Deep Learning in Alexa

AI Frontiers

Santa Clara, CA

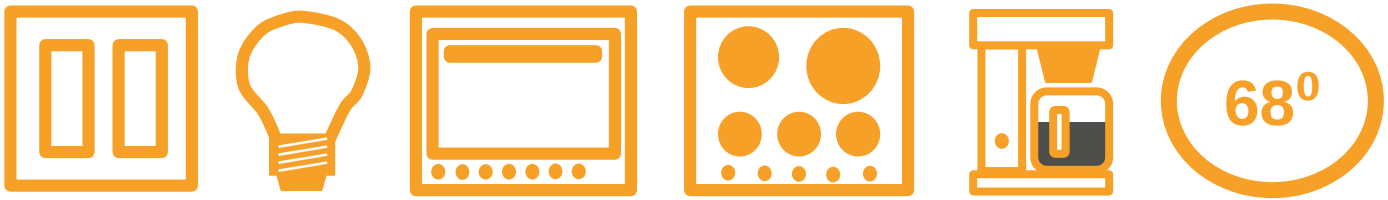
January 11, 2017

Nikko Strom, Sr. Principal Scientist, Alexa Machine Learning

Alexa’s growing family

Alexa in the wild

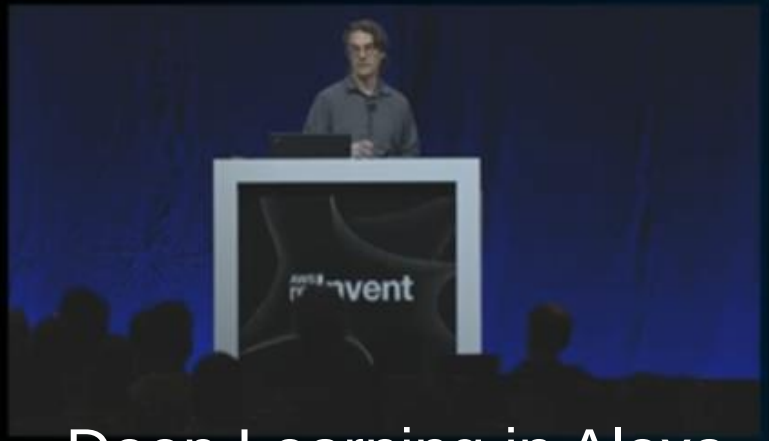
Alexa’s friends



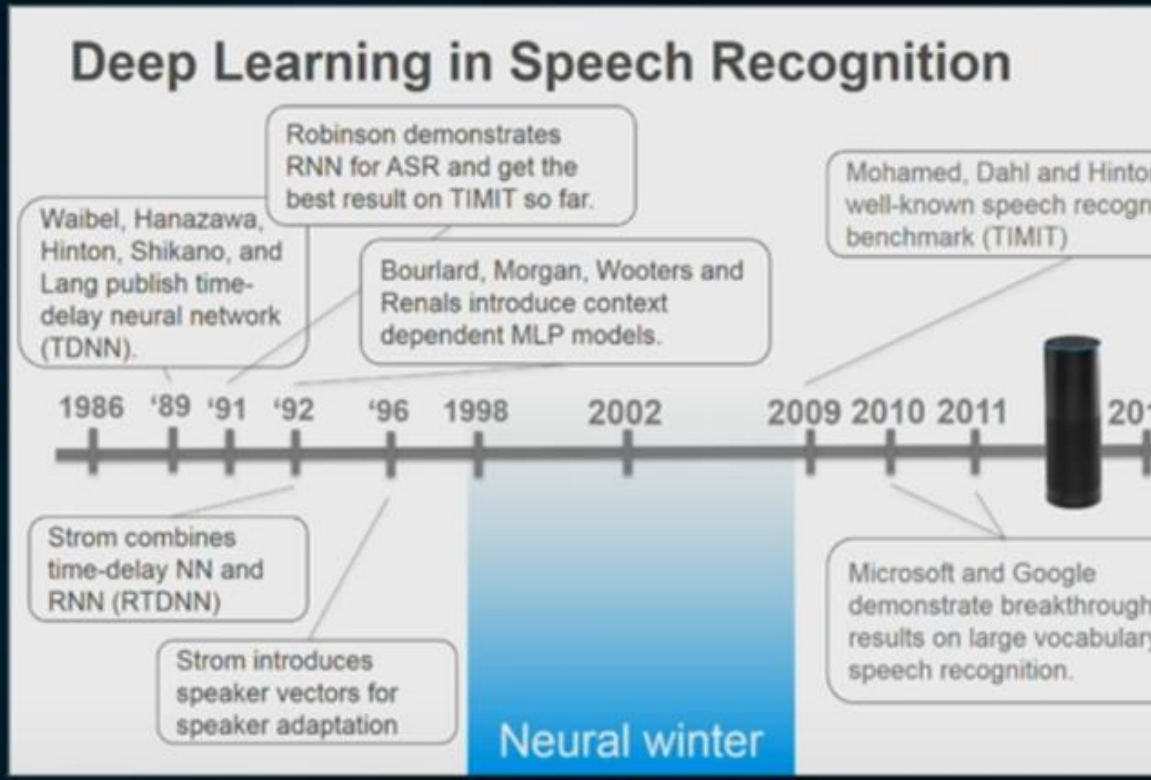
Alexa’s skills

Deep Learning at scale

Longer-form talk at AWS re:Invent 2016



Deep Learning in Alexa (MAC202)



