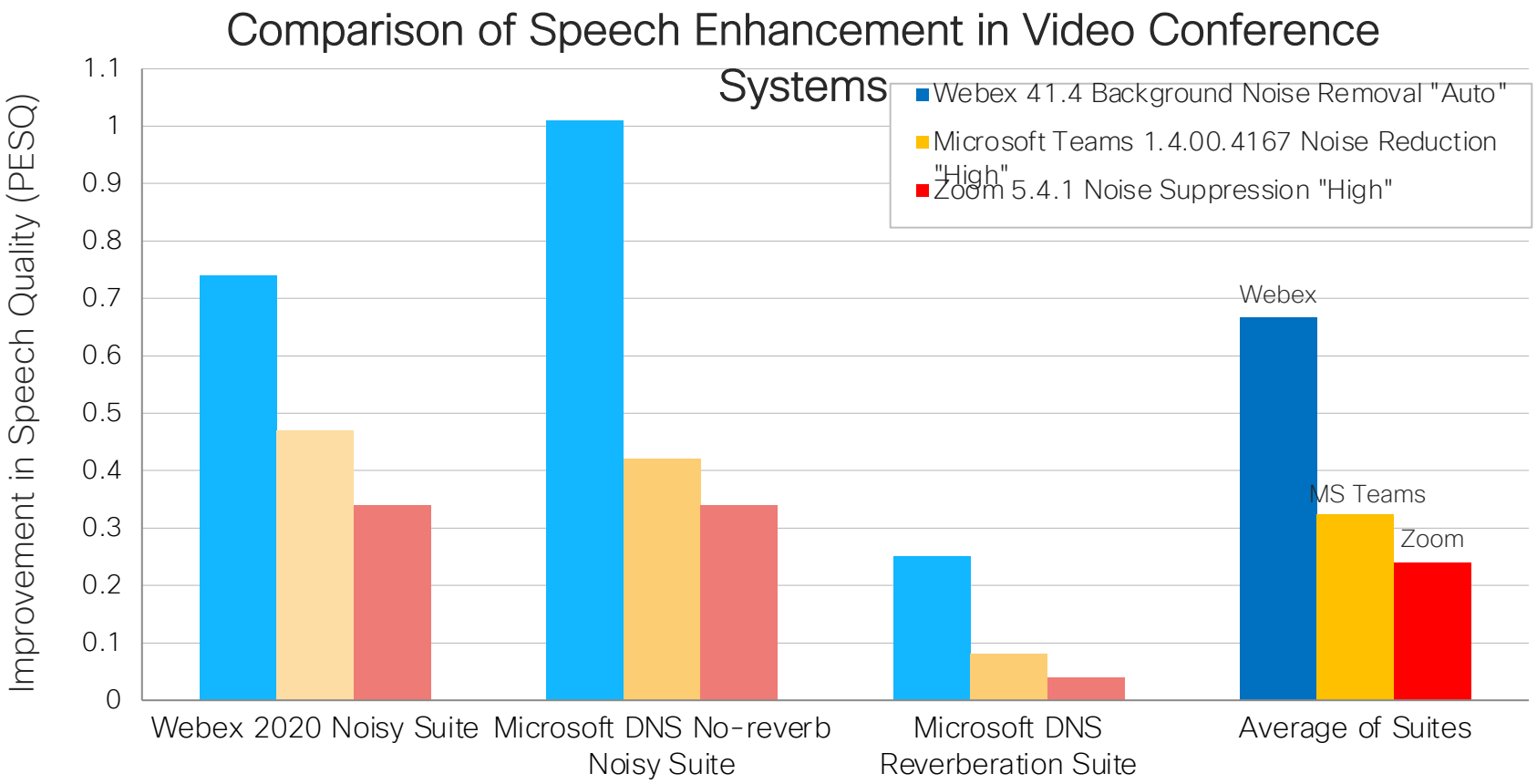
Working from home
with standard audio technology

Remove reverberation to bring speaker ”closer”

