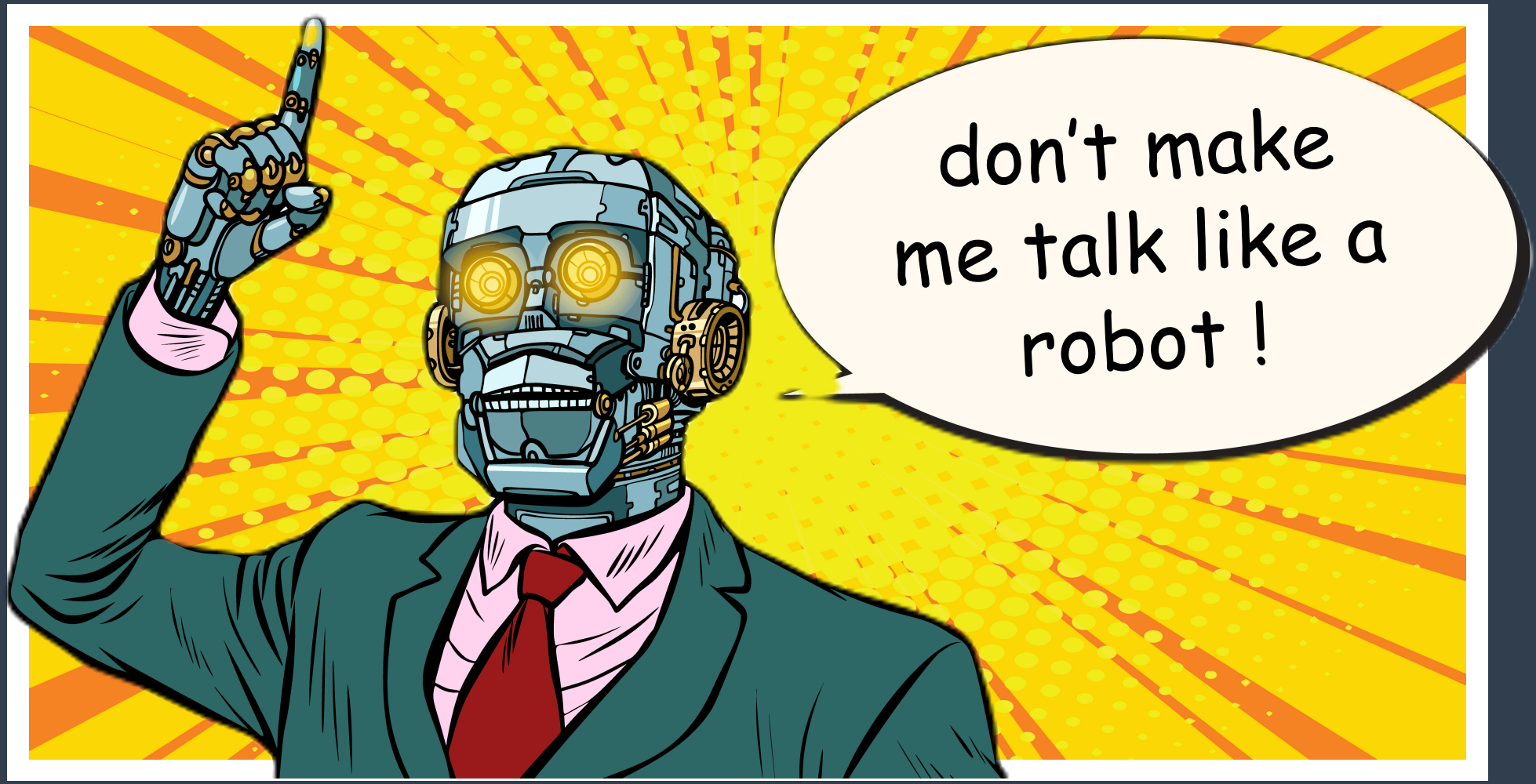
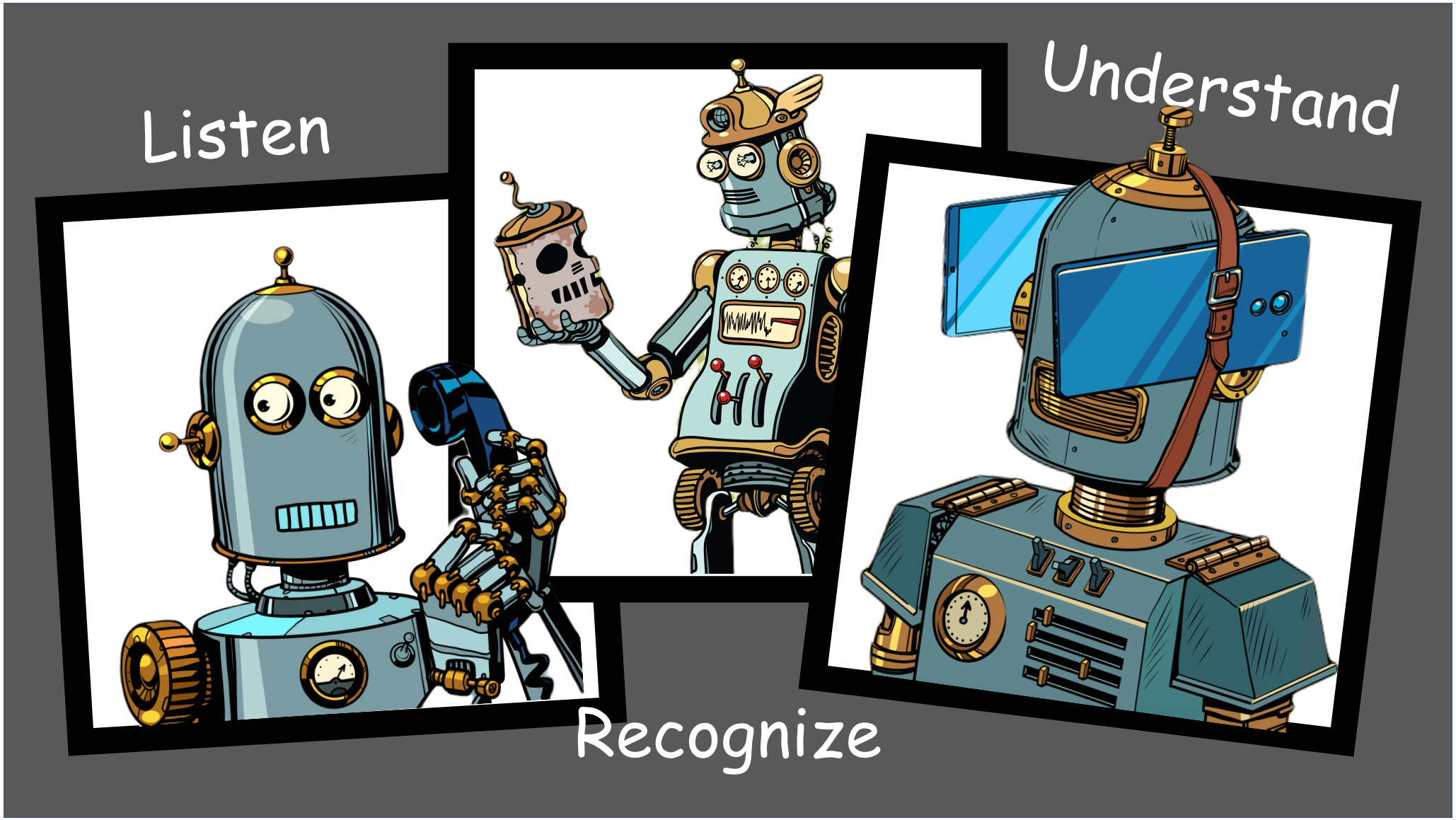
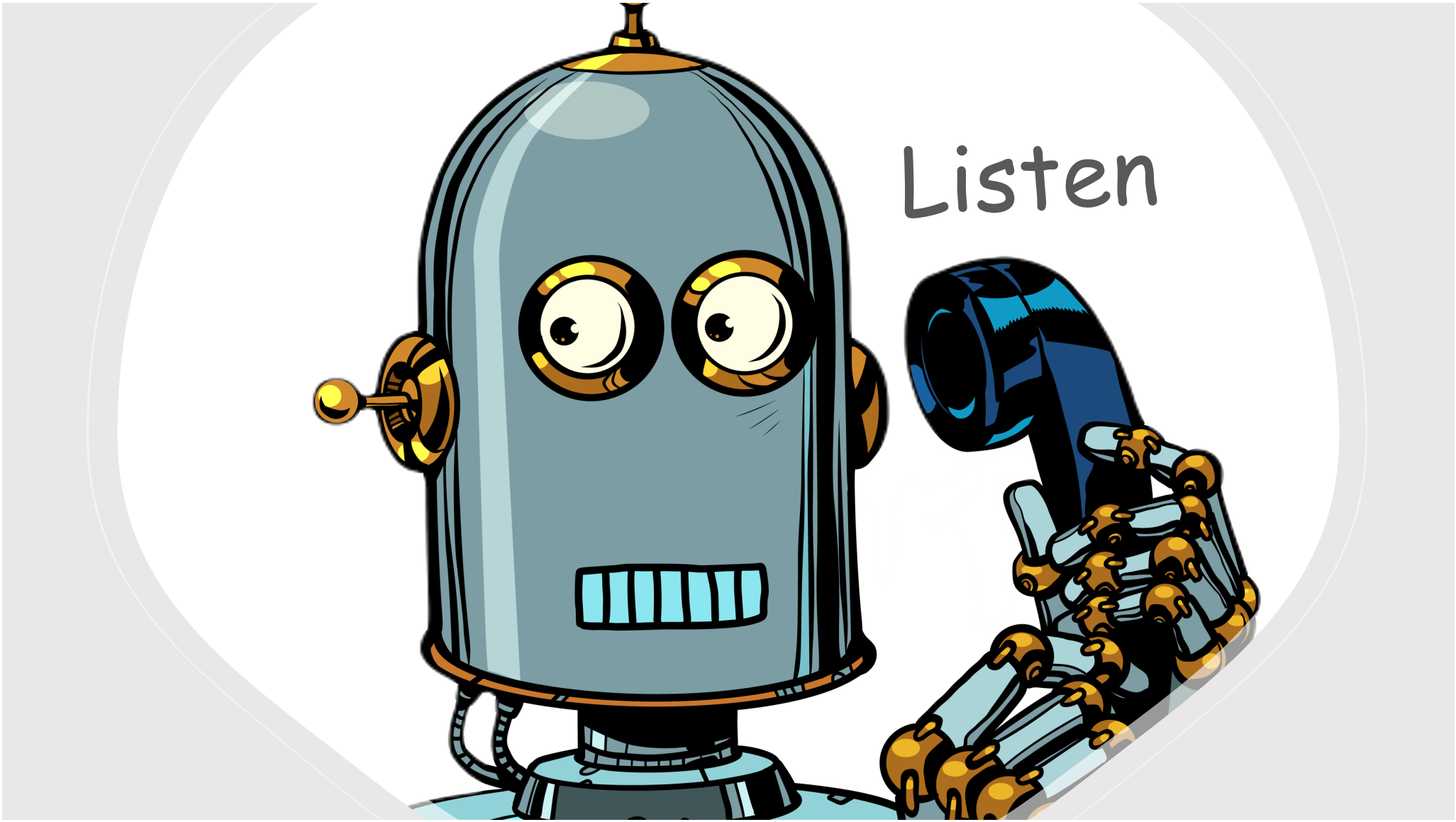
What call centers can teach us about good conversational AI
PROSODICA