https://www.**youtube**.com/watch?v=TYRckcVm4WE

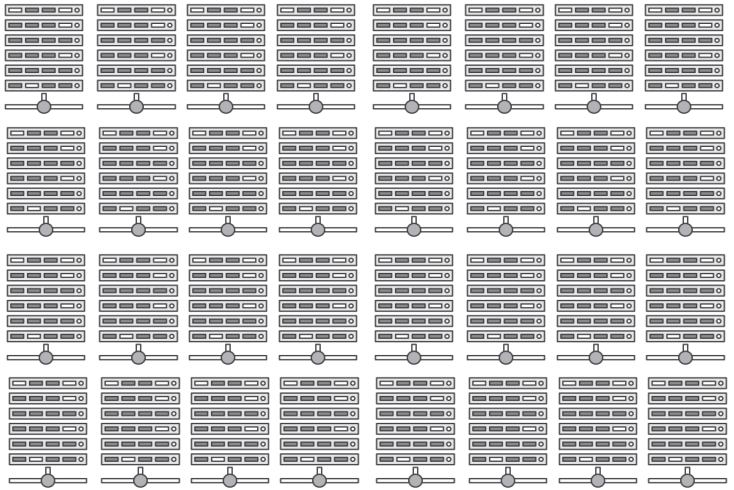
Speech data

16 years = 140,160 hours

≈14,016 hours of speech

Large-scale distributed training

Thousands of hours of speech training data stored in S3



Up to 80 EC2 g2.2xlarge GPU instances working in sync to train a model

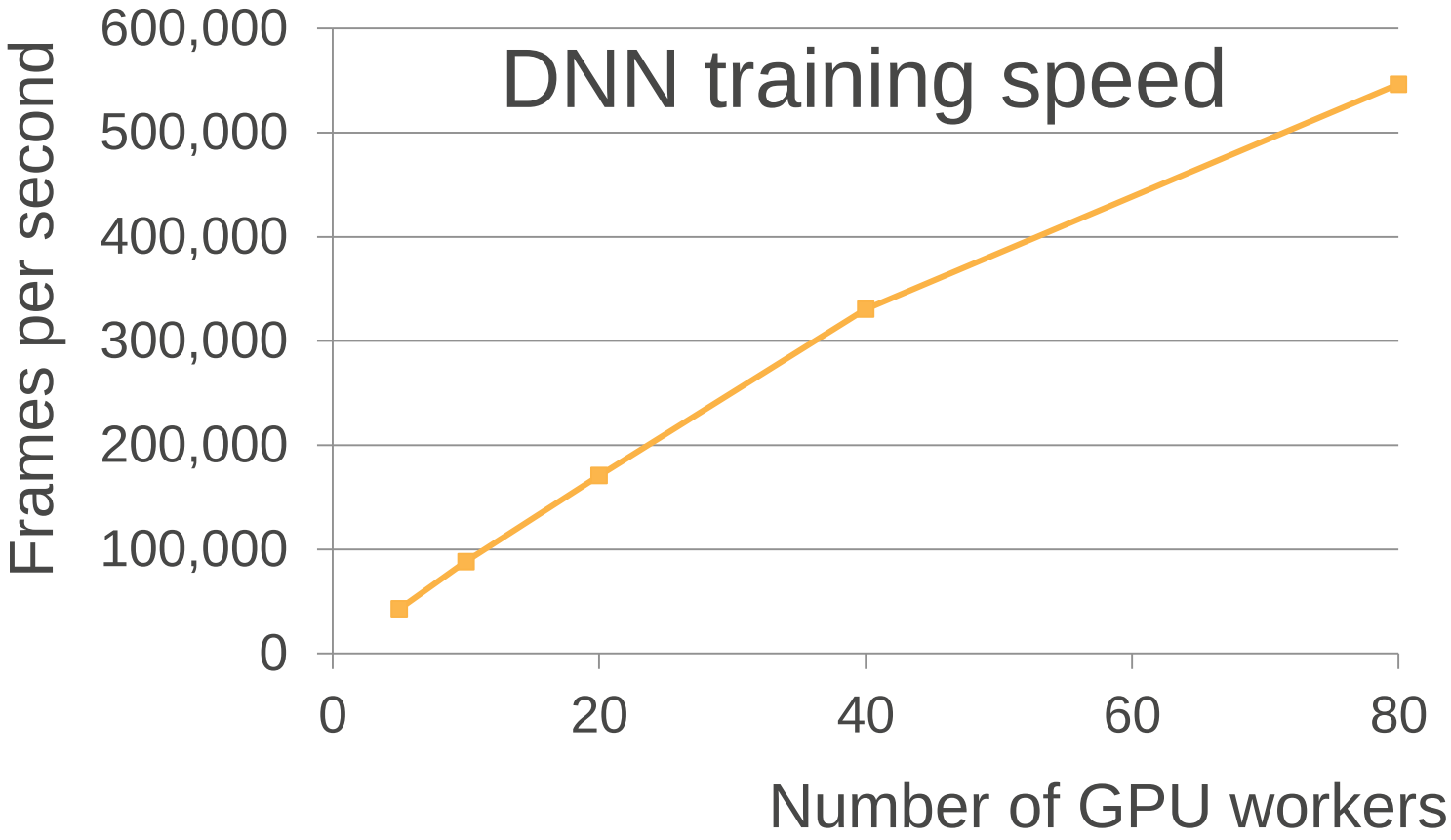
