

On Speech Processing

Yannis Stylianou

yannis@csd.uoc.gr



BioSketch

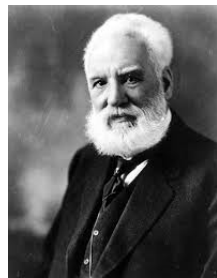
[Yannis Stylianou](#) is Professor of Speech Processing at University of Crete, in Greece and Research Manager at Apple, Cambridge UK.

From 1996 until 2001 he was with AT&T Labs Research (Murray Hill and Florham Park, NJ, USA) and until 2002 he was with Bell-Labs Lucent Technologies, in Murray Hill, NJ, USA. He is with University of Crete since 2002.

From 2013 until 2018 (July) he was Group Leader of the Speech Technology Group at Toshiba Cambridge Research Lab in Cambridge UK. He joined Apple in Aug 2018. He holds MSc and PhD from ENST-Paris on Signal Processing and he has studied Electrical Engineering at NTUA Athens Greece (1991).

He is an IEEE Fellow, an ISCA Fellow and an AAIA Fellow.

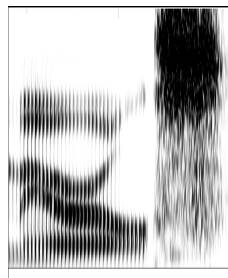
Speech has a central position in human communication



Bell (1876)
discovery of
telephone



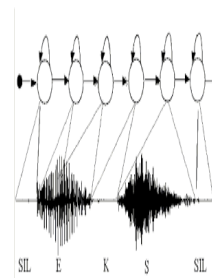
Rayleigh (1900)
theory of sound



Speech
spectrogram
(1946)



Shannon (1948)
speech & language
transmission



Markov chain
(Baum, 1960)



Békésy (1961)
frequency coding



Itakura (1970)
Autoregressive
modelling



Turing (1950)
thinking machine

Understanding speech production and acoustics led to ...

