MIS|TI™ PRESENTS

InfoSec World

Conference & Expo 2018

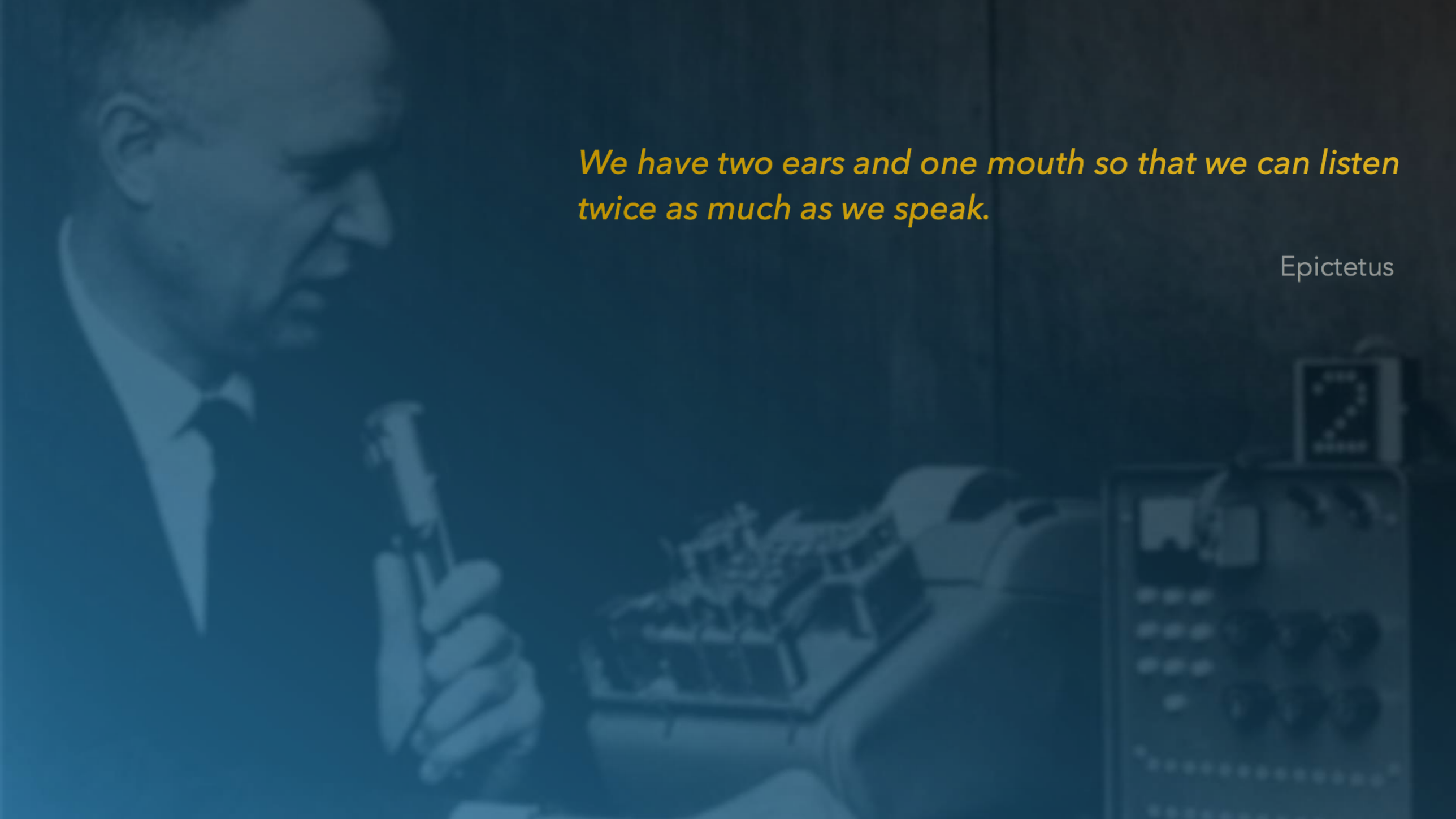
OMNICHANNEL ATTACKS

How Voice Could be the Next Frontier for Attackers

Sam Rehman

CTO, Arxan Technologies

@codemonkeysam


We have two ears and one mouth so that we can listen twice as much as we speak.
Epictetus

WHY VOICE?