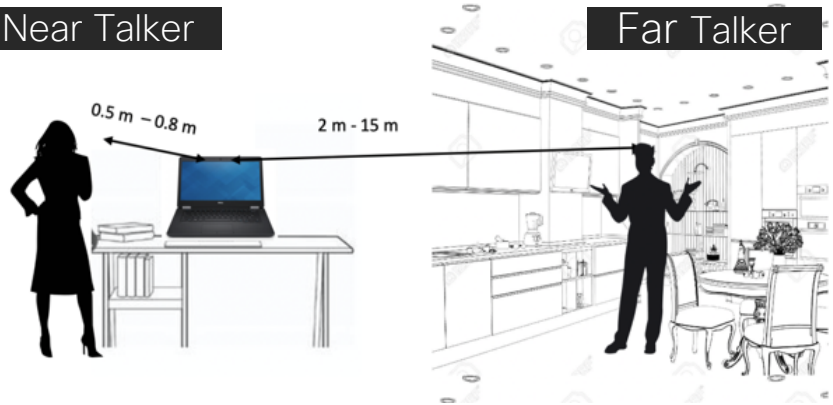
Distant speaker in a video call

9

Leadership in End-to-End Noise Removal



Near and Far Talkers



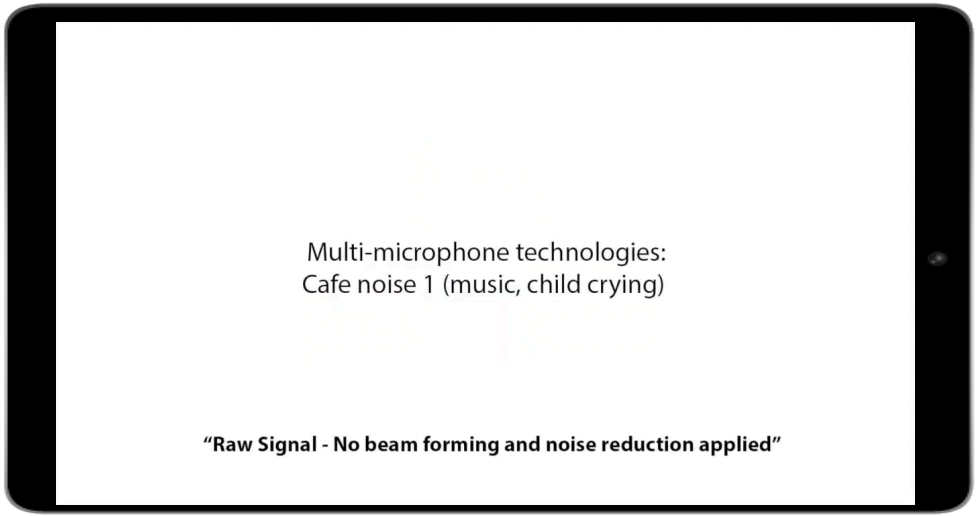
Scenario	Today	Coming Soon
Single-talker – suppress background talkers Near field talker = 0.5m(Savita) Far field talker = 3m		
Multi-talker – enhance background talkers Near field talker = 0.5m (Samer) Far field talker = 4m		