✓ Improved communication



✓ Enhanced hearing



✓ Advanced speech technologies



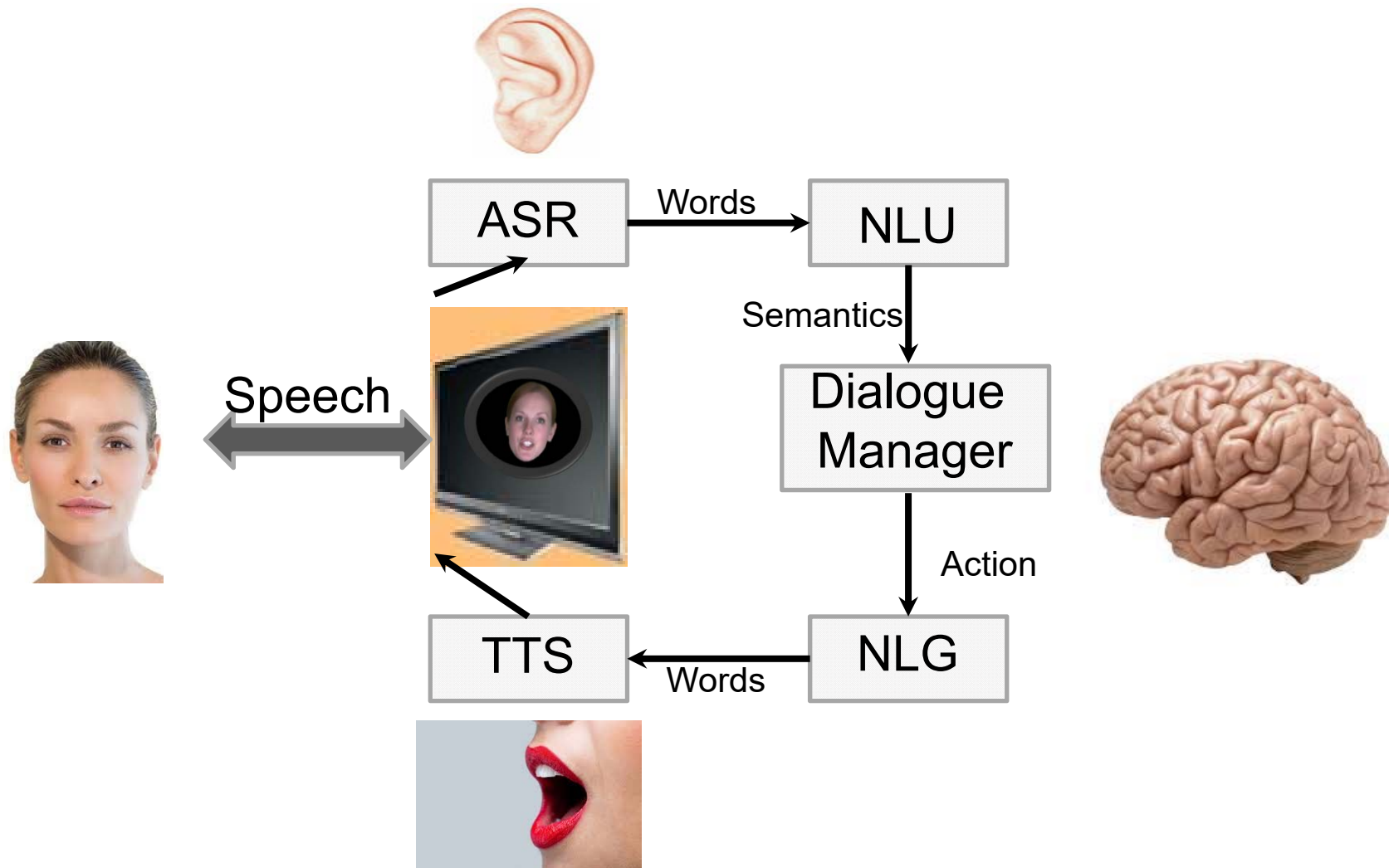
- Text-to-Speech Synthesis (TTS)
- Automatic Speech Recognition (ASR)

Example of natural human-machine communication



... with the CRL statistical spoken dialogue manager

Main Speech Technologies



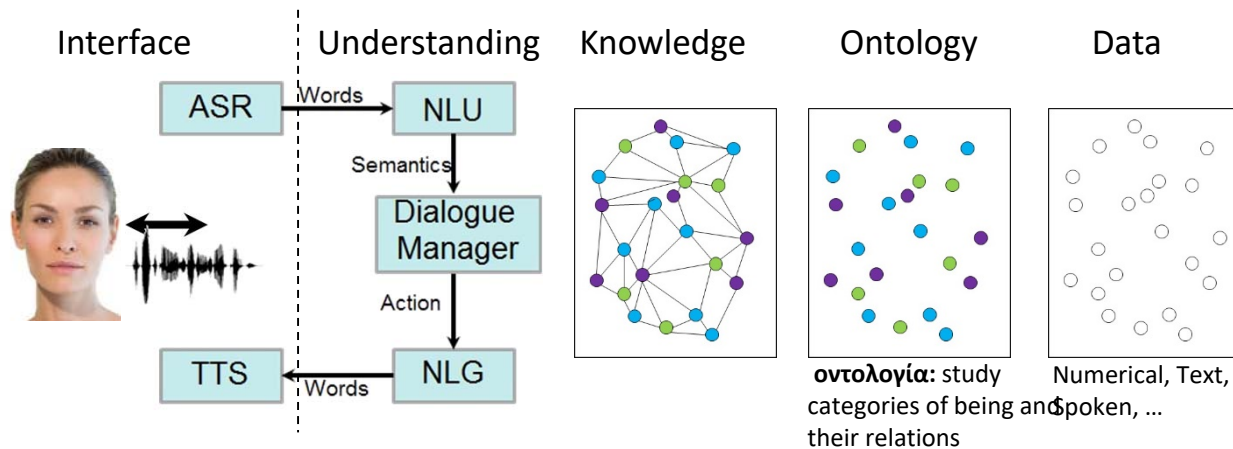
Trend: End-to-end deep learning solutions