An ambient medium

A more natural interface

Far more pervasive end-points

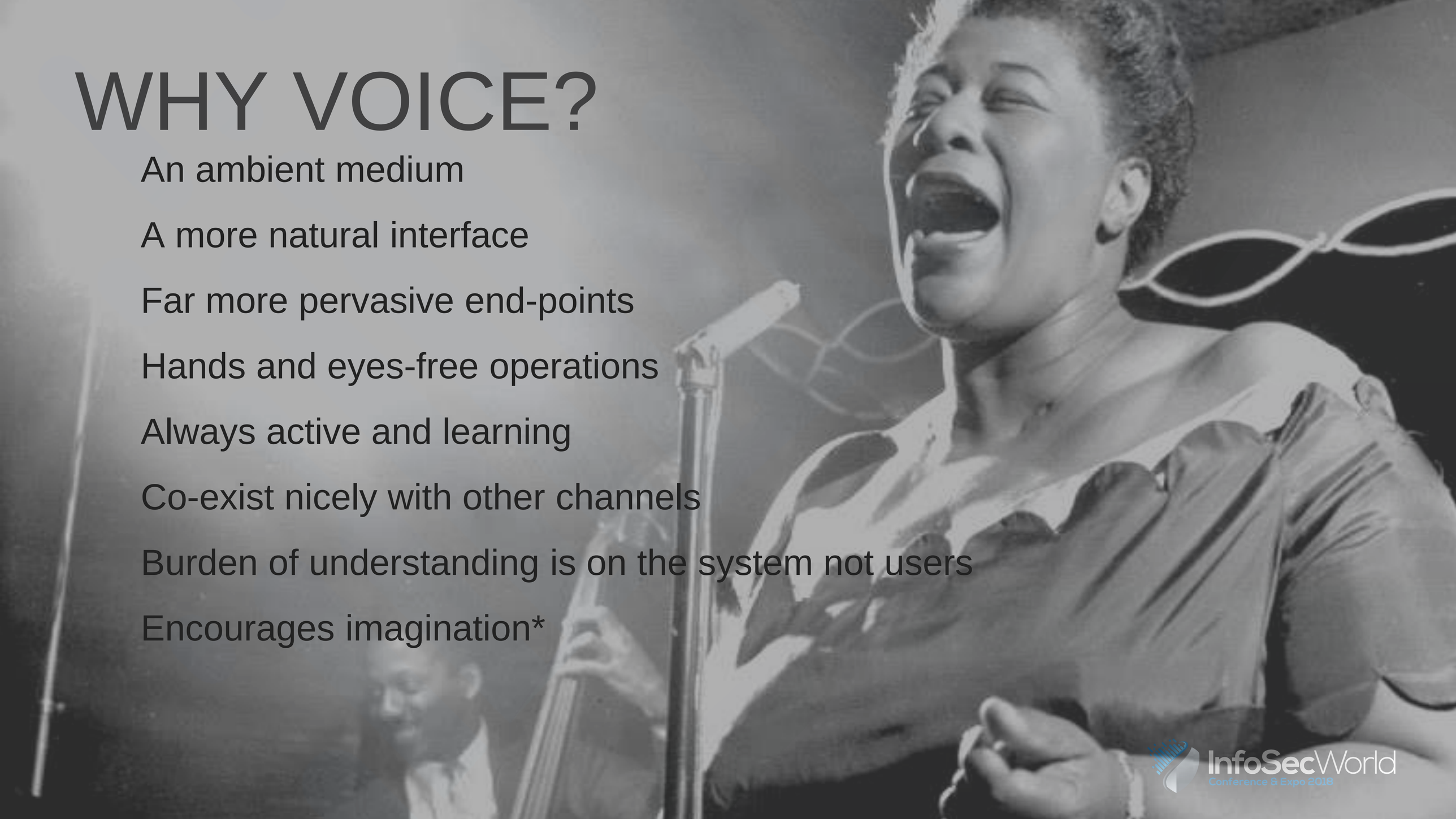
Hands and eyes-free operations

Always active and learning

Co-exist nicely with other channels

Burden of understanding is on the system not users

Encourages imagination*



WHAT IS A CONVERSATION

An exchange of “ideas” by spoken words
between two or more “learning systems”