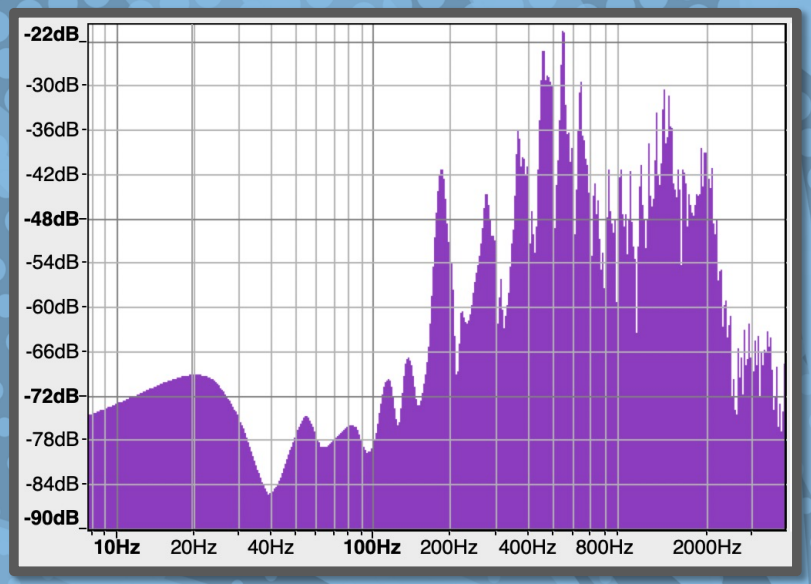
this is the third time I've called...
?!
cancel my account !


don't make
me talk like a
robot !