Large-scale distributed training

All nodes must communicate updates to the model to all other nodes.

GPUs compute model updates fast – Think updates per second

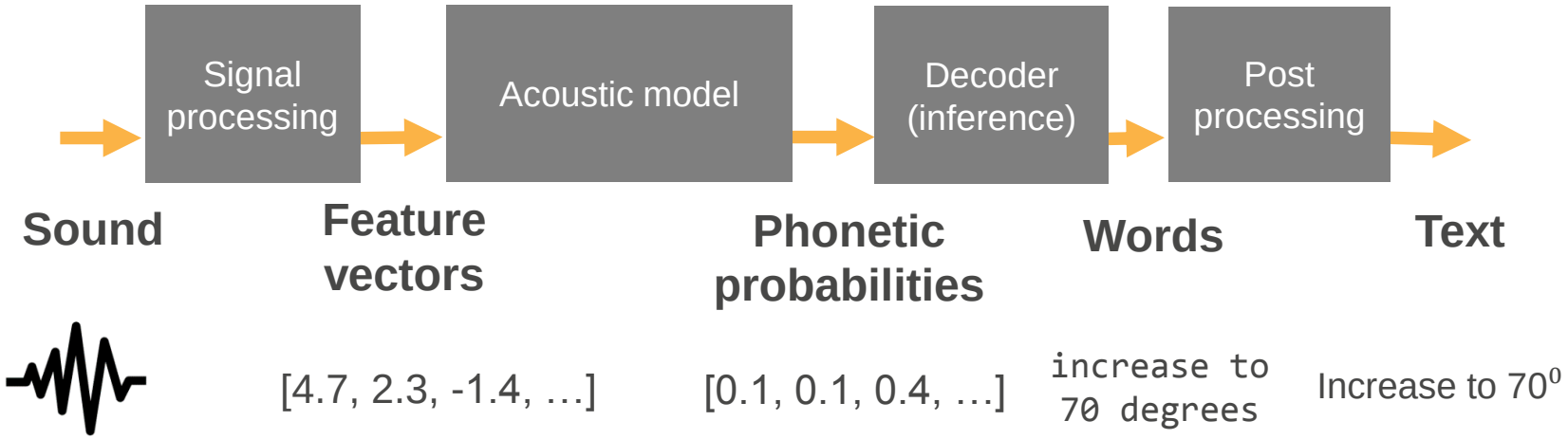
A model update is hundreds of MB

Strom, Nikko. "Scalable Distributed DNN Training using Commodity GPU Cloud Computing." *INTERSPEECH*. Vol. 7. 2015.