Speaking
Listening, facial expressions, gestures, tone, experimentation, tempo, etc

Listening
Attention, memory, context, historical data, identity, emotions, etc.

A LITTLE HISTORY

1952 Aubrey recognizes single digits

1962 IBM Shoebox undstands 16 words

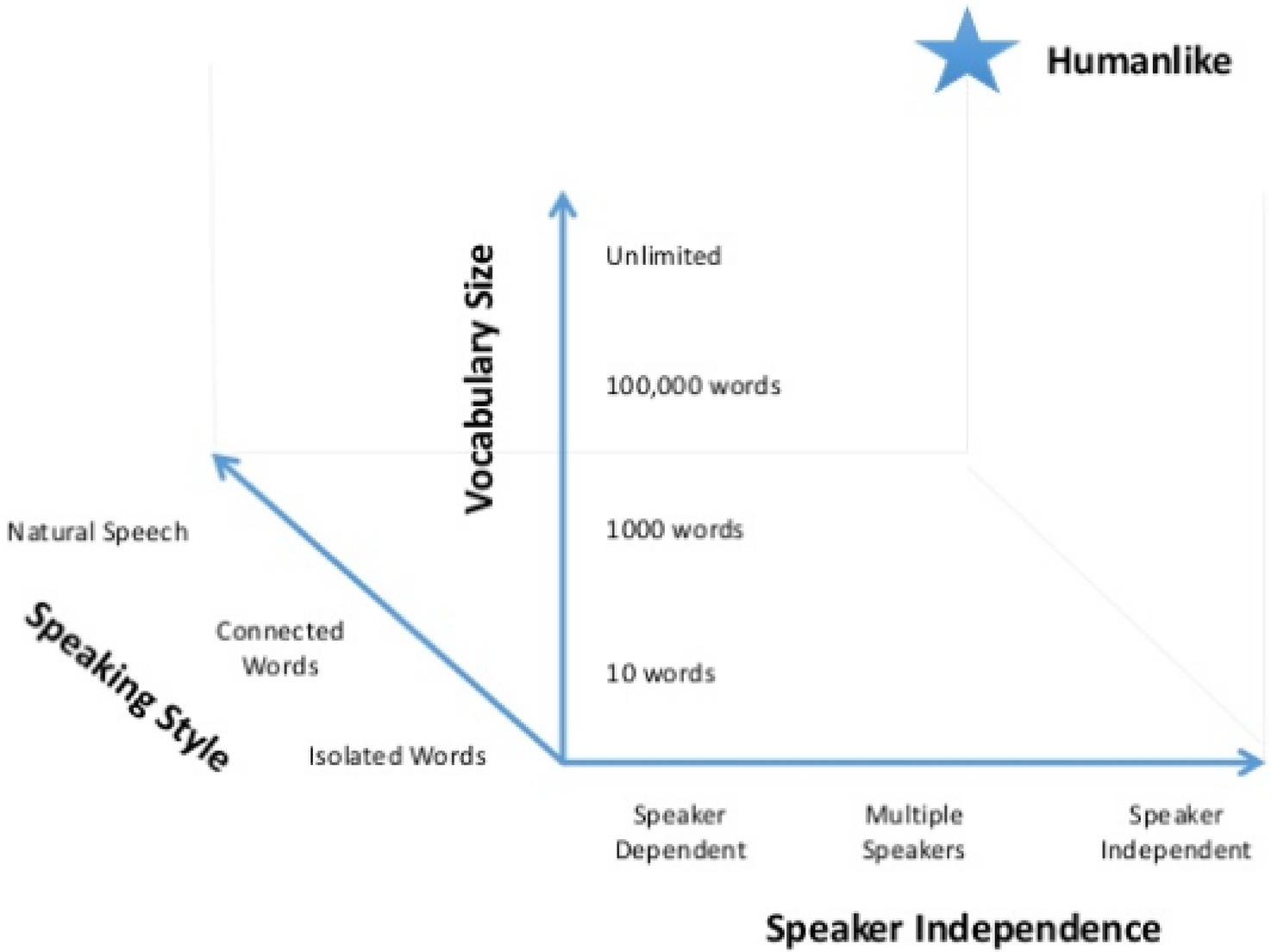
1971 “Harpy Speech” recognize 1011 words