Multi-microphone Devices

Leveraging full power of leading-edge laptops, room devices, phones

Webex Desk Navigator

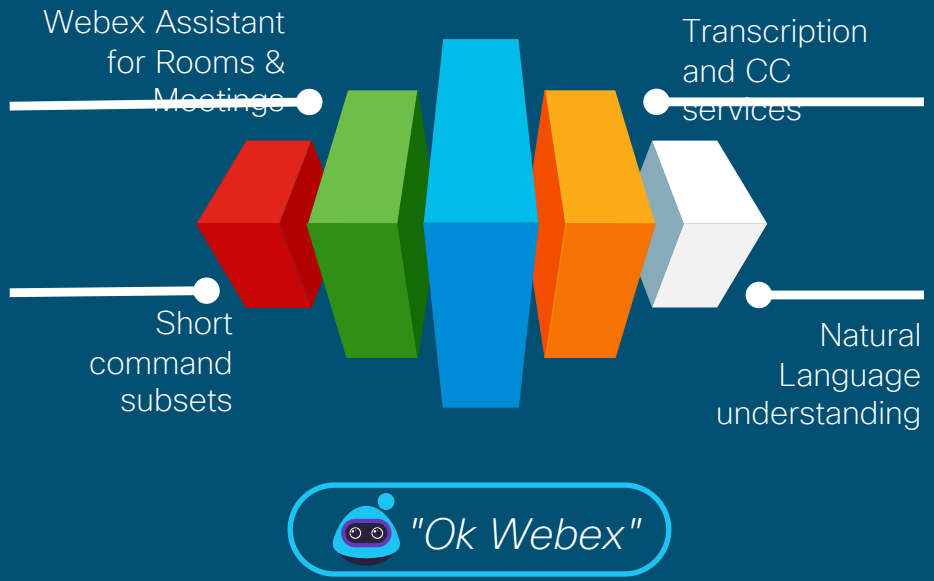
Webex Desk Plus

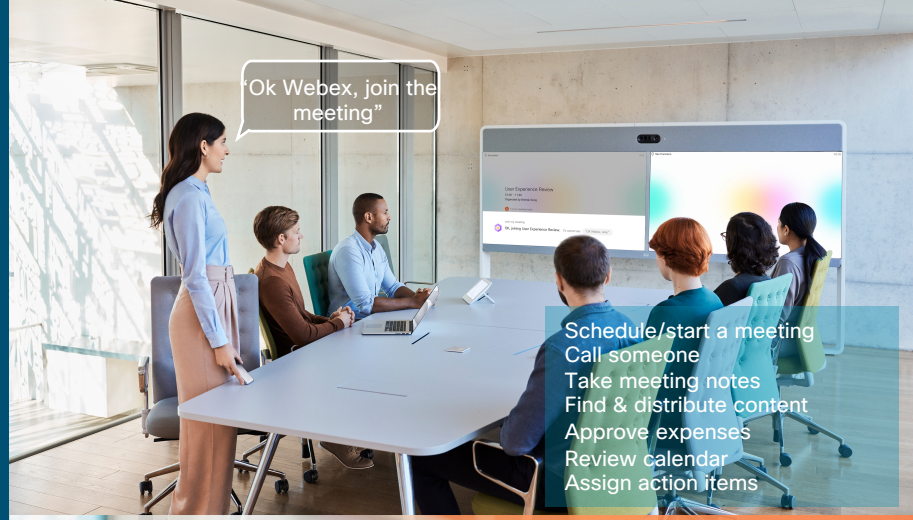


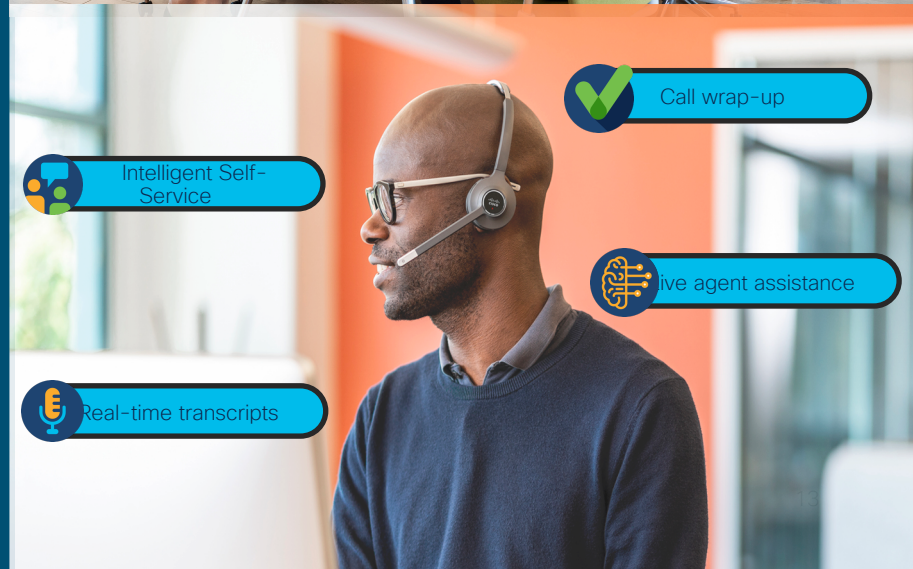
Comparison:

1. Single microphone audio
2. Multi-microphone audio

Speech Recognition





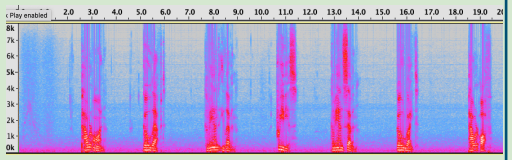
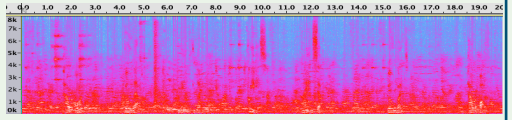


What's coming next?
Robust Local Speech Recognition

Recognition rates degrade sharply under real-world noisy conditions

Local recognition can be **faster**, more **accurate**, and more **private**.