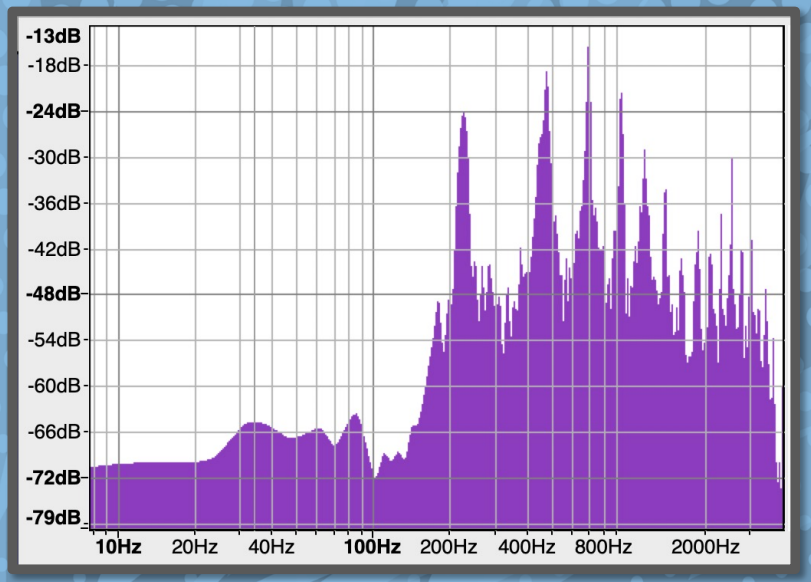

Listen
Understand
Recognize

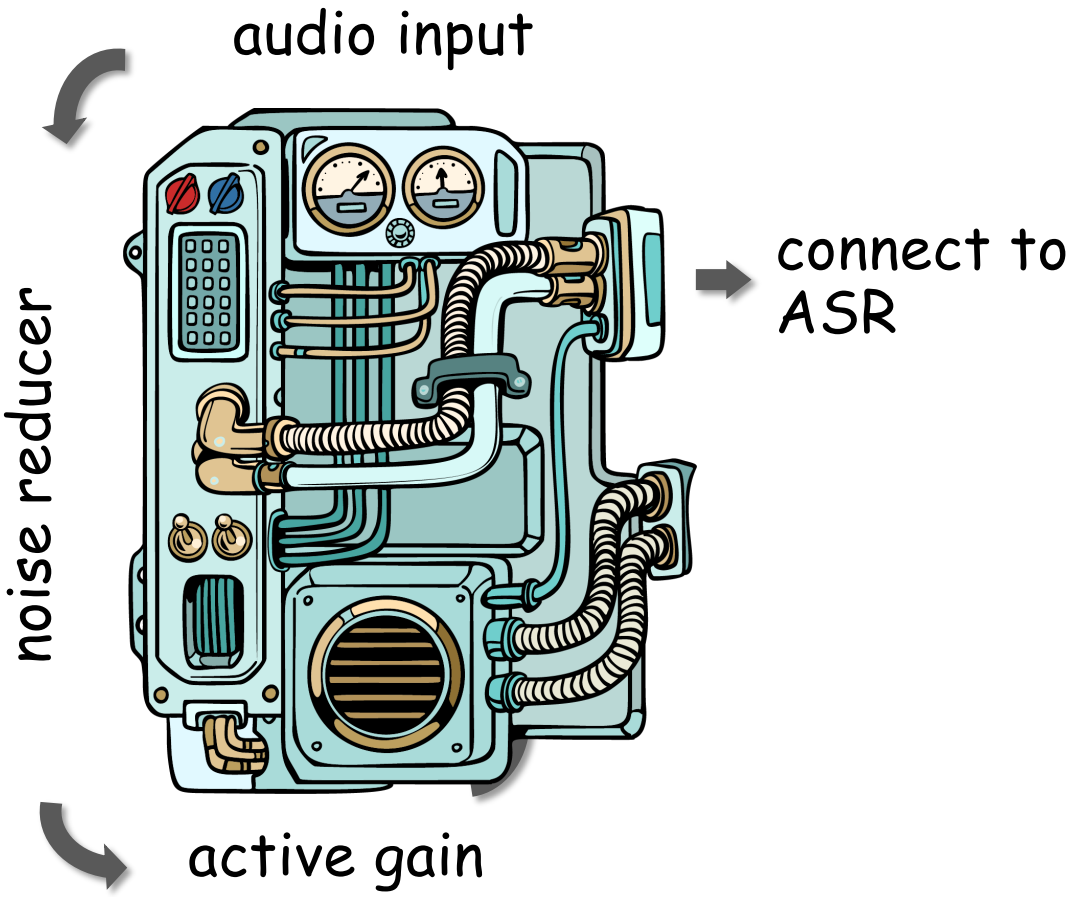

Listen

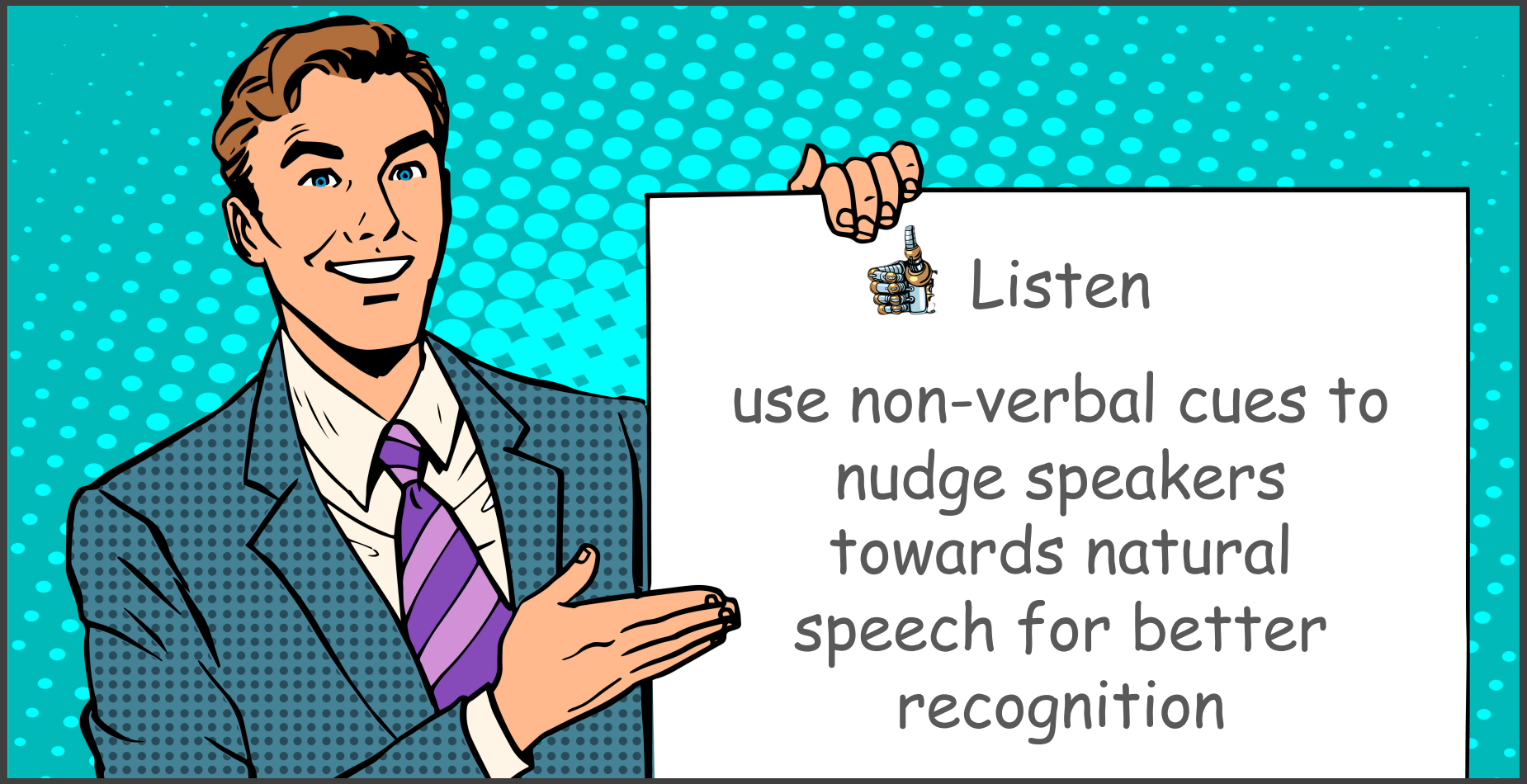
Normal

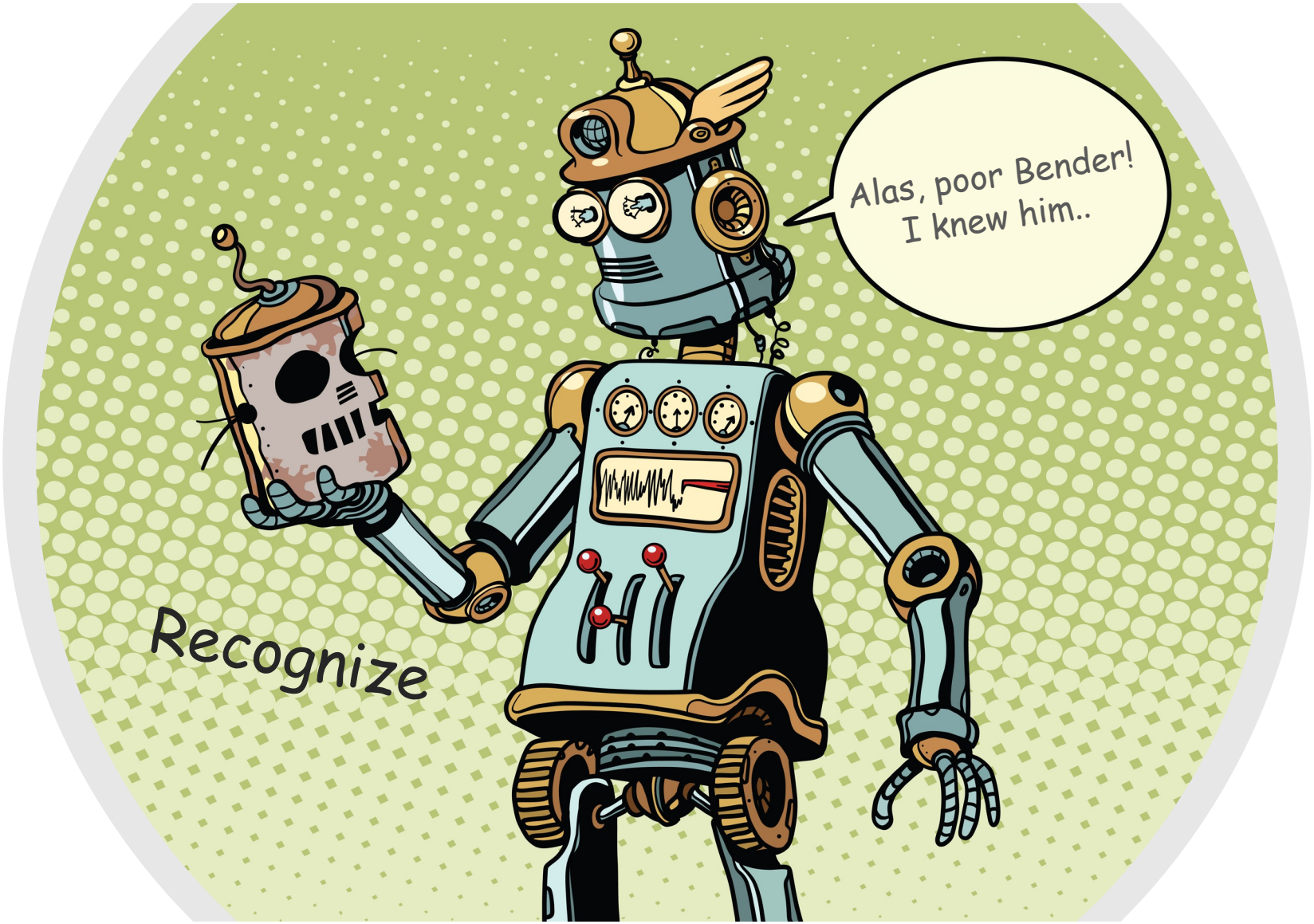


Elevated

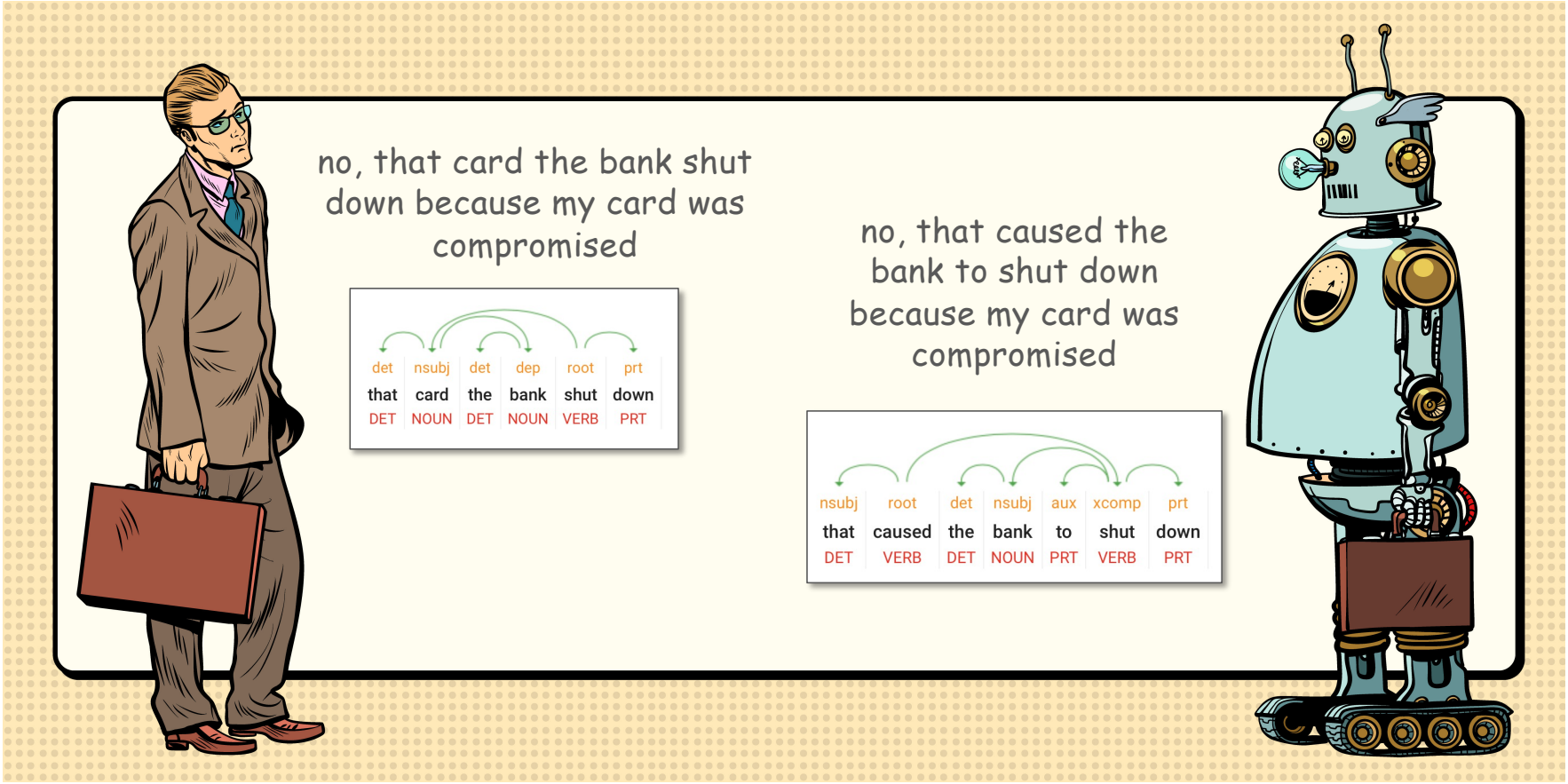


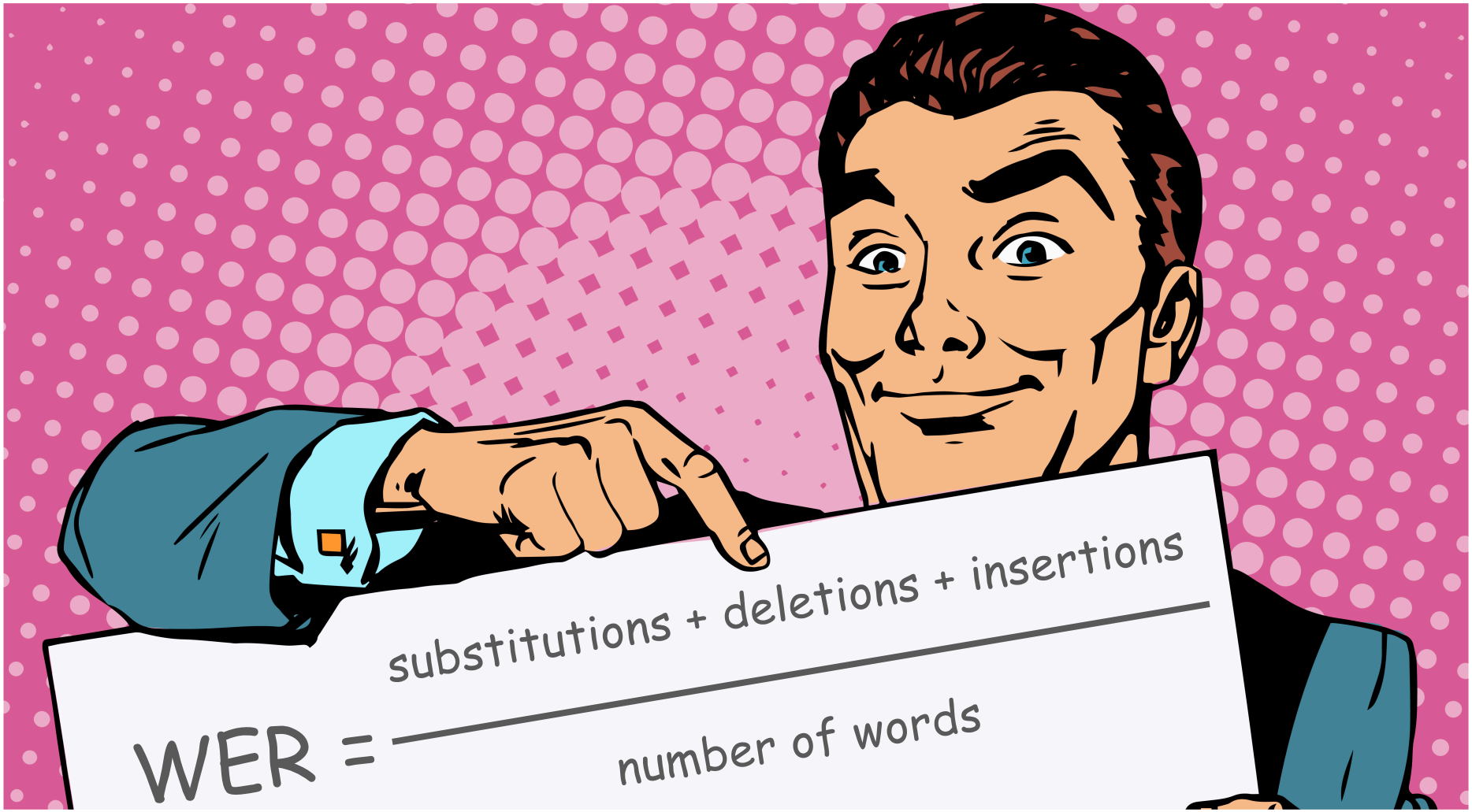

audio input
connect to ASR
active gain
noise reducer

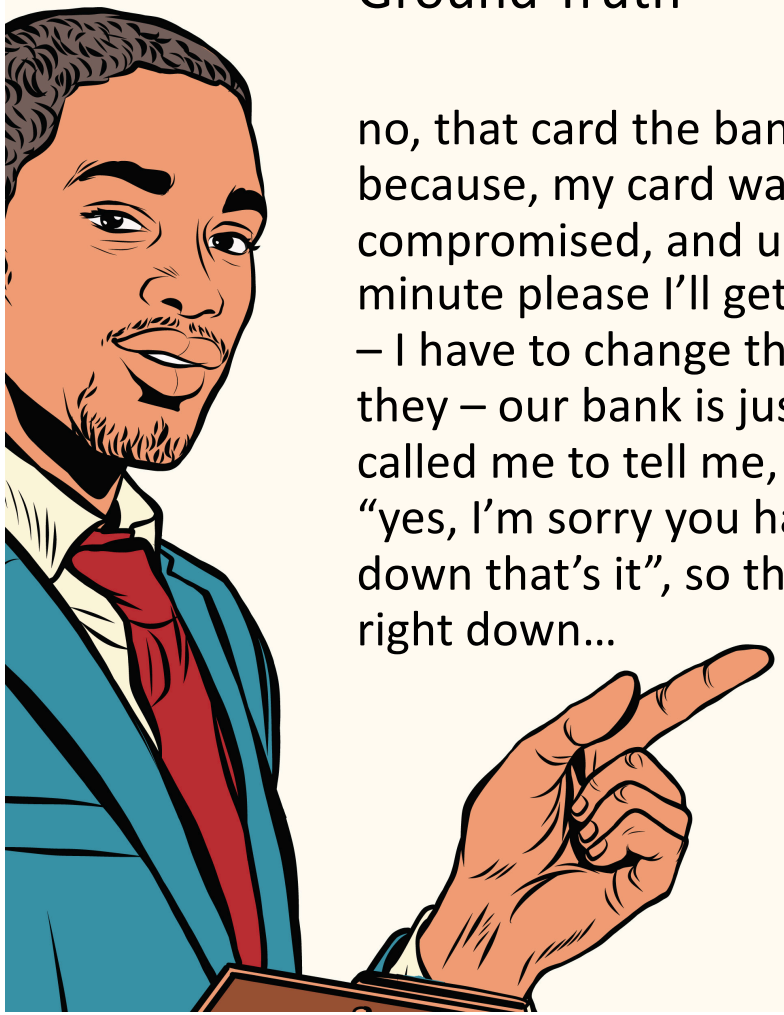

Listen
use non-verbal cues to nudge speakers towards natural speech for better recognition


Alas, poor Bender!
I knew him..
Recognize

no, that card the bank shut down because, my card was compromised, and um – hold on a minute please I'll get this card here – I have to change the card umm – they – our bank is just good, they called me to tell me, I said "no no", "yes, I'm sorry you have to shut it down that's it", so they closed it right down…


no, that card the bank shut down because my card was compromised
det nsubj det dep root prt
that card the bank shut down
DET NOUN DET NOUN VERB PRT
no, that caused the bank to shut down because my card was compromised
nsubj root det nsubj aux xcomp prt
that caused the bank to shut down
DET VERB DET NOUN PRT VERB PRT