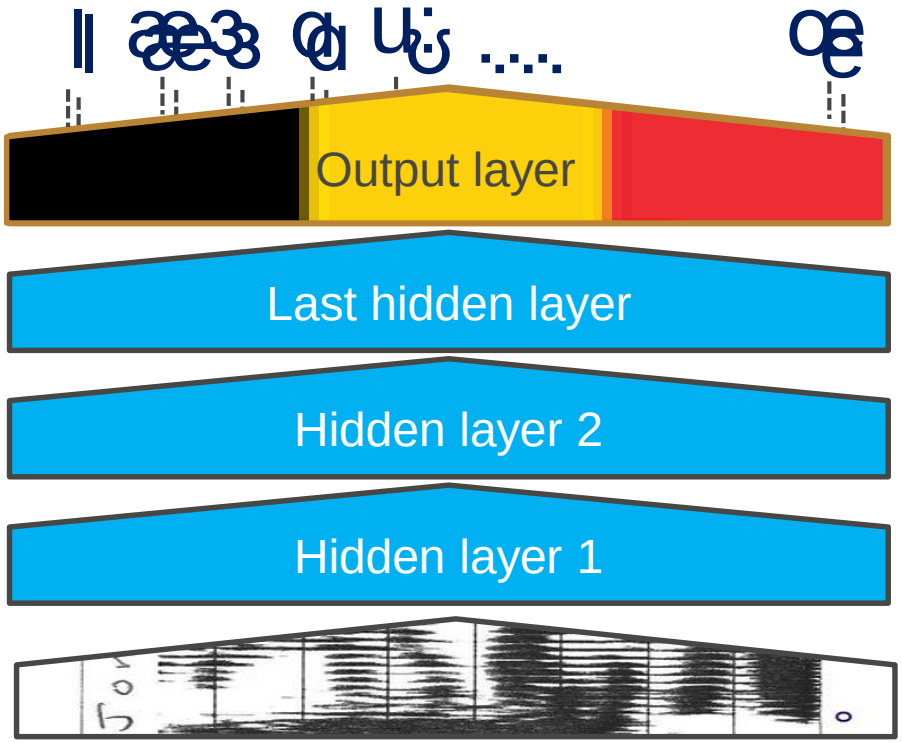


Speech Recognition

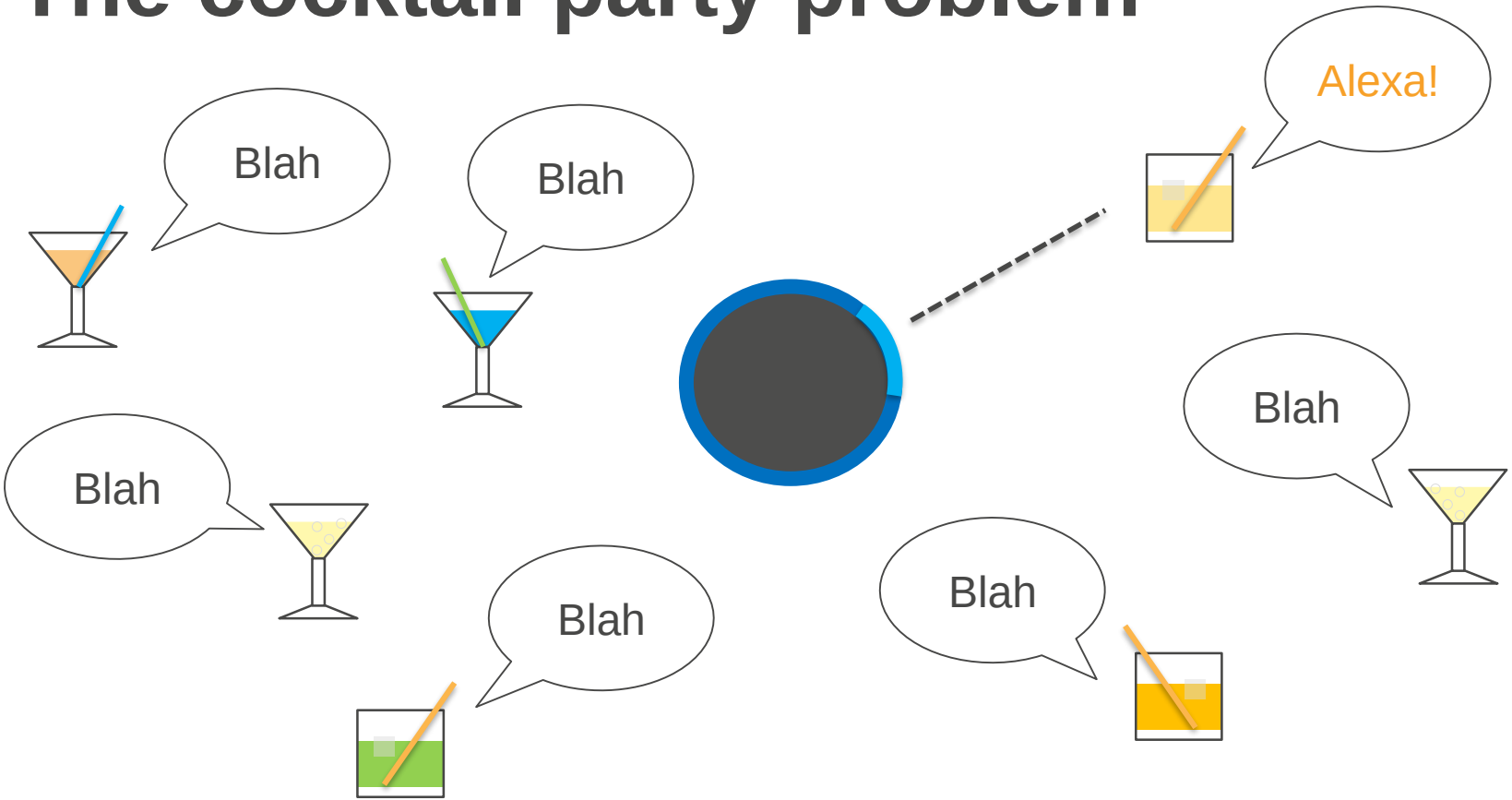