- Accurately recognize single word/phrase to wake up the system.
- Accurately recognize command set without re-training the primary network

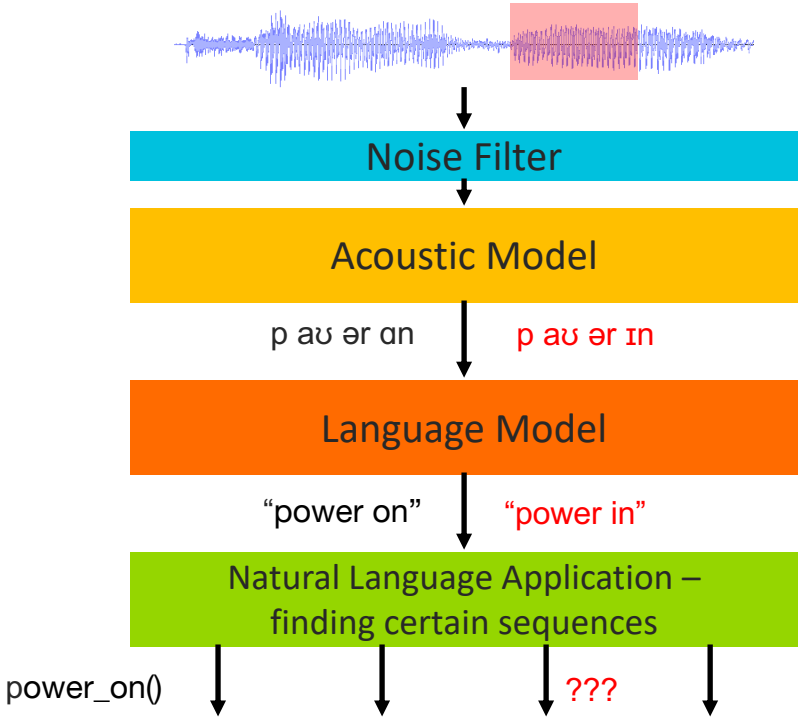
Signal to Noise Ratio	Spectrogram	IBM Watson	Webex Local Command Recognition
>20 dB		Power on Go left Command eleven Previous Command seven Go right Power off	Power on Go left Command eleven Previous Command seven Go right Power off
0 dB		— Yes — — — — —	Power on Go left Command eleven Previous Command seven Go right Power off

e.g. : Local Customization of Conference Controls

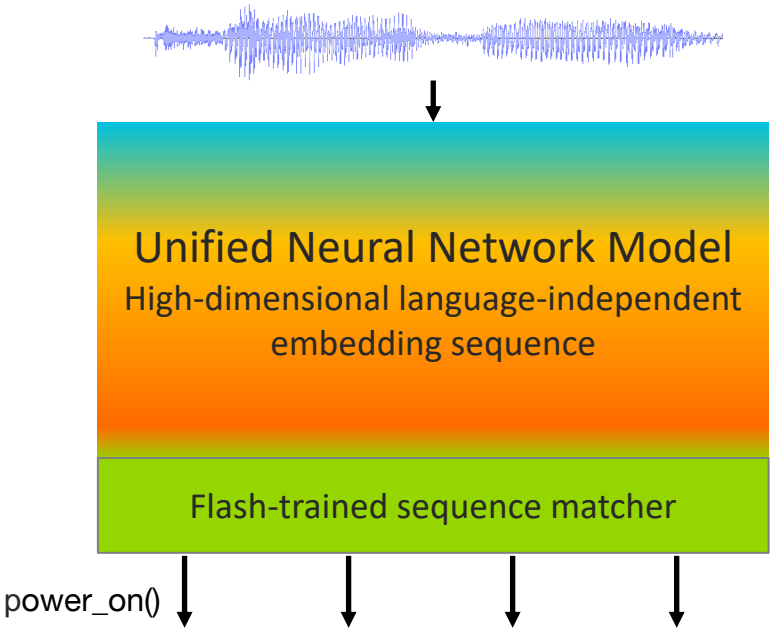
14

What's happening?