WER = substitutions + deletions + insertions / number of words



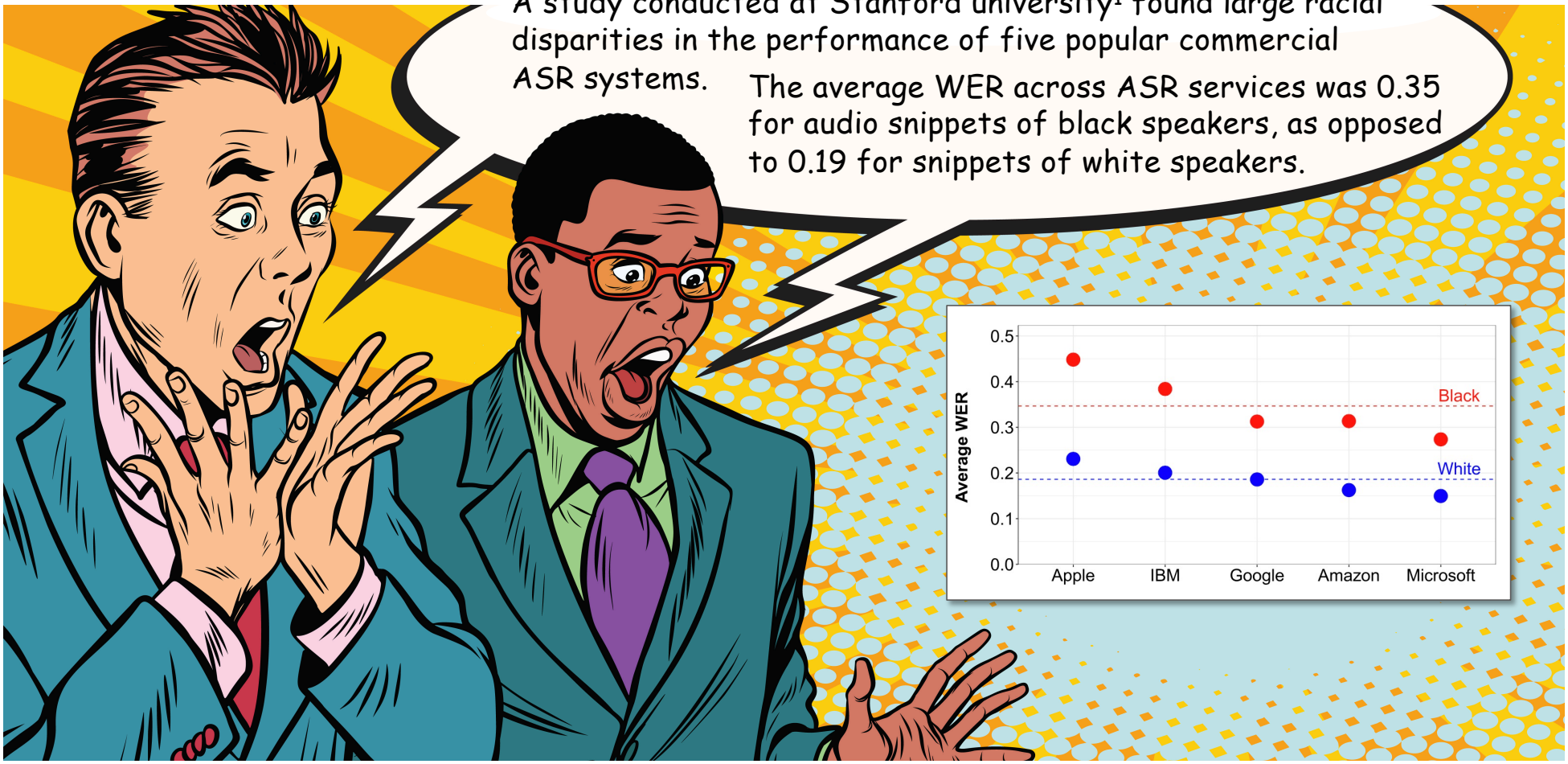
Ground Truth

no, that card the bank shut down because, my card was compromised, and um – hold on a minute please I'll get this card here – I have to change the card umm – they – our bank is just good, they called me to tell me, I said "no no", "yes, I'm sorry you have to shut it down that's it", so they closed it right down…

Transcription

No that caused the bank to shut down because my card was compromised . And ___ hold on a minute please . Because I have to change the card and they are bank as ___ ___they call me to tell me actually __ ___ No no ___ I'm sorry you had to set __ down this way so they closed it right down…

substitution insertion deletion



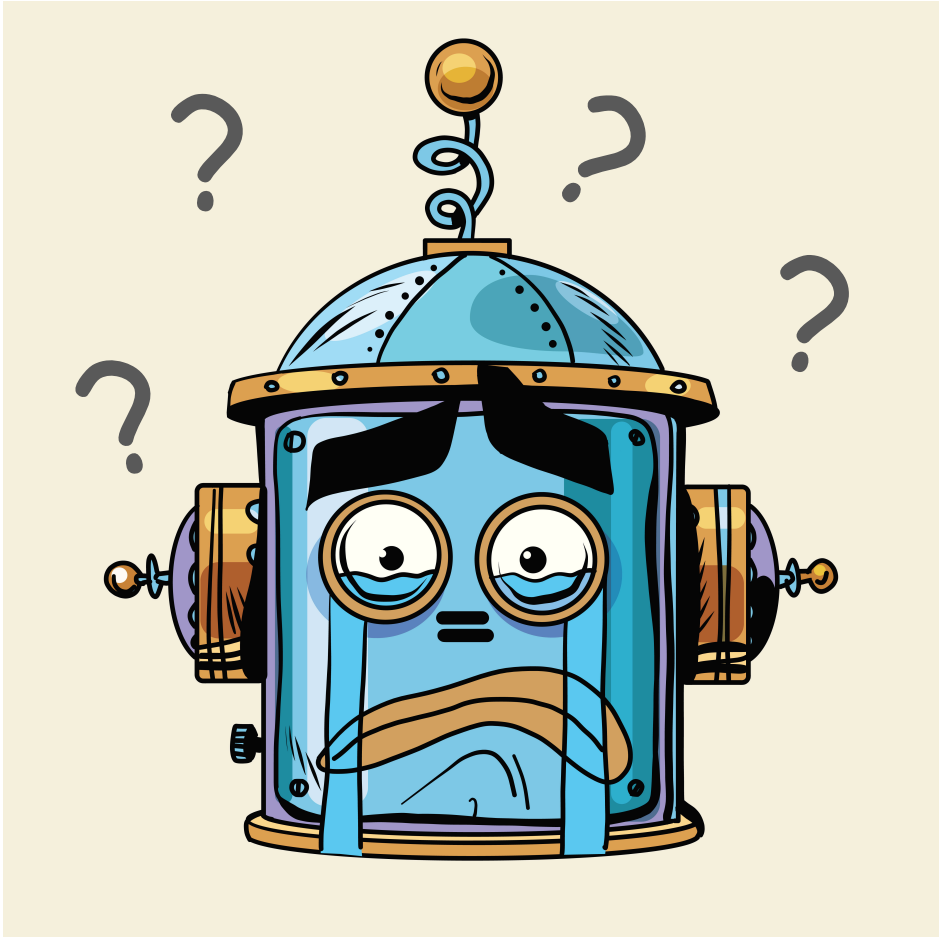
[1] Racial disparities in automated speech recognition - 2020 Apr 7;117(14):7684-7689. doi: 10.1073/pnas.1915768117. Epub 2020 Mar 23 - Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R Rickford, Dan Jurafsky, Sharad Goel