Speech recognition



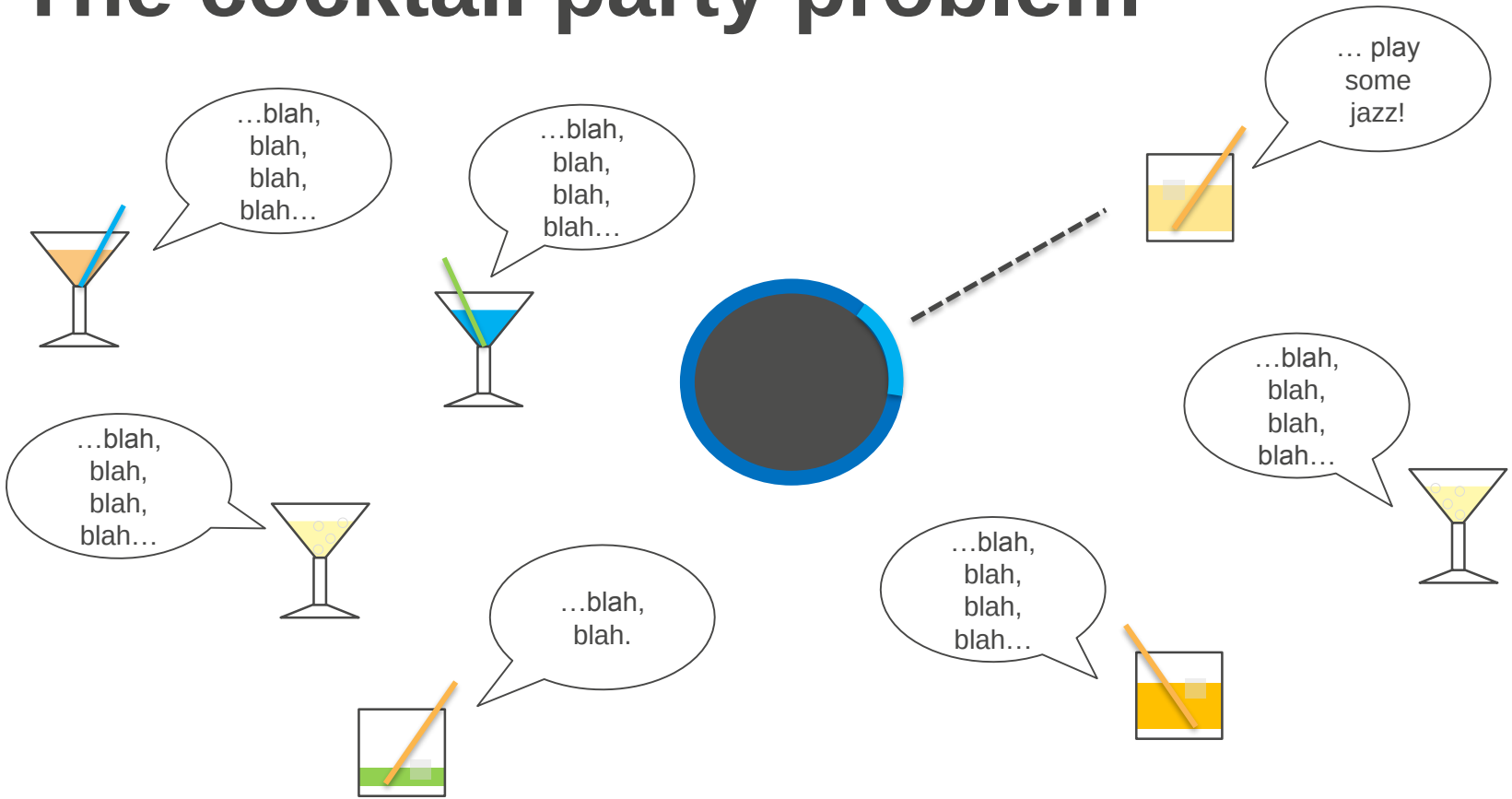
Transfer learning from English to German