1986 IBM Targora moves towards phonemes predication

2006 NSA starts isolating key words with recorded speech

2008 Google launches voice search app on mobile devices

2011 Apple announces Siri



TOWARDS A FAR MORE NATURAL EXPERIENCE

VOICE FIRST ERA HAS ALREADY STARTED

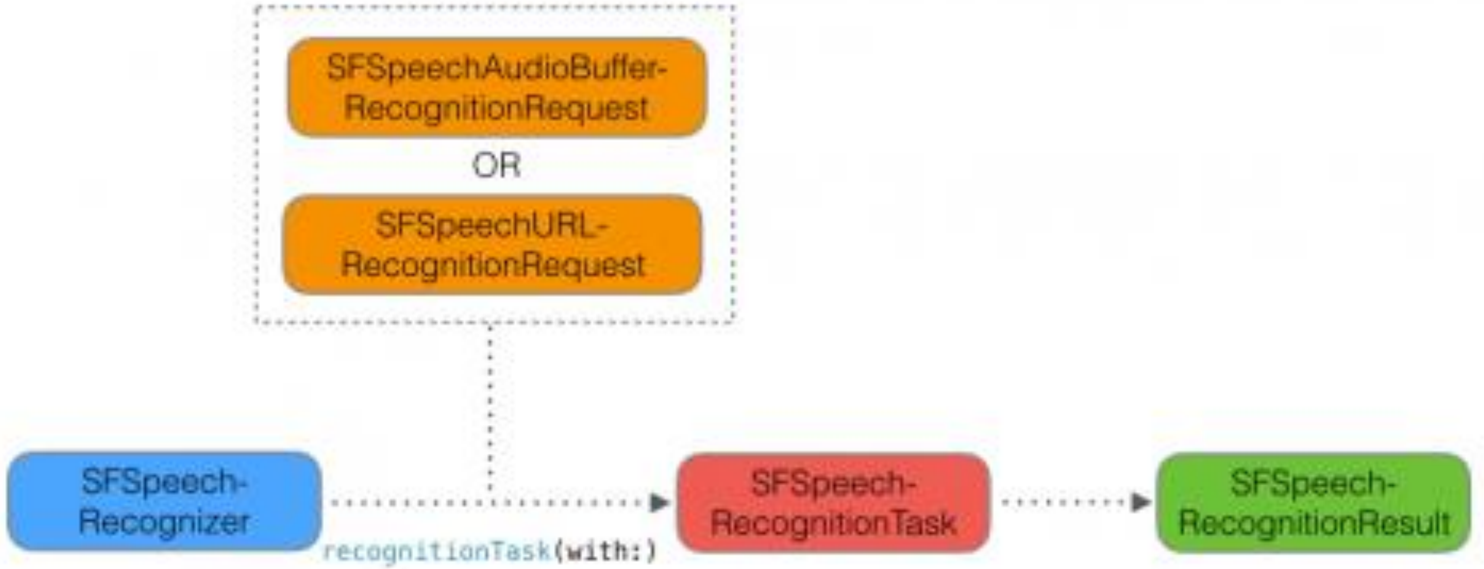
Alexa in 4% US household already
Siri handles over 2 billion commands a week
20% of Google search on Android are by voice
Growing choices of endpoints and platforms:
Siri, Alexa, Google Now, Cortana, Jibo, UBI, Assistant.ai, etc.