race

age

gender

language

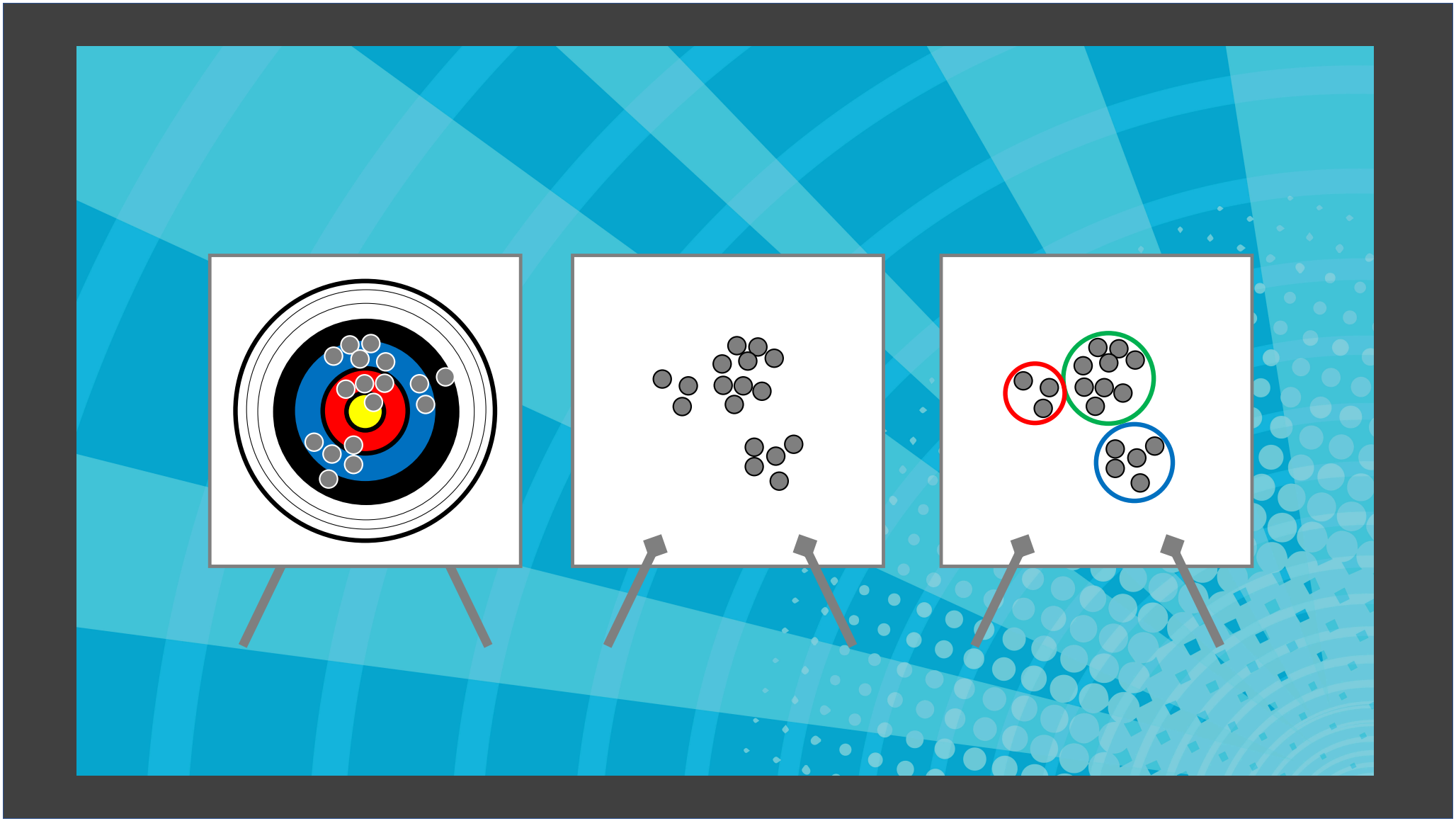


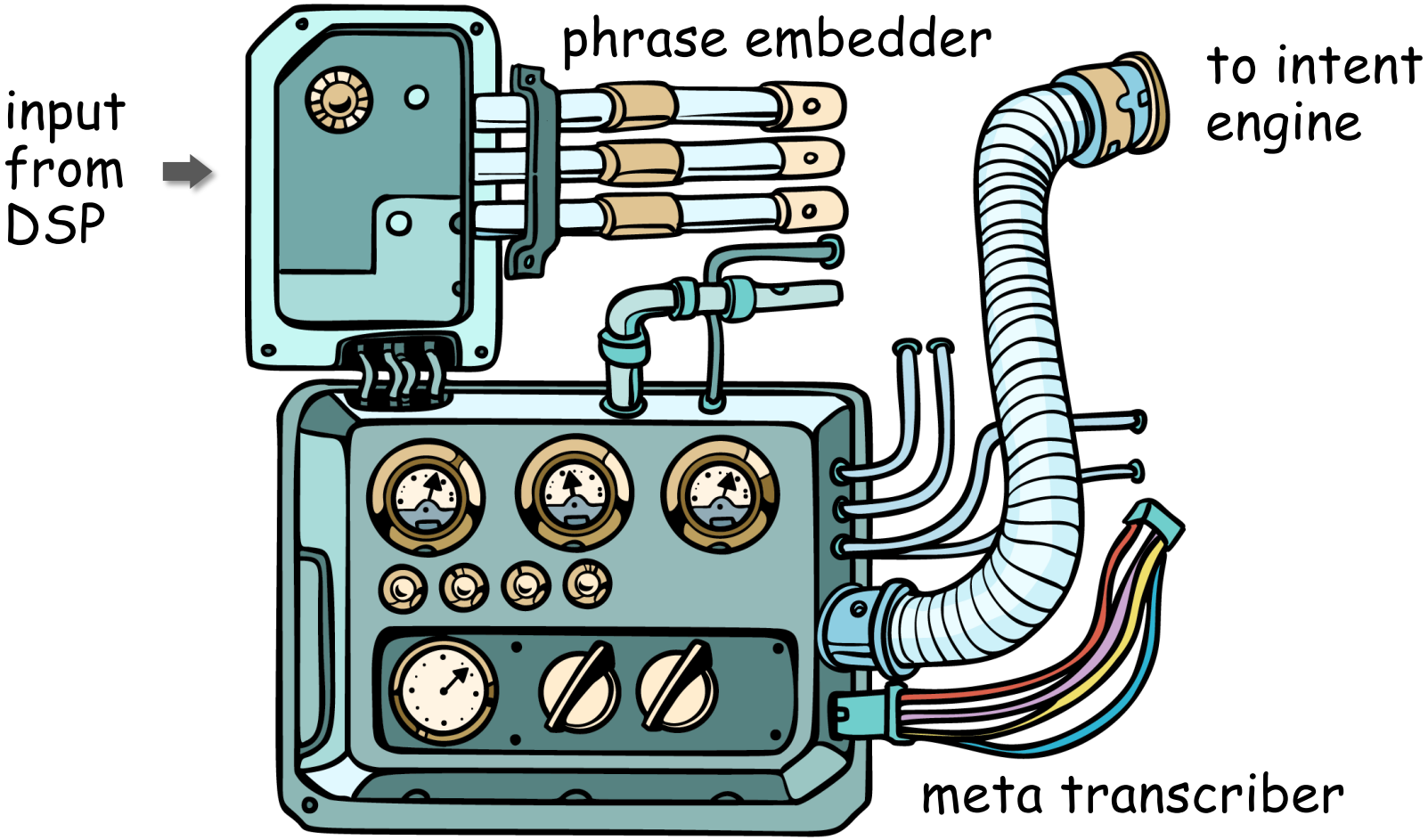
rapid tempo

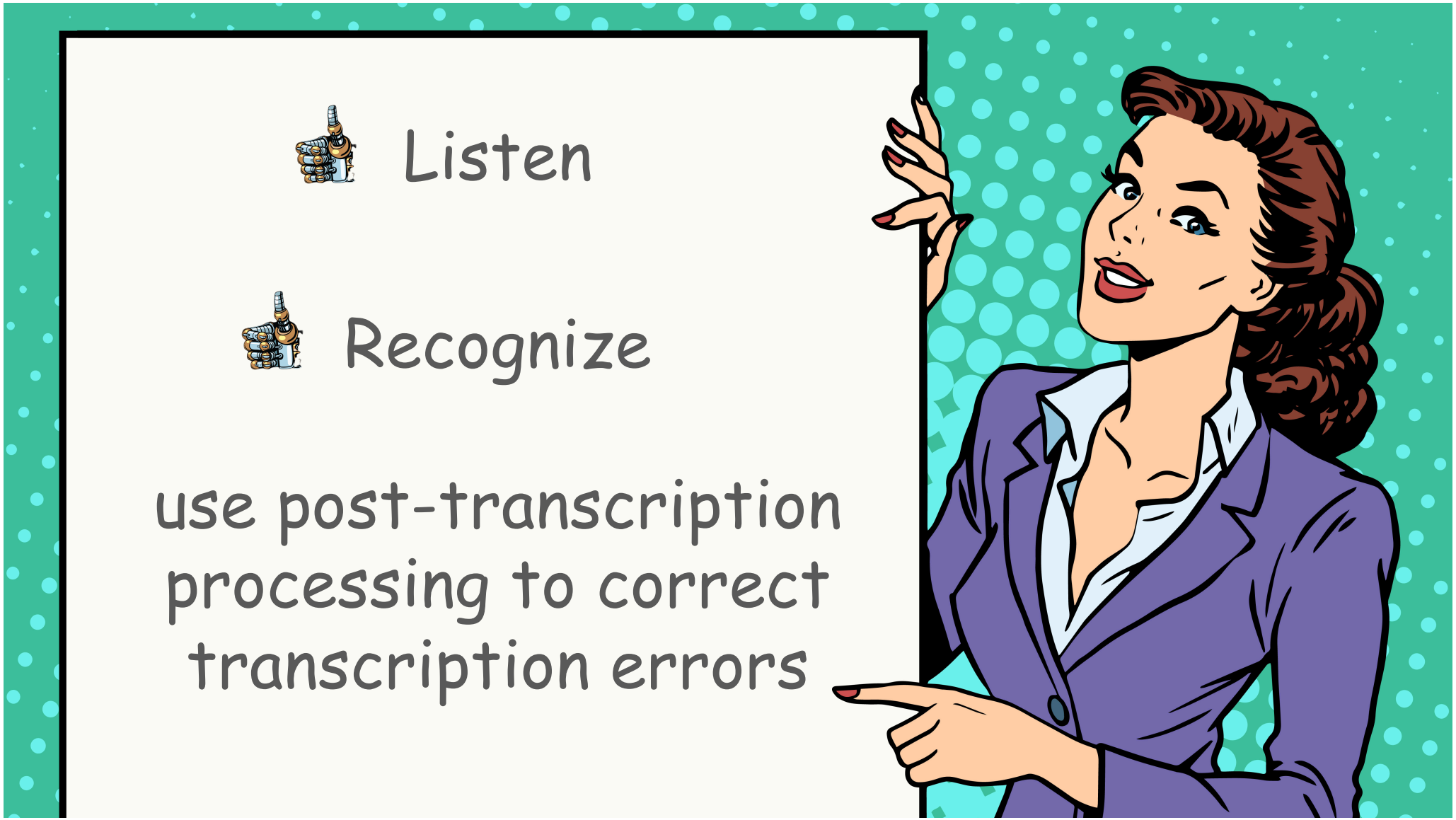
sentence structure

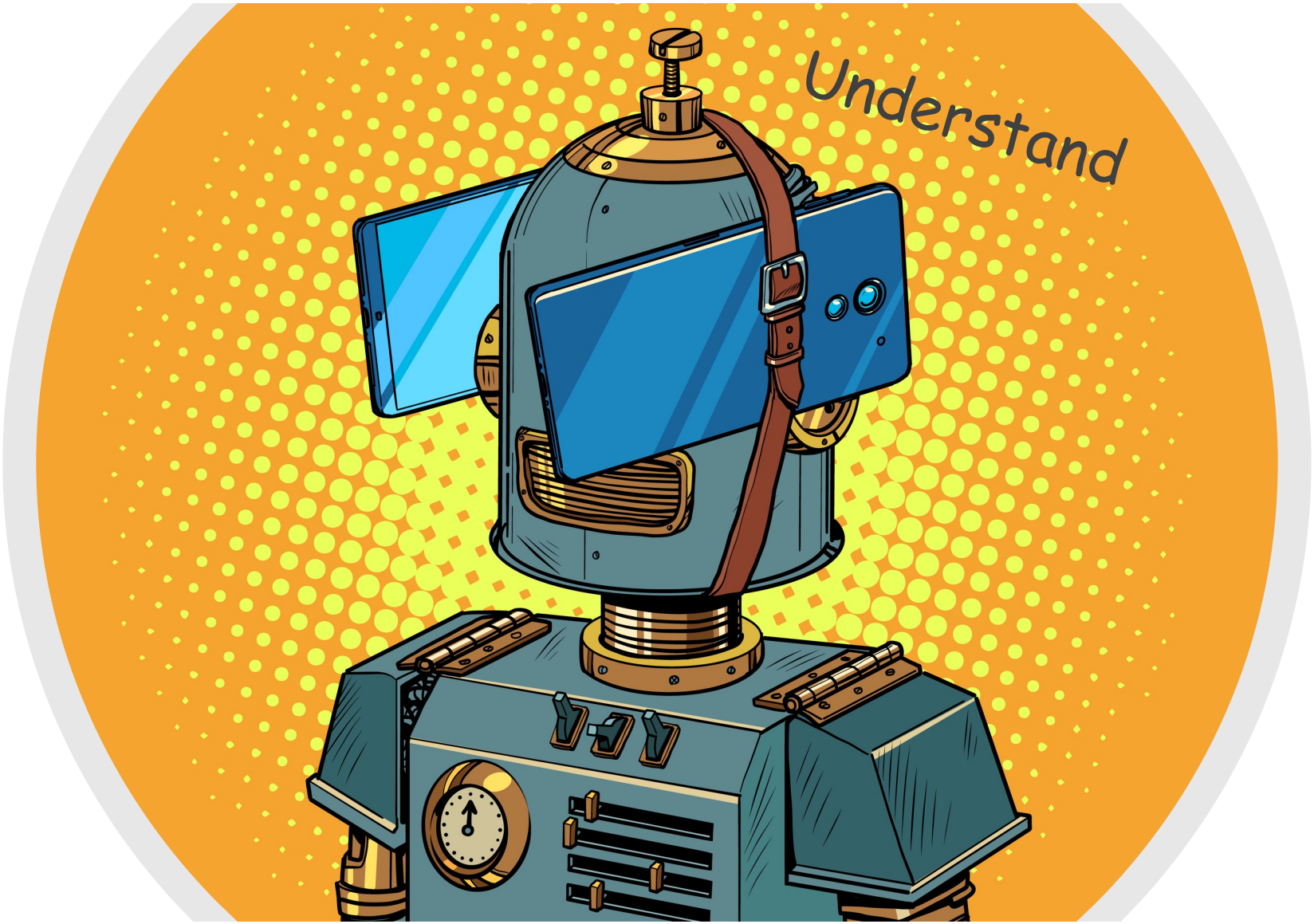
low intensity

hesitations

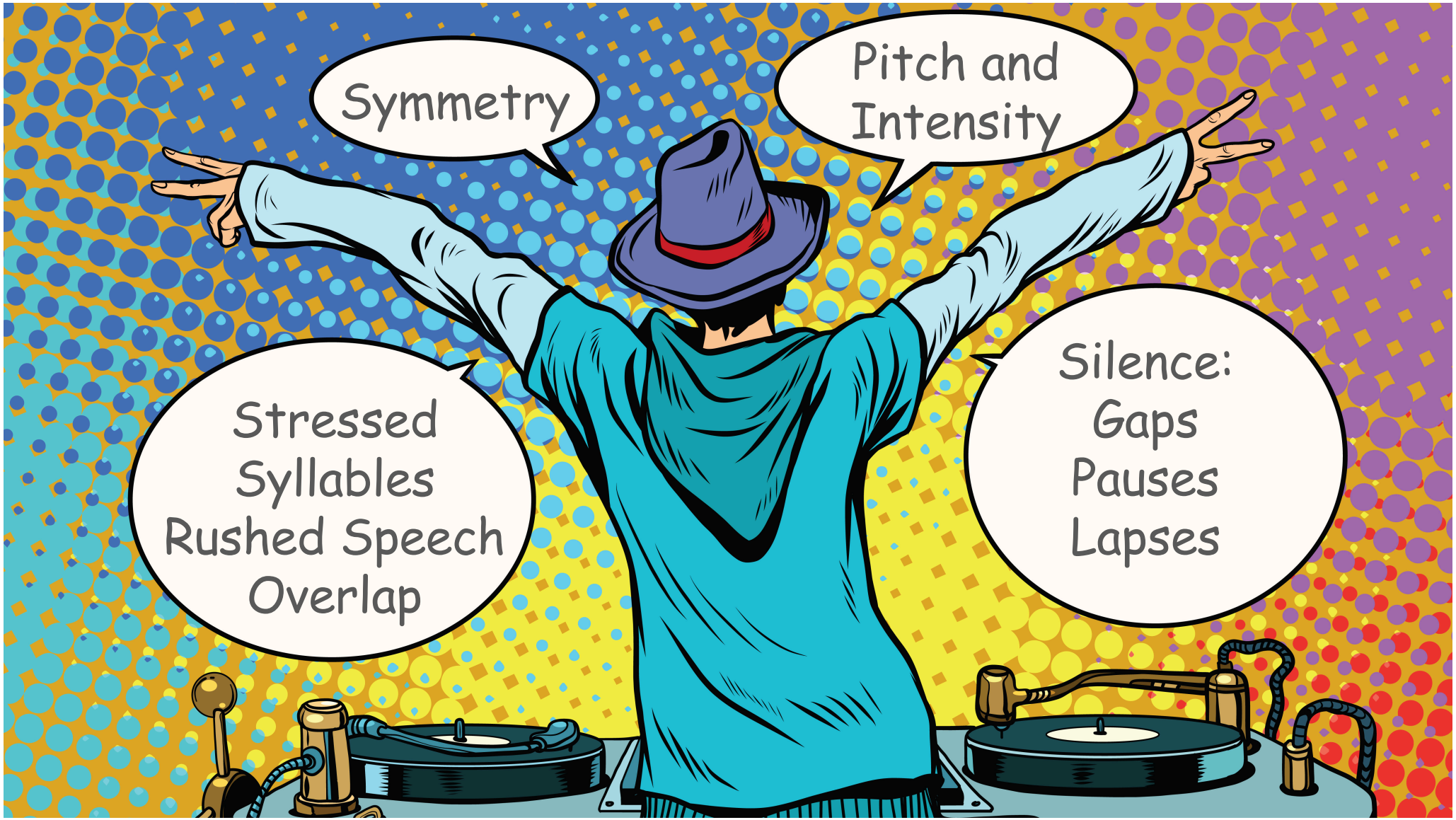



input
from
DSP
phrase embedder
to intent
engine
meta transcriber


Listen
Recognize
use post-transcription processing to correct transcription errors


Understand

no, that card the bank shut down because, my card was compromised, and um – hold on a minute please I'll get this card here – I have to change the card umm – they – our bank is just good, they called me to tell me, I said "no no", "yes, I'm sorry you have to shut it down that's it", so they closed it right down…


Symmetry
Pitch and Intensity
Stressed Syllables Rushed Speech Overlap
Silence: Gaps Pauses Lapses

increased intensity

no, that card the bank shut down because, my card was compromised, and um – hold on a m

hesitation

please I'll get this card here – I have to change the card umm – they – our bank is just good, they called me to tell me, I said "no no", "yes, I'm sorry you

