

Voice and speech perception. Human vs automatic. How to build clinical tools ?

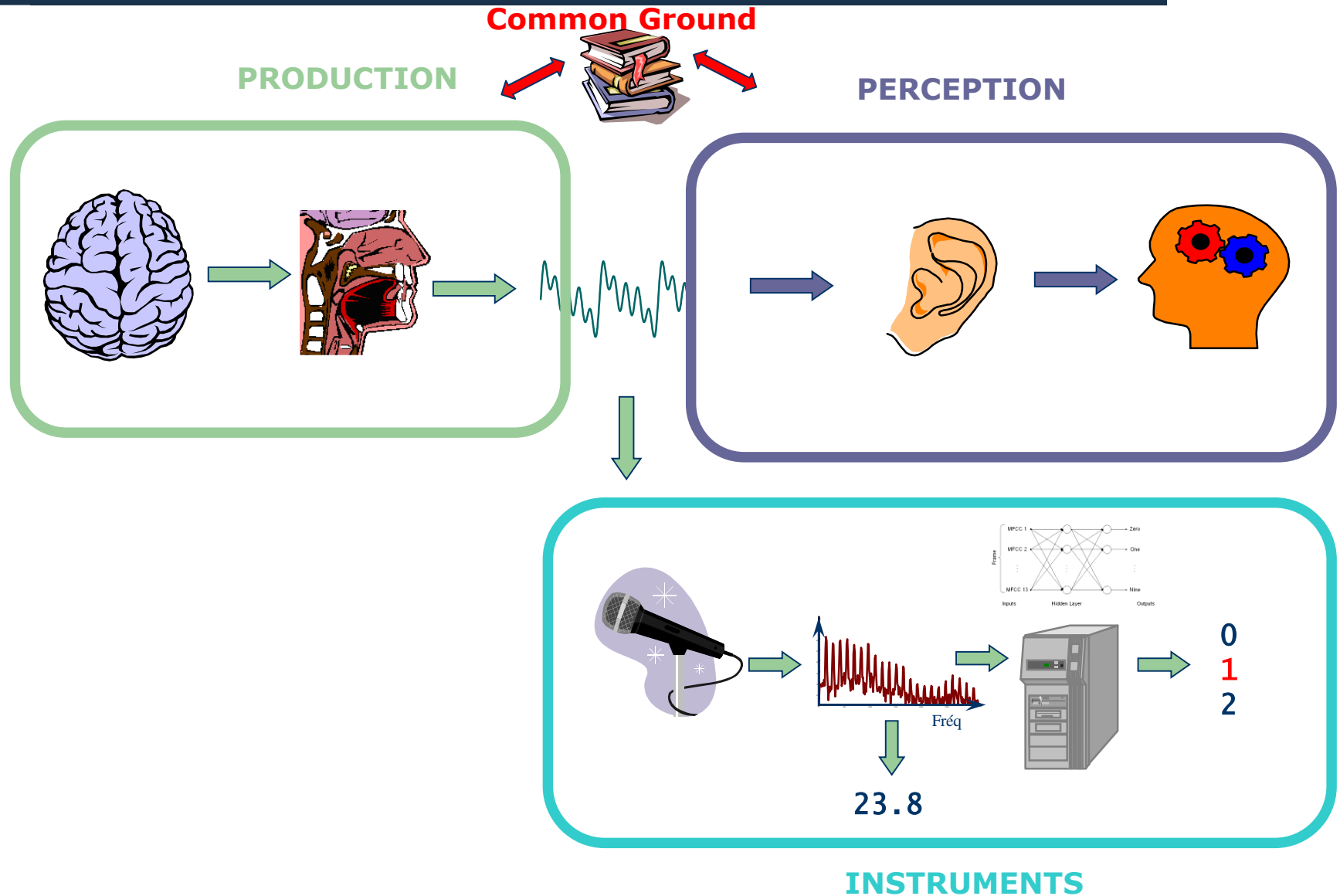
Alain Ghio

**Laboratoire Parole et Langage
Aix-Marseille University & CNRS
France**

**UEP 2018
Helsinki**



Spoken communication



Instrumental voice/speech assessment

- Old antagonism between perceptual vs. instrumental voice/speech assessment

Kent et al., 2000, "Voice quality measurement", Singular

p1. "Some would argue that voice quality can be described and assessed only by auditory-perceptual means... "

p1. "Still others might claim that auditory-perceptual methods are prone to various source of error..."

p117. "It is generally accepted in current research that the impressionistic nature of perceptual description should be augmented by more objective nature of instrumental analysis where possible."

p117. "It is important however not to assume that instrumental approaches are somehow correct and infallible. Limits on accuracy exist with all techniques and the interpretation of results is not always straightforward"

An old story

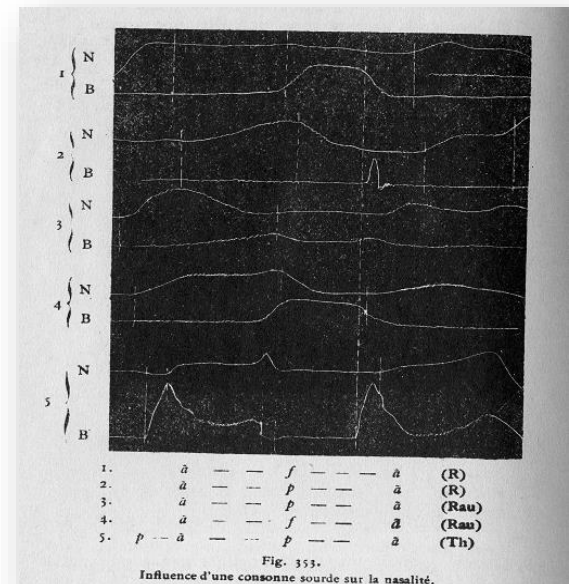
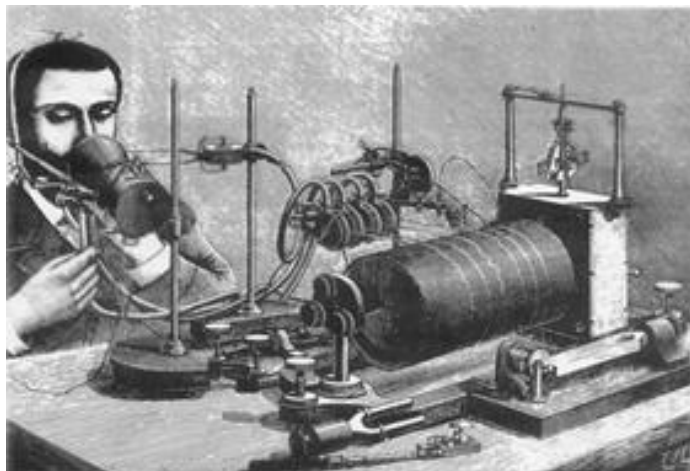


- 1875 : Société de Linguistique de Paris consults Etienne-Jules Marey (French physiologist who studied movements of animals and humans) in order to...

« apply the graphic method to the study of the complex and varied movements that occur in speech [...] to obtain an objective trace of the movements of articulatory organs, rib cage, larynx, tongue, lips, soft palate, during the articulation of different phonetic unit»

Because « The ear, the sight and the palpation are not sufficient to apprehend the successive and simultaneous movements necessary for the realization of a phoneme »

- Devices developed by Abbé Rousselot



when the instruments seem to be wrong

- Basic measurement errors (ex : octave jumps, bad calibration...)
- Sound quality vs noise
- Stability
 - ✓ Of instrument : OK (deterministic process)
 - ✓ Of voice/speech : no warranty
 - ✧ Instruments too sensitive
- Instruments can detect phenomena that cannot be perceived because of perceptual top down effects (phonemic restauration,phonological deafness)

World Health Organisation : *International classification of Impairments, Disabilities and Handicaps, a manual of classification relating to the consequences of disease*

[illegible]

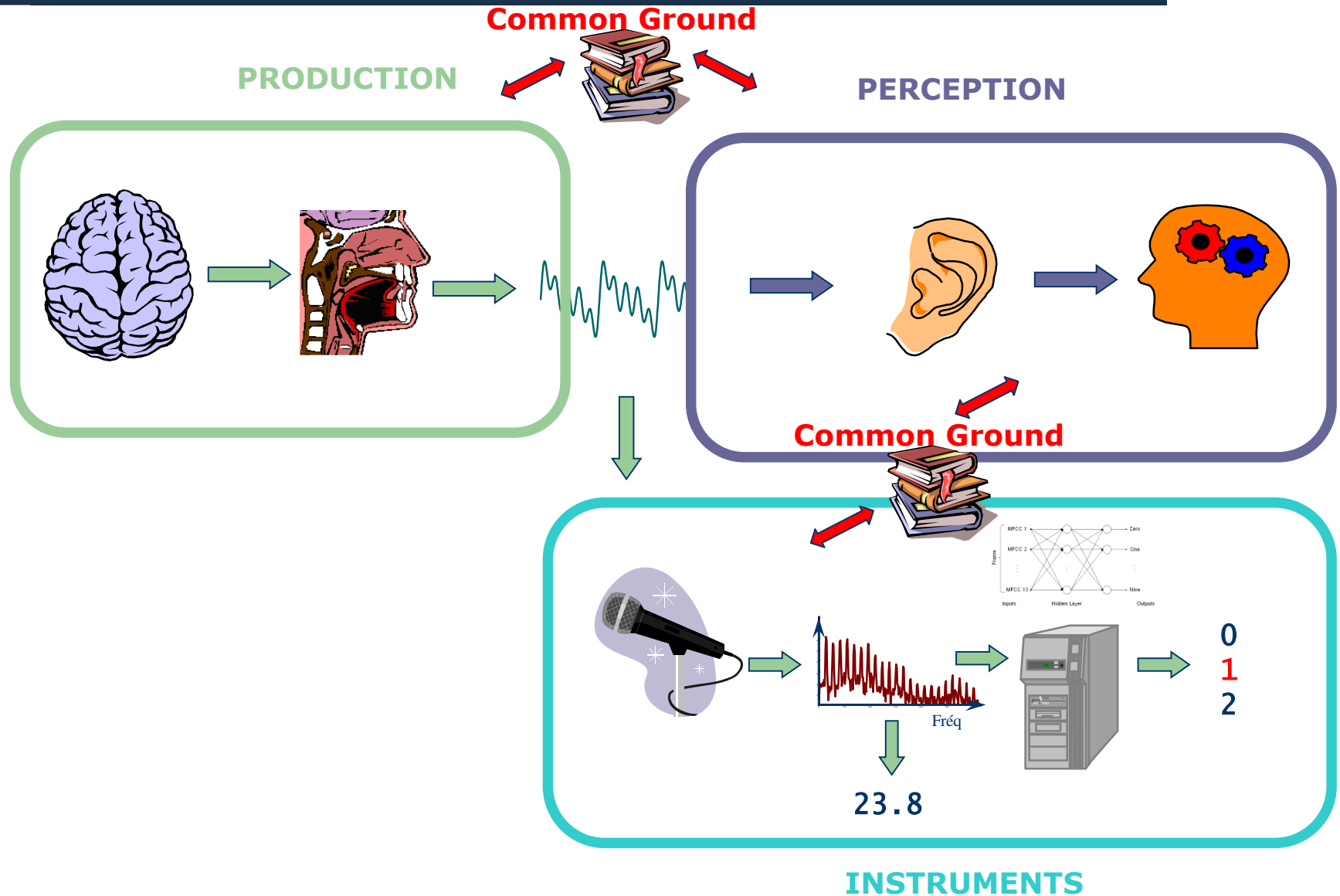
Perfect world/ real world

- Assessments are consistent
 - ✓ Abnormal values (ex: HNR, speech rate...)
 - ✓ Poor intelligibility
 - ✓ High Handicap Index
- Real world
 - ✓ Inconsistent relation between Handicap Index vs perception/measure




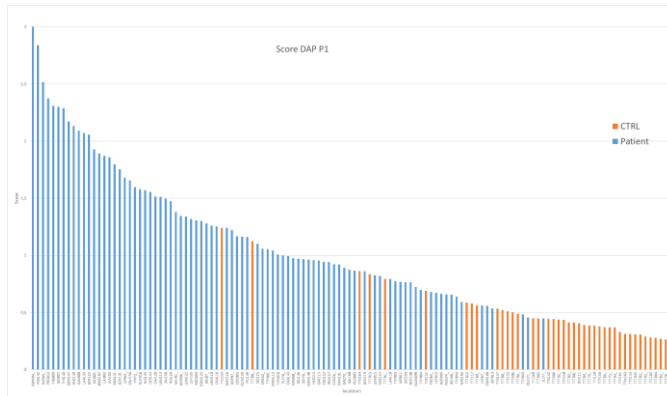
- ✓ Inconsistent relation between measures/perception

Spoken communication



Acoustic-Phonetic Decoding by humans

- Project C2SI : Carcinologic Speech Severity Index 
- Speaker pronounce pseudo-words (a word without meaning)
 - ✓ Ex: tafi, respo...
- Listeners transcribe what they « understand » (decode)
- Transcription is compared to expected target :
 - ✓ The distance between them can be seen as a severity score



Transcription « tovène »

	s	t	o	l	5
t	1	0	9	2	9
o	9	9	0	7	4
v	2	3	8	3	8
E	9	9	3	7	1
n	4	3	8	1	7

Distance locale

Distance cumulée

12

Cible « stolain »

Assessment by machines

- Automatic acoustic-phonetic decoding based on Automatic speech recognition techniques
- Comparison with human acoustic-phonetic decoding
- Others speaking tasks
- Accepted if not a blackbox
- Useful if we can improve our knowledge
 - ✓ Deep Neural Network could be a good candidate



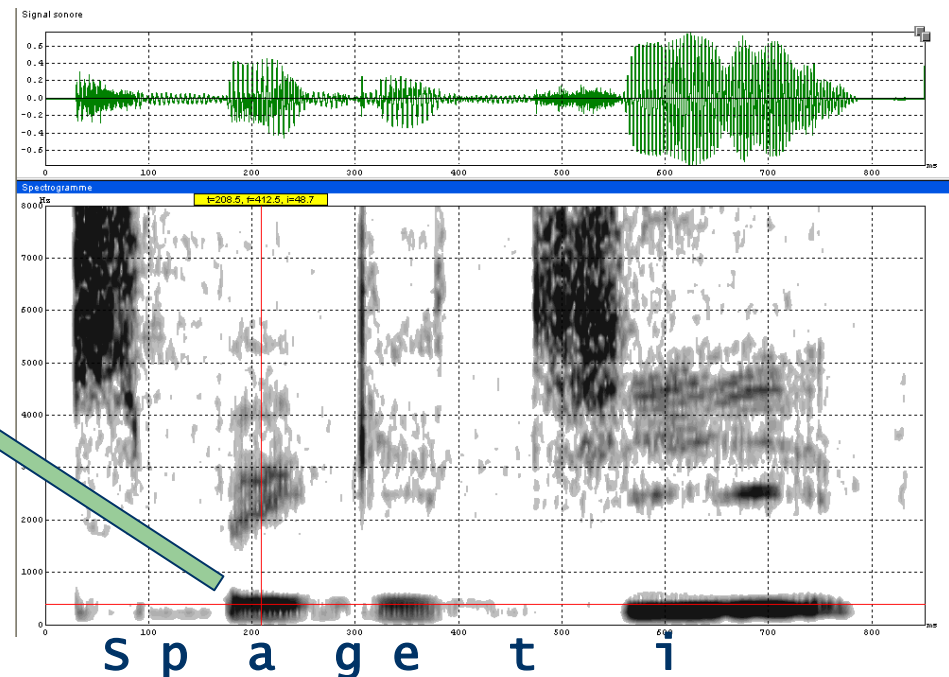
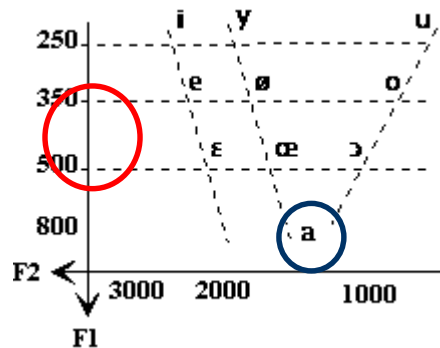
THANK YOU

when the instruments seem to be wrong

- “j’aime bien manger des spaghettis” ⇔ I like to eat spaghetti produced by a French speaker in spontaneous speech



/sp~~a~~geti/



- F1 = 400 Hz
- Instrument is inaccurate ?
- No !!

- spaghetti spaghetti

Meunier C., Floccia, C. (1999) "Syllabe ou mot : quelle unité permet d'identifier les catégories phonétiques?", Actes des 2èmes Journées d'Etudes Linguistiques, "Syllabes", Nantes, pp. 87-92.

TDNN and speech

- Nagamine et al. Exploring How Deep Neural Networks Form Phonemic Categories, INTERSPEECH 2015