HIGHLY EXTENSIBLE AND INTEGRATED

e.g: Alexa Skills processing

ASR AND MOBILE/IOT - PERFECT MARRIAGE

Requires minimum physical footprint, perfect for IoT and wearables
Most mobile devices now can handle offline speech recognition
Both Apple (SFSpeechRecognizer) and Android (SpeechRecognizer) now opened up their layered APIs



FROM VIBRATIONS TO MEANING

ACOUSTICS

ASR PHONETICS

MORPHOLOGY

LEXICON

SYNTAX

NLU

SEMANTICS

WORLD KNOWLEDGE

ATTACK SURFACES

Mimicry

Speech Synthesis

Voice Conversion

Replay and Sampling

Inaudible Attacks

Etc…

ACOUSTICS

ASR PHONETICS

MORPHOLOGY

LEXICON

SYNTAX

NLU

SEMANTICS

WORLD KNOWLEDGE

Voice/data injection

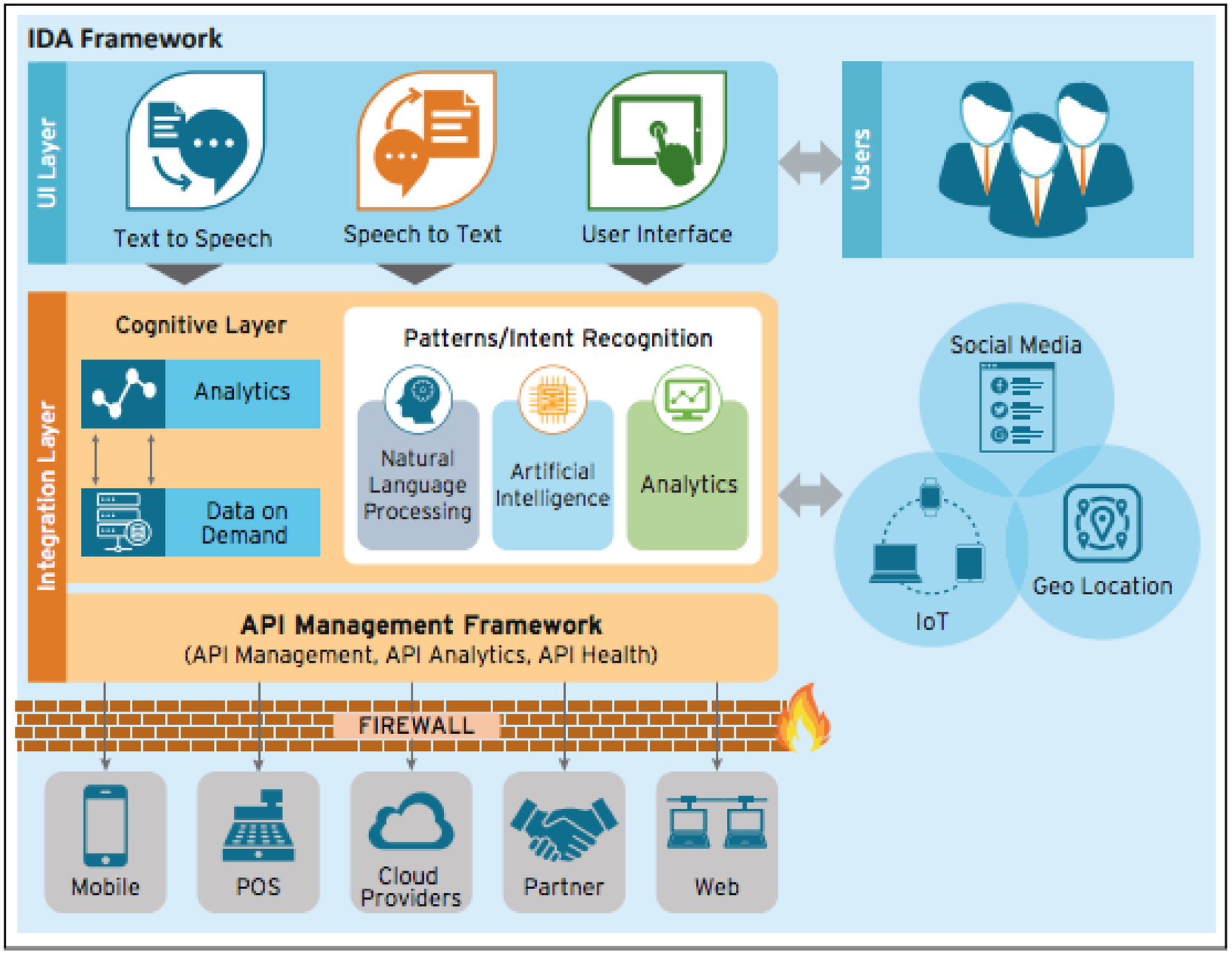
Piggy backing

Impersonation

Conversation exploits

Etc…

How Platforms Can Embrace IDA Evolution



INTELLIGENT ASSISTANT

SECURITY CONCERNS

SPEECH IS AMBIGUOUS AND HARD TO SECURE

If true or false is now a gray scale, then integrity is substantially harder to enforce

IN ORDER TO BE USEFUL, AI HAS TO KNOW TOO MUCH

State, context, and historical data all has to processed or stored somewhere

INAUDIBLE ATTACKS (“DOLPHIN”)

- Translate voice commands shifted into inaudible frequency to induce nefarious actions
- Attacks types for both activation and recognition
- Message snippets can be embedded in other audio channels (radio, tv, broadcast, etc.)
- In most cases requires unlocked devices

Table 3: Experiment devices, systems, and results. The examined attacks include *recognition* (executing control commands when the SR systems are manually activated) and *activation* (when the SR systems are unactivated). The modulation parameters and maximum attack distances are acquired for recognition attacks in an office environment with a background noise of 55 dB SPL on average.