The cocktail party problem



The cocktail party problem

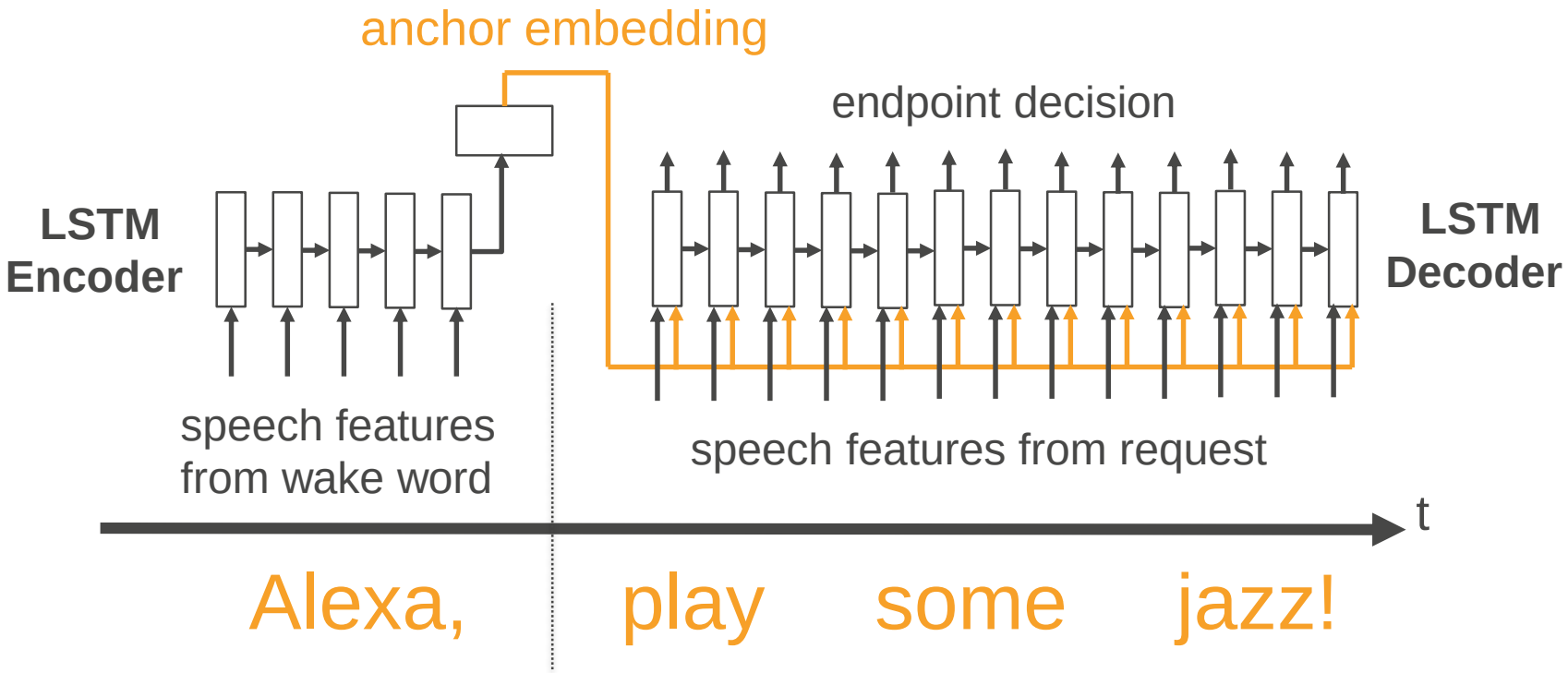


Anchored speech detection

Wake word	Request
“Anchor”	Speech consistent with anchor
Encoder	Decoder

Roland Maas, Sree Hari Krishnan Parthasarathi, Brian King, Ruitong Huang, Björn