



Human Voice Representation – further Progress

Robert Mores
Hamburg

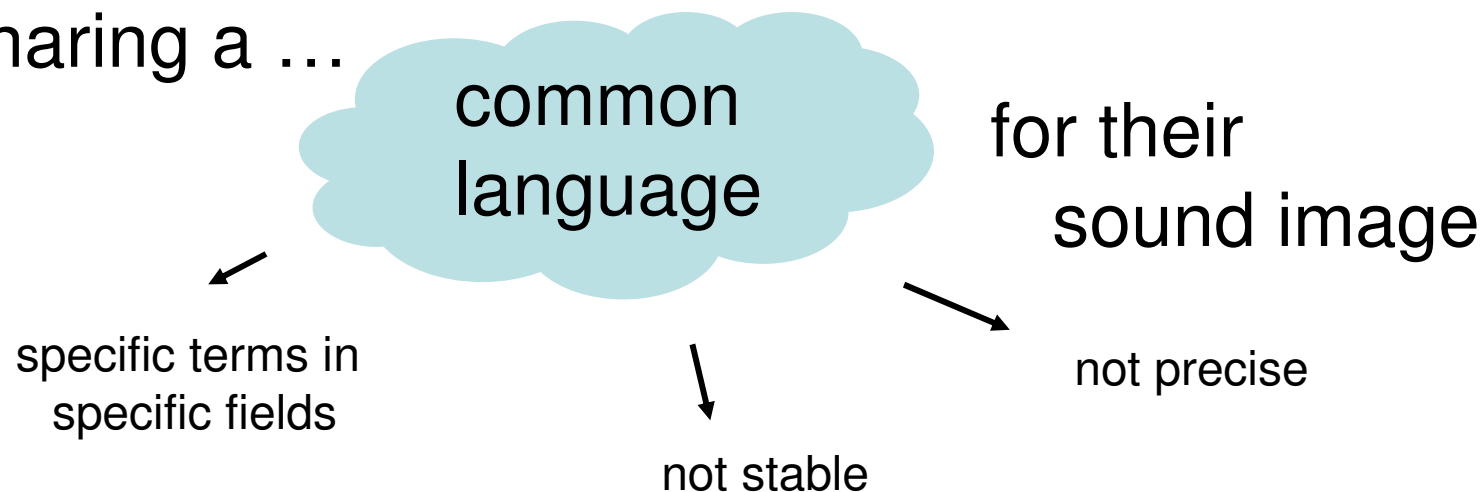
Foto: © Zippo Zimmermann, www.designladen.com

Why don't we use features of human voice to describe most of what we perceive when we listen to sounds?

- Dilemma of bridging two worlds
- Capability of human voice features
- Examples
- Value of the representations
- Summary

World of musicians ...

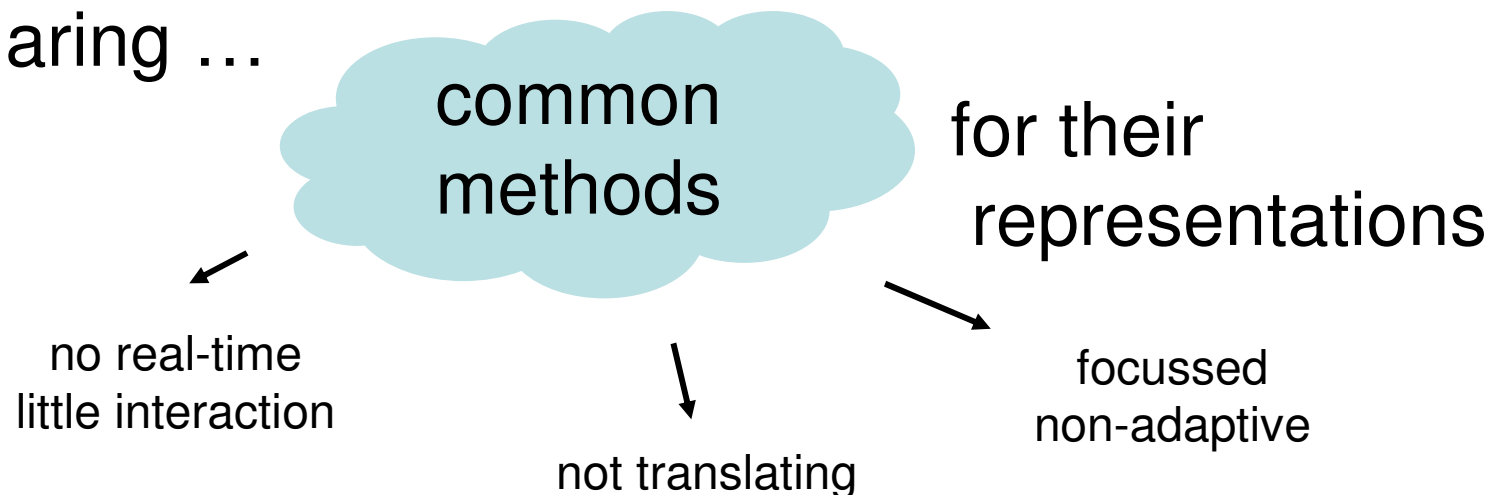
- makers of musical instruments, recording engineers, composers, ...
- sound-centered
- sharing a ...