Manuf.	Model	OS/Ver.	SR System	Attacks		Modulation Parameters		Max Dist. (cm)	
				Recog.	Activ.	f_c (kHz) & [Prime f_c] ‡	Depth	Recog.	Activ.
Apple	iPhone 4s	iOS 9.3.5	Siri	√	√	20–42 [27.9]	≥ 9%	175	110
Apple	iPhone 5s	iOS 10.0.2	Siri	√	√	24.1 26.2 27 29.3 [24.1]	100%	7.5	10
Apple	iPhone SE	iOS 10.3.1	Siri	√	√	22–28 33 [22.6]	≥ 47%	30	25
			Chrome	√	N/A	22–26 28 [22.6]	≥ 37%	16	N/A
Apple	iPhone SE †	iOS 10.3.2	Siri	√	√	21–29 31 33 [22.4]	≥ 43%	21	24
Apple	iPhone 6s ⋆	iOS 10.2.1	Siri	√	√	26 [26]	100%	4	12
Apple	iPhone 6 Plus ⋆	iOS 10.3.1	Siri	×	√	— [24]	—	—	2
Apple	iPhone 7 Plus ⋆	iOS 10.3.1	Siri	√	√	21 24-29 [25.3]	≥ 50%	18	12
Apple	watch	watchOS 3.1	Siri	√	√	20–37 [22.3]	≥ 5%	111	164
Apple	iPad mini 4	iOS 10.2.1	Siri	√	√	22–40 [28.8]	≥ 25%	91.6	50.5
Apple	MacBook	macOS Sierra	Siri	√	N/A	20-22 24-25 27-37 39 [22.8]	≥ 76%	31	N/A
LG	Nexus 5X	Android 7.1.1	Google Now	√	√	30.7 [30.7]	100%	6	11
Asus	Nexus 7	Android 6.0.1	Google Now	√	√	24–39 [24.1]	≥ 5%	88	87
Samsung	Galaxy S6 edge	Android 6.0.1	S Voice	√	√	20–38 [28.4]	≥ 17%	36.1	56.2
Huawei	Honor 7	Android 6.0	HiVoice	√	√	29–37 [29.5]	≥ 17%	13	14
Lenovo	ThinkPad T440p	Windows 10	Cortana	√	√	23.4–29 [23.6]	≥ 35%	58	8
Amazon	Echo ⋆	5589	Alexa	√	√	20-21 23-31 33-34 [24]	≥ 20%	165	165
Audi	Q3	N/A	N/A	√	N/A	21–23 [22]	100%	10	N/A

‡ Prime f_c is the carrier wave frequency that exhibits highest baseband amplitude after demodulation. — No result
† Another iPhone SE with identical technical spec.
⋆ Experimented with the front/top microphones on devices.