Hoffmeister. “Anchored Speech Detection.” *INTERSPEECH*. 2016.

Anchored speech detection



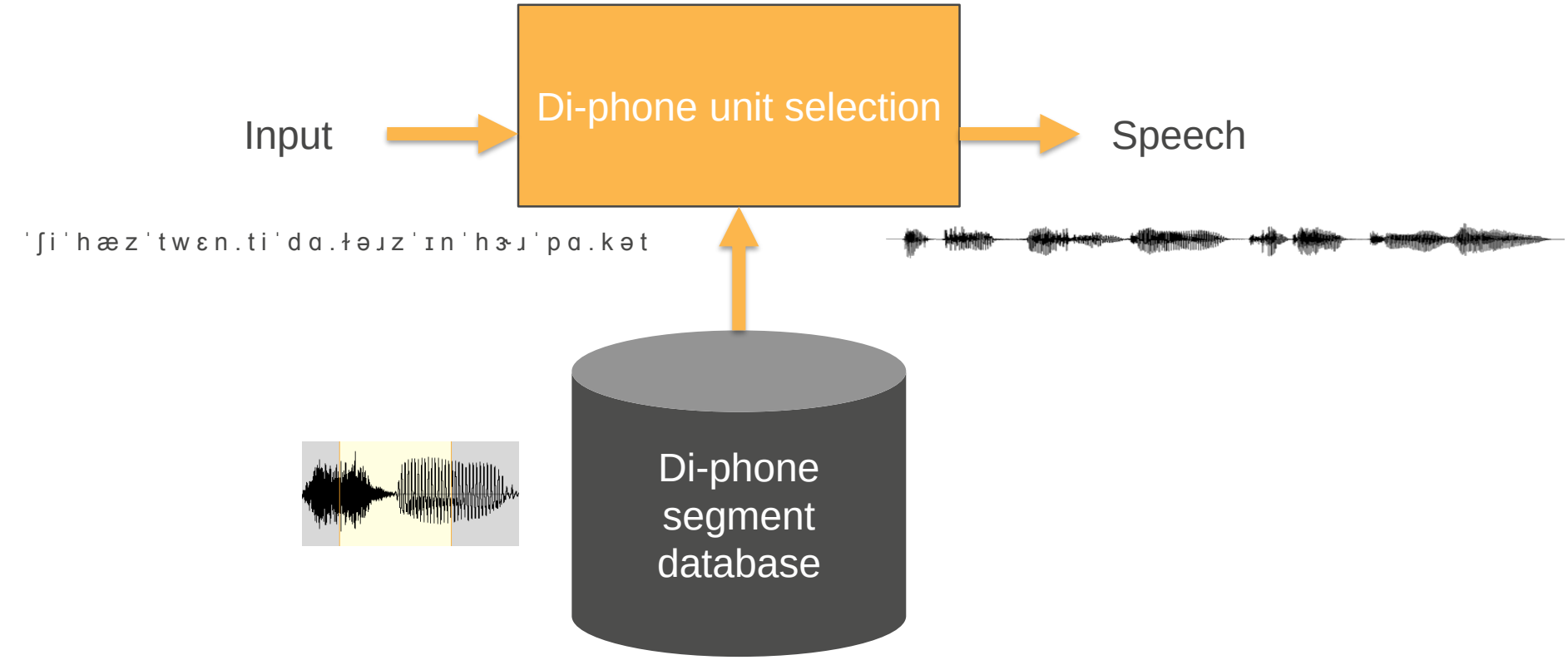
Roland Maas, Sree Hari Krishnan Parthasarathi, Brian King, Ruitong Huang, Björn Hoffmeister. “Anchored Speech Detection.” *INTERSPEECH*. 2016.