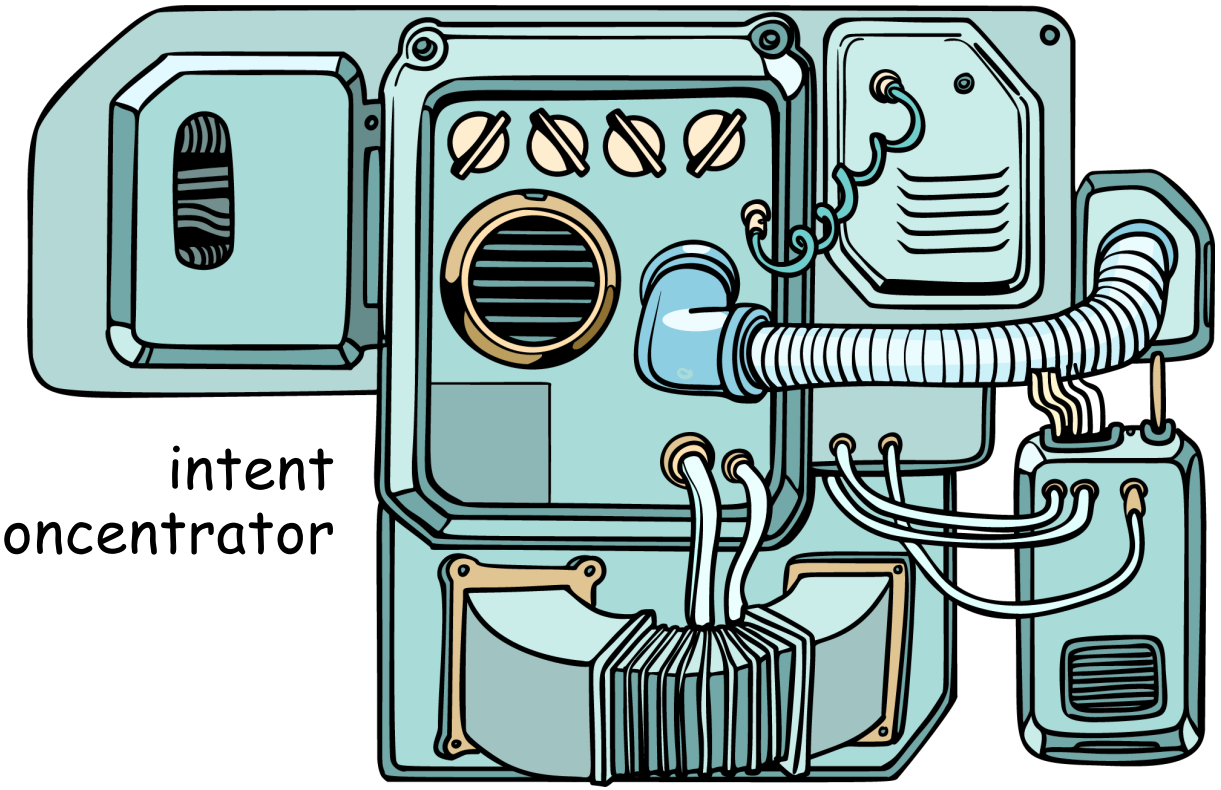
low volume

rushed speech

ut it down that's it", so they closed it right down…

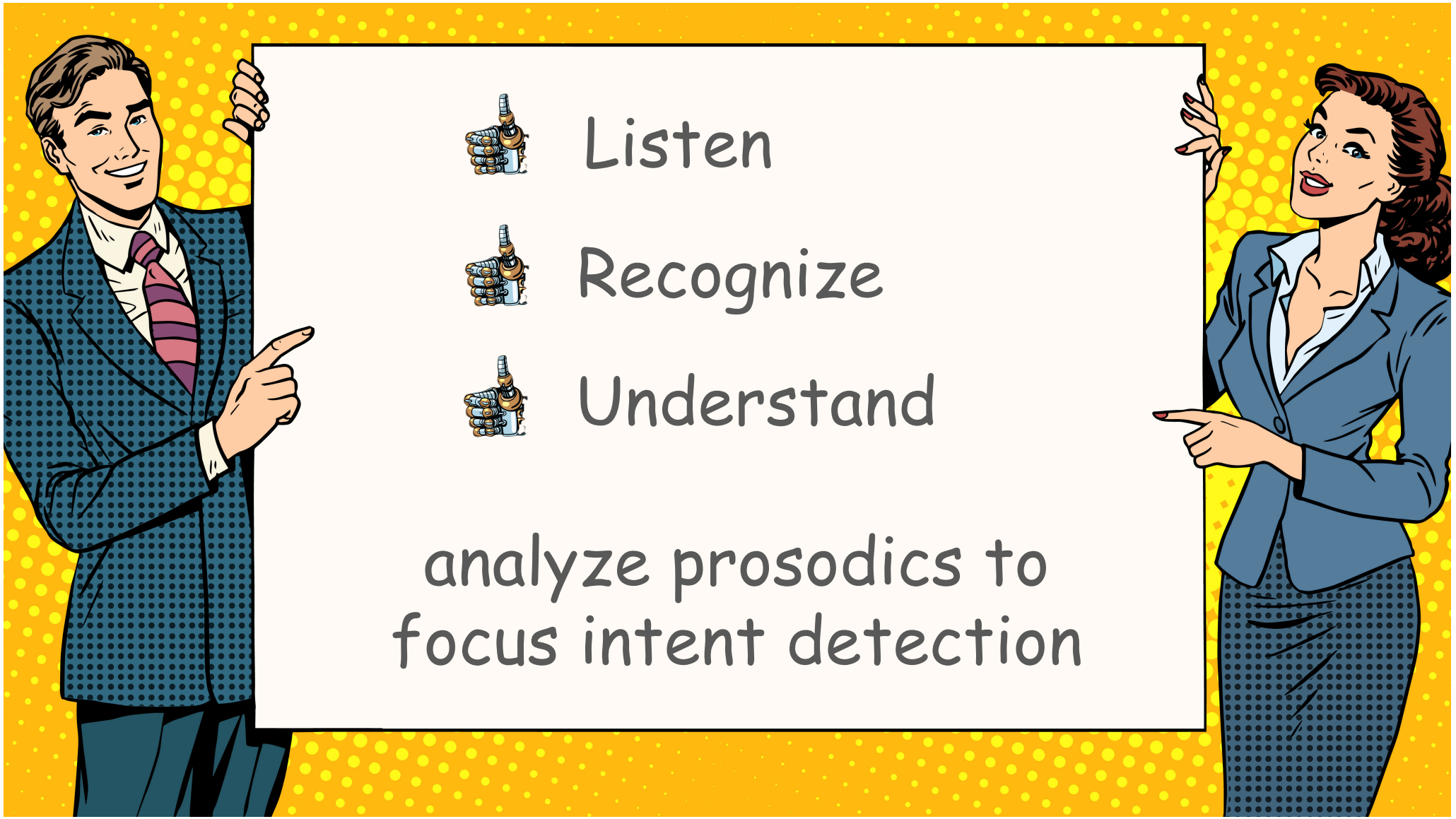
elongation

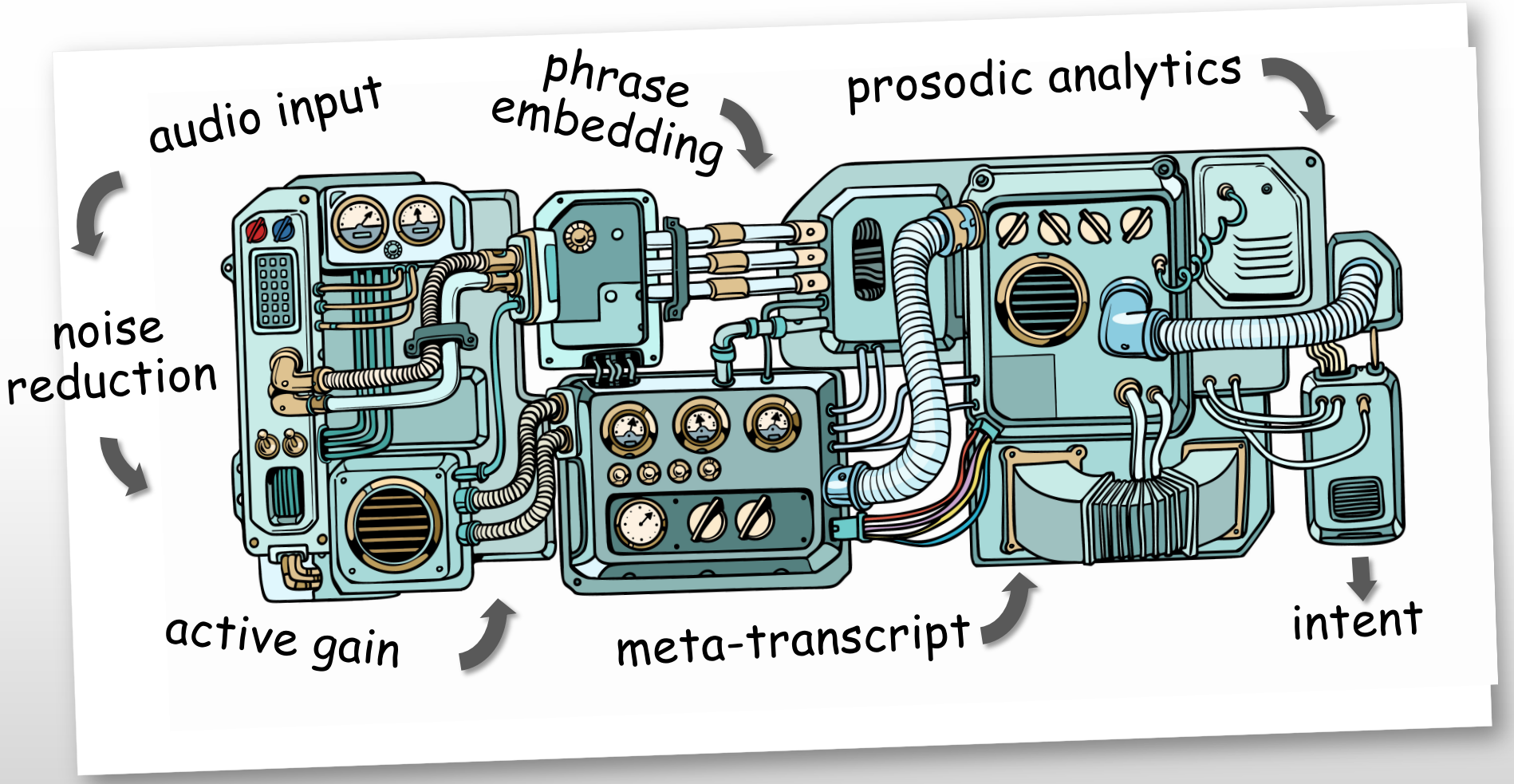
prosodic analyzer



intent concentrator

distilled intent


Listen
Recognize
Understand
analyze prosodics to
focus intent detection


audio input
phrase embedding
prosodic analytics
noise reduction
active gain
meta-transcript
intent

Mariano Tan

President, CEO Prosodica LLC

A VAIL COMPANY

mariano.tan@prosodica.com

www.prosodica.com

linkedin.com/in/marianotan

Twitter: @MarianoTan

Grafiker61, CC BY-SA 4.0 <https://creativecommons.org/licenses/by-sa/4.0>, via Wikimedia Commons