The usual way



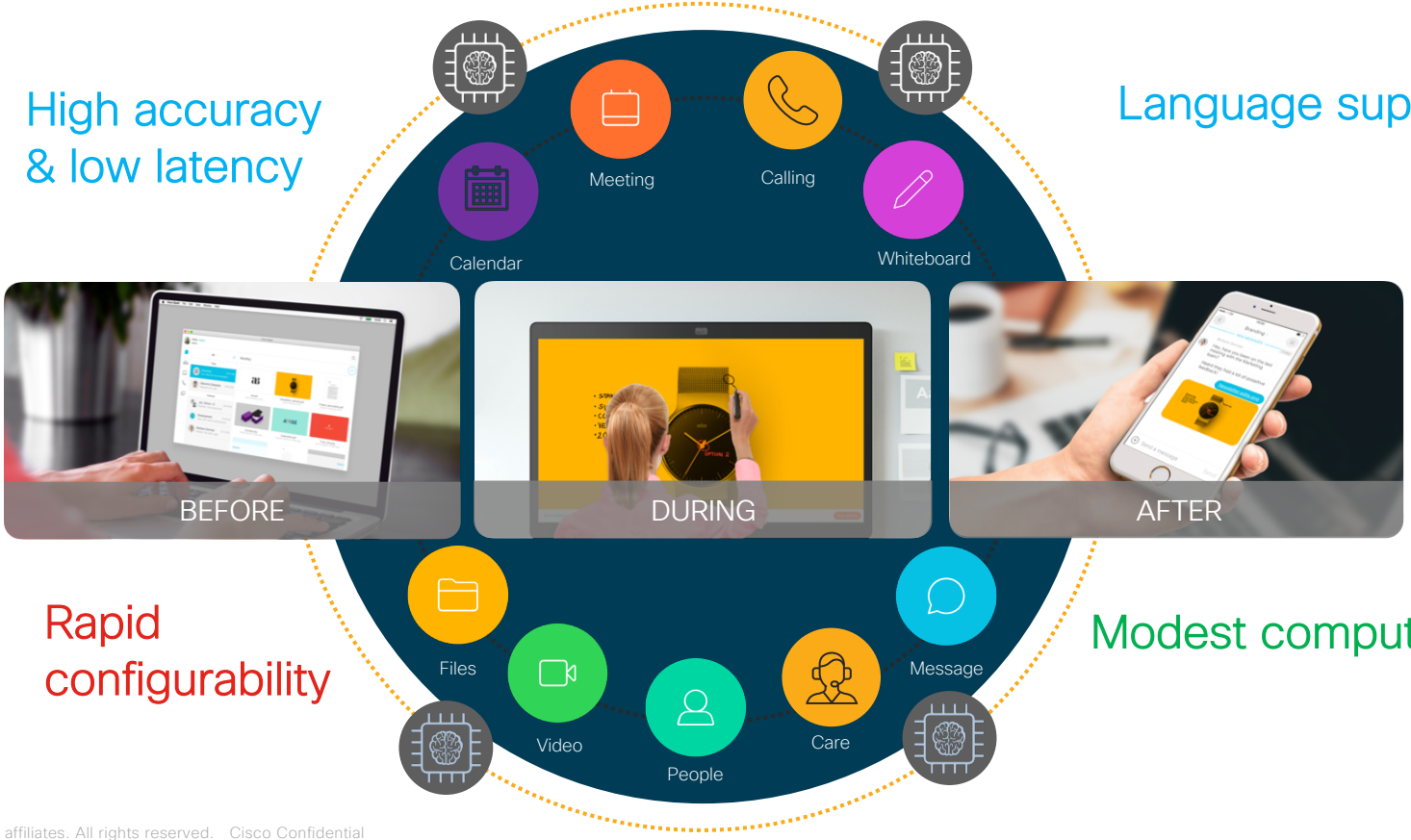
A better way



Enabling speech recognition's proliferation

High accuracy
& low latency

Language support



Rapid
configurability

Modest compute


Thank You!