PRIVACY

Always on

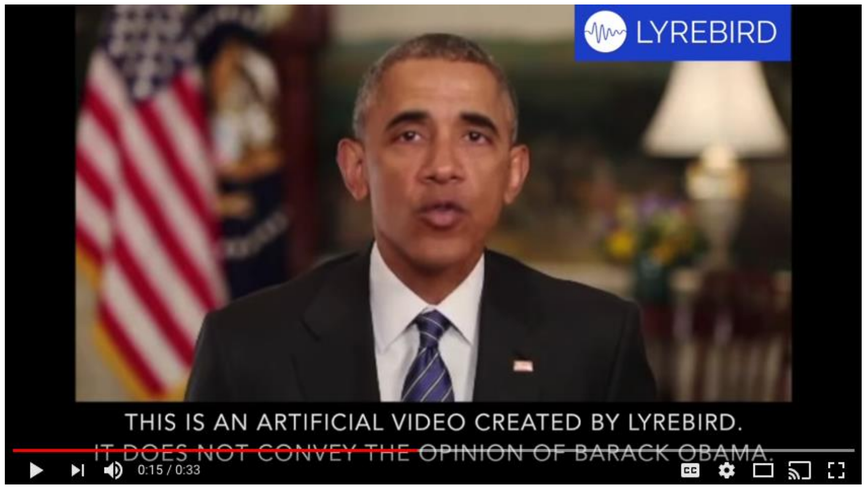
Data stored and cached

Data needed for NLU

Large interface endpoints

Intrinsically omni directional

VOICE SPOOFING



- Replay Attacks: Everyone is sampling!
- Voice mimicking: voice synthesis technology is advancing in incredible rate
- Slicing and conversion
- Readily available platforms: Lyrebird, Adobe project Voco, Candyvoice, Festvox just to name a few
- Can be used for both speaker recognition (authN/authR) and for countering CSR (Continuous Speech Recognition)
- False acceptance is up to 77%

USER JOURNEY

- One level automatic speaker recognition
- Allow high level of context retrieval
- No way to reset or fixed by persona
- No time to live on context
- Lack of server identification
- Misleading interfaces

Apple ID Sign In Requested

If you have an Apple ID, enter it here. If you don't, or forgot your Apple ID or password, go to appleid.apple.com.

Hold to spell

SPACE a b c d e f g h i j k l m n o p q r s t u v w x y z

1 2 3 4 5 6 7 8 9 0 . _ - @ .com .net .edu

ABC abc #+-

Continue

Understanding commands and conversations
Learning and optimize scope

SPEECH RECOGNITION

SPEAKER RECOGNITION

Identify users automatically
Authentication
Maintain context across sessions
Saves onramp time

EXECUTION AND INTEGRATION

Execution internal and external commands
Share data with external actors
Cross system identities

DON’T SAY ”YES”

- CallerId or simple phrases are extremely easy to spoof
- Used for both speaker verification and covering recorded trails
- Voice prints spidering: automated system trying to get you to say “Yes” or common phrases
- MITM type voice recorders

SOCIAL HACKING AGAINST AI

RSE(reverse social engineering) attacks on learning systems
Relationship hacking and piggy -backing
Staged phishing/whaling/vishing optimization
The other NLP and…
Noise

MIS|TI™ PRESENTS

InfoSec World

Conference & Expo 2018

THANK YOU

PLEASE FILL OUT YOUR EVALUATIONS!

Sam Rehman
CTO, Arxan Technologies
@codemonkeysam