Speech
synthesis

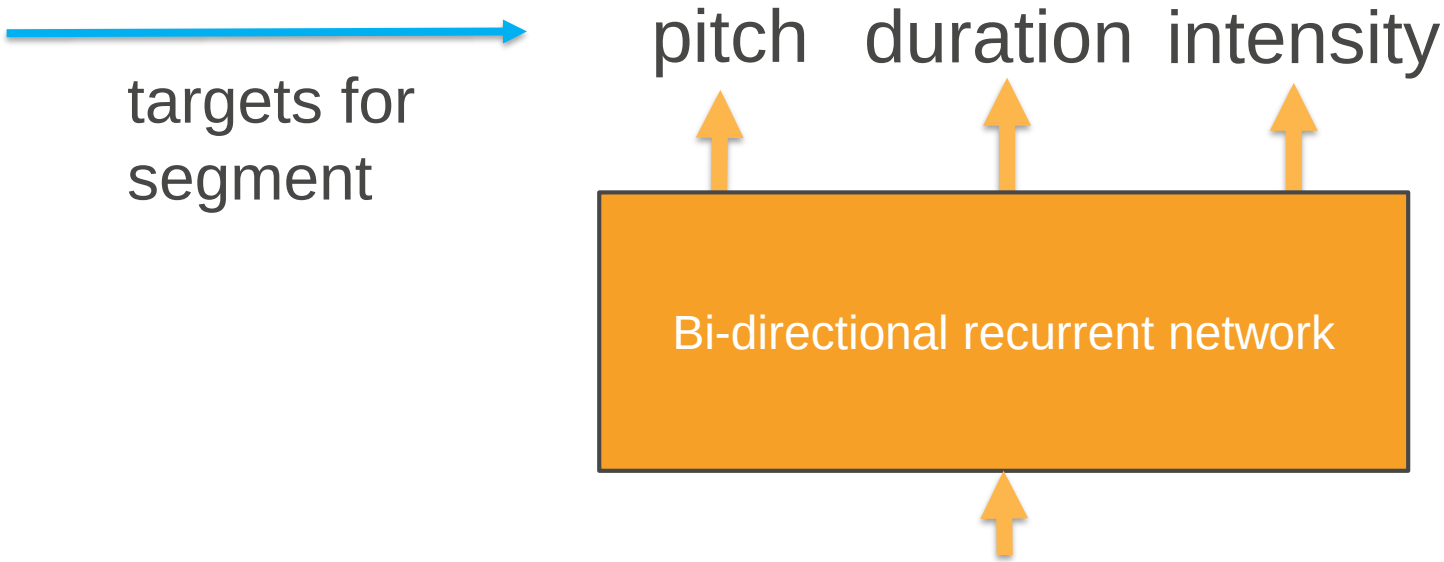
Speech synthesis

Text	She has 20$ in her pocket.
Text normalization	she has twenty dollars in her pocket
Grapheme-to-phoneme conversion	ˈ ʃ i ˈ h æ z ˈ t w ɛ n . t i ˈ d ɑ . ɬ ə ɻ z ˈ ɪ n ˈ h ɝ ɻ ˈ p ɑ . k ə t
Waveform generation	
Speech	

Concatenative synthesis



Prosody for natural sounding reading



- **Phonetic features**
- **Linguistic features**
- **Semantic word vectors**

Long-form example

“Over a lunch of diet cokes and lobster salad one balmy fall day in Boston, Joseph Martin, the genial, white-haired, former dean of Harvard medical school, told me how many hours of pain education Harvard med students get during four years of medical school.”

Before

After

Thank you!