A. Liebe 1952: investigating 1600 words, 16th to 19th century

World of engineers ...

- engineers, science, information technology, ...
- technology-centered
- sharing ...



Rare ‚successful‘ examples

perception	techn. representation	who
roughness ↔	modulation parameters	Aures, Daniel
loudness ↔	energy in bands, statical and dynamical masking	Zwicker Moore, Glasberg
sharpness ↔	weighted specific loudness	Bismarck, Aures
brightness ↔	spectral average (centroid)	Beauchamp, Grey
singing ↔ ?		
noble ↔ ?		

do we trust ?

Dilemma

- 500 JASA publ. on piano: only few with human factors
- 1100 JASA publ. on strings: a few incorporate human hearing
- 3000 CATGUT publ. on musical acoustics: 3 with musicians, another 5 consider hearing

Problems

- verbal descriptions used to capture perception
- natural sounds -> complex answers
synth. sounds -> little application value
- heterogenous learning base

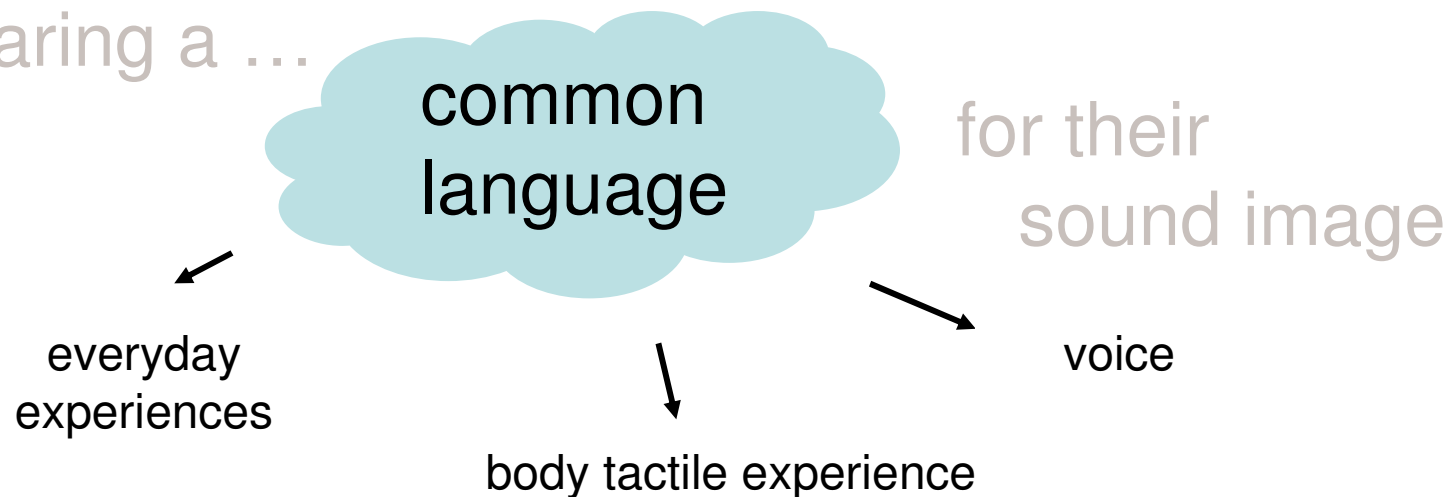
Trend: capture perception and emotion

A recent publication in ACTA 95

- Nykänen, Johansson, Berg, Lundberg:
perceptual dimensions of saxophone sounds
- 16 subjects started with 140 verbal descriptions
- after tests, ANOVA and PCA, 9 descriptors
revealed significance:
full-toned, rough, warm , soft, sharp, sharp/keen,
[a]-like, [o]-like, nasal





World of musicians ...

- makers of musical instruments, recording engineers, composers, ...
- sound-centered
- sharing a ...