From information retrieval to thinking machines

Combining speech with machine learning will lead to effective human-machine communication

Learn from human:

Data driven approaches



1. natural, speech enabled, human-machine interface for information retrieval

2. learn human's procedures

- ❖ Design human centric information processing algorithms and services to create and access knowledge effectively, for improving productivity and quality of life

ASR: Automatic Speech Recognition; **NLU:** Natural Language Understanding;
NLG: Natural Language Generation; **TTS:** Text-to-speech

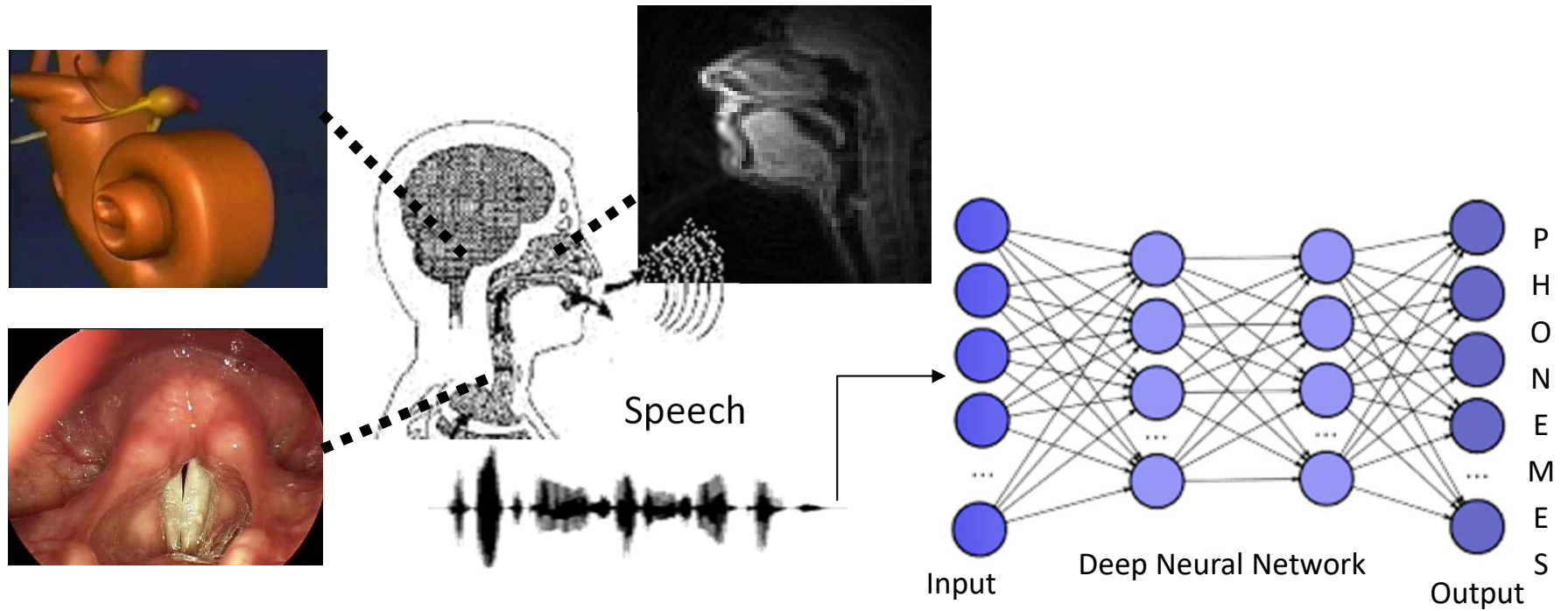
Human-like:




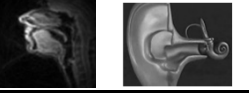
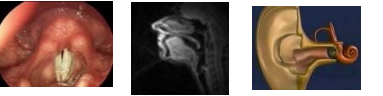
thinking machine



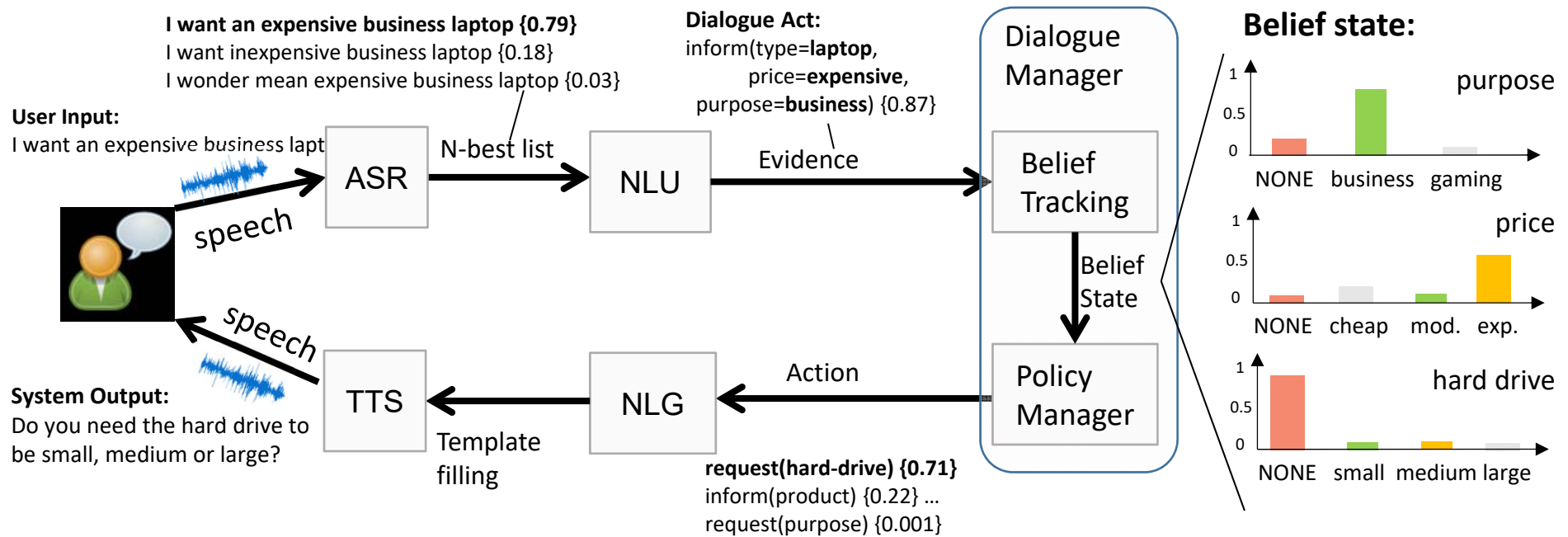
- make suggestions, compare, planning

Automatic speech recognition: speech to text

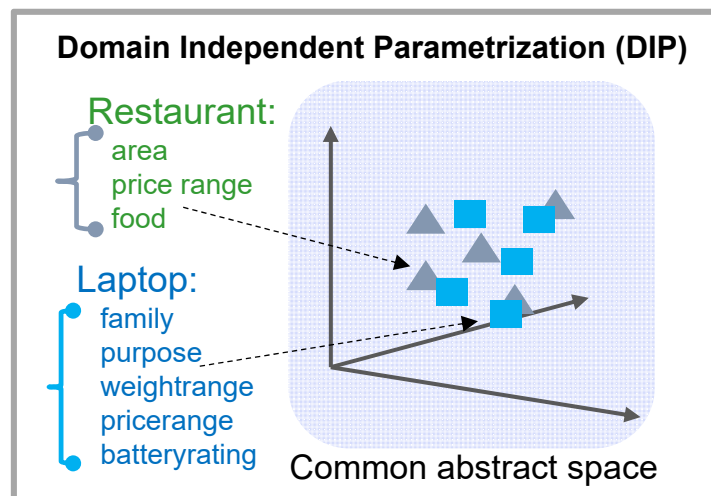


Approach	WER(%) 	WER(%) 
Waveform 	42	58
State of art 	14	34
CRL 	12	33

Statistical Dialogue Manager



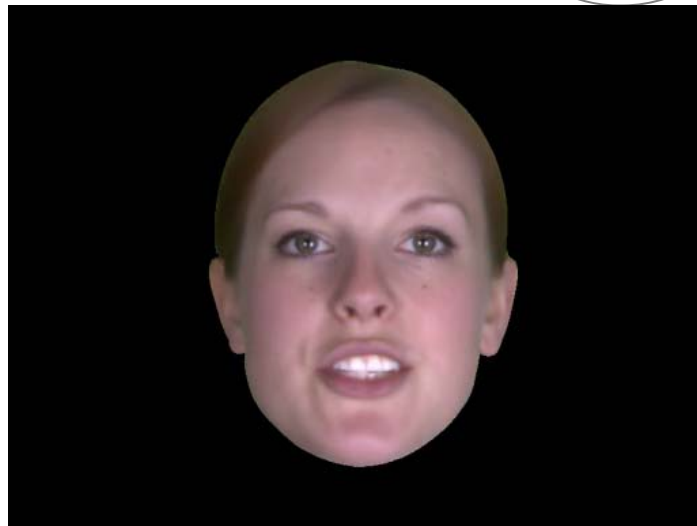
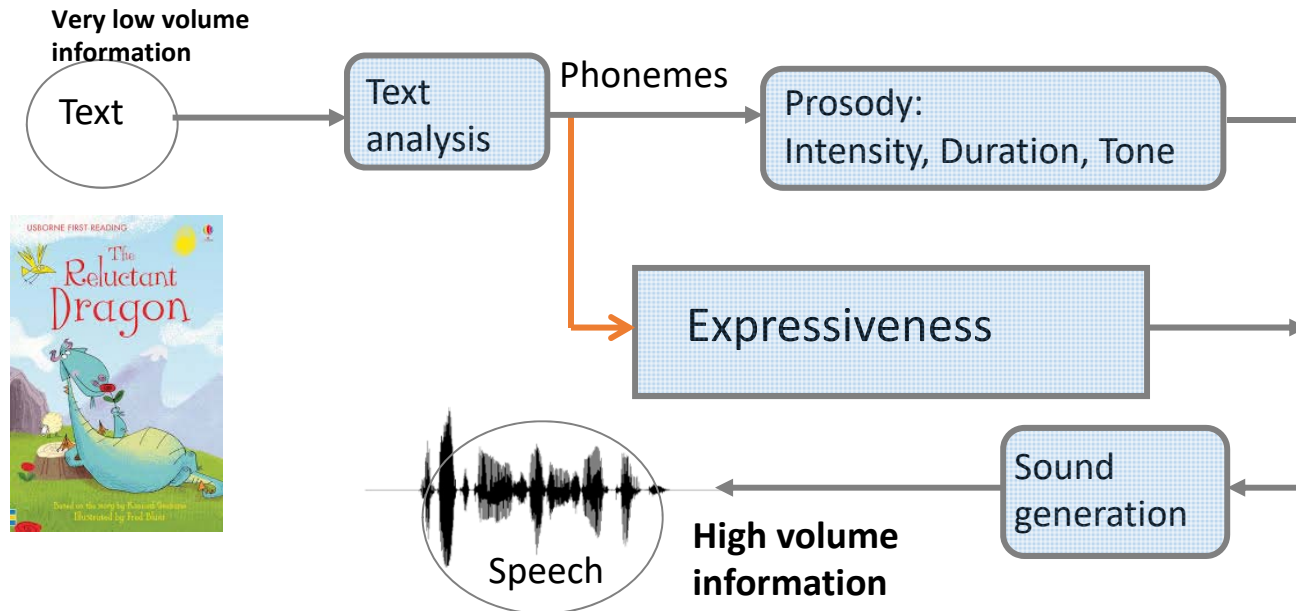
➤ Transfer learning



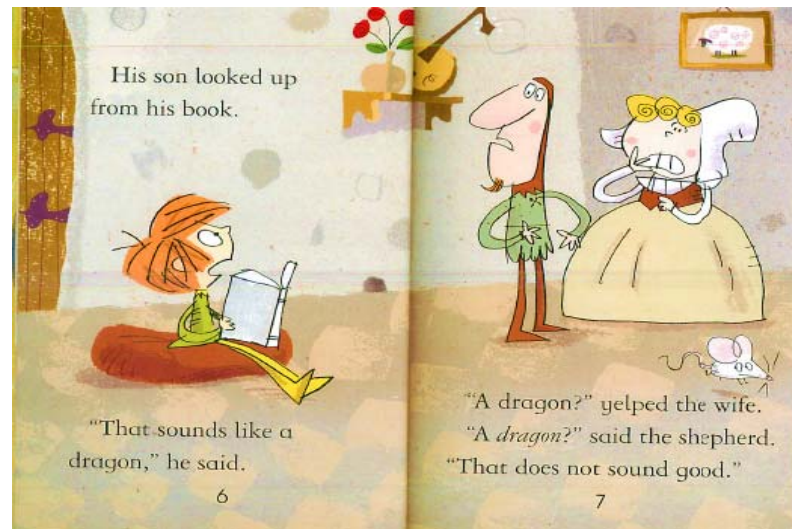
	In Domain	Transfer learning
Success rate	85%	82%

ASR: Automatic Speech Recognition; **NLU:** Natural Language Understanding; **NLG:** Natural Language Generation; **TTS:** Text-to-speech

Flexible and high quality visual text-to-speech synthesis



Xpressive Talk™

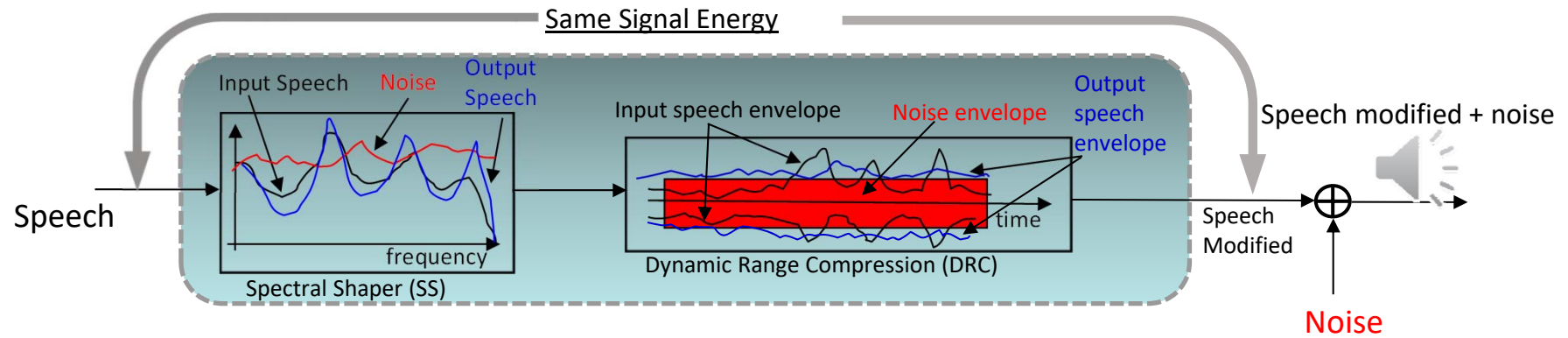


Intelligibility of speech in noise

➤ **Problem:** Speech Perception in Noise



➤ **Solution:** Spectral Shaping and Dynamic Range Compression (SSDRC)



➤ **Applications:**



Example of natural human-machine communication



... with the CRL statistical spoken dialogue manager