A. Liebe 1952: investigating 1600 words, 16th to 19th century

Capable in terms of translation

- sound imitation,
translation,   imitation   original
classification

... learning the mother language

... fun of imitating people

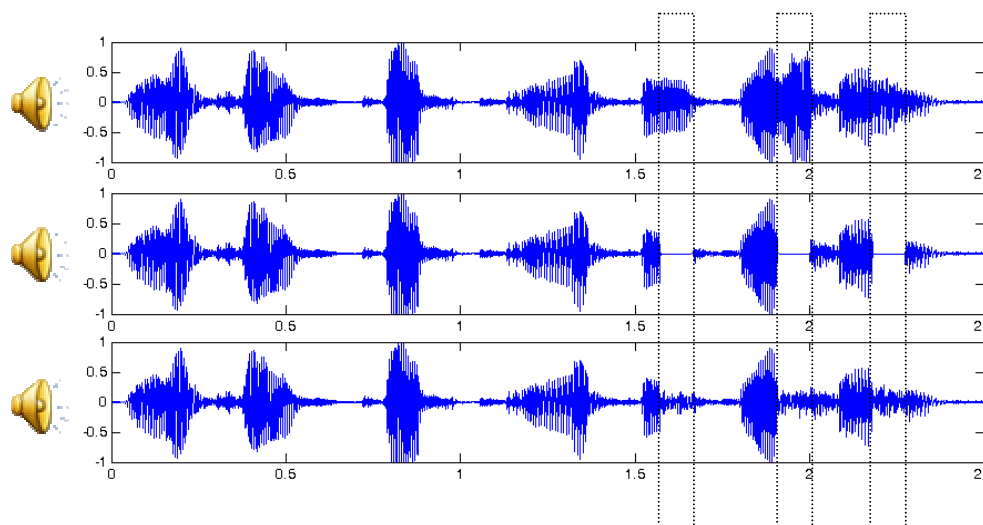
... imitating sound

Capable in terms of construction

- reconstruction,
construction,
imagination

[Visual Example](#)

[Visual Example](#)



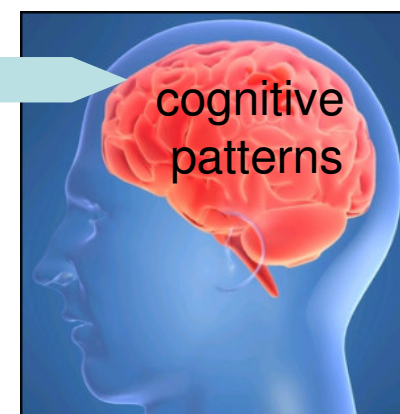
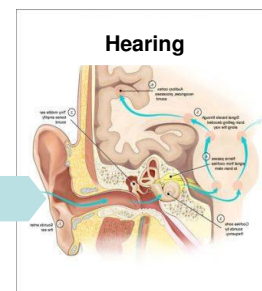
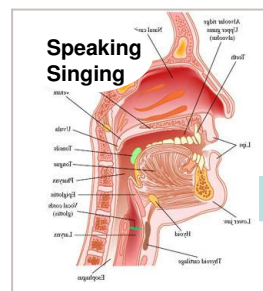
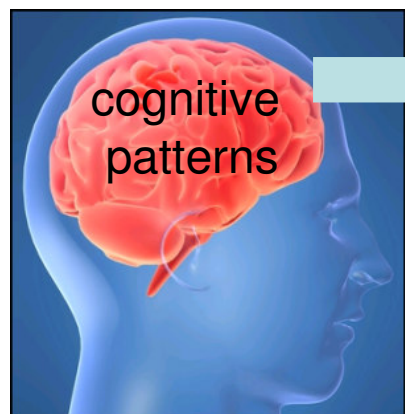
... immense training intensity

... start at age 0 yrs and continue >12h/d

Translation and construction

original sounds

learning base



physical
encoding

physical
decoding

cognitive
decoding

A few dimensions

high-level
semantic

musical and literal: symbolism, drama
interpretation, dynamics, irony, wit, ...

mid-level
features

vowel quality, nasality, articulation, rhythm
prosody, speaking – sprechgesang – singing

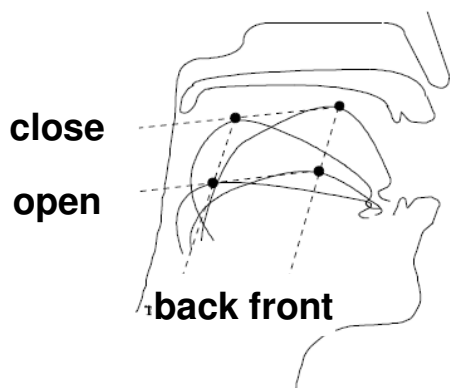
low-level
physical
properties

pitch, loudness, spectrum, formants,
vibrato, glottal flow, directivity

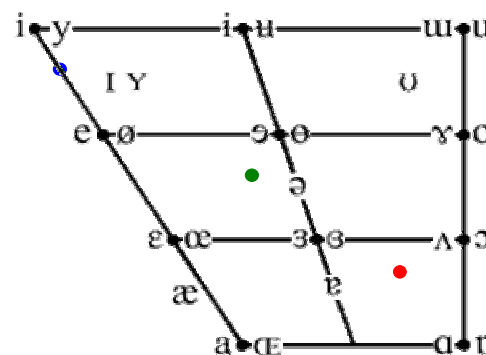
A few human voice features

- vowel quality
 - tongue position : backness and height
 - primary and secondary cardinal vowels:
rounded/unrounded
- nasality
 - open and / or closed
- singing character – speaking character

Vowel quality

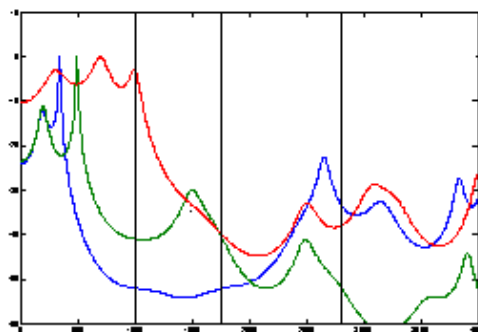


learning base



Jones diagram

physical
params



LPC-spectrum

cognition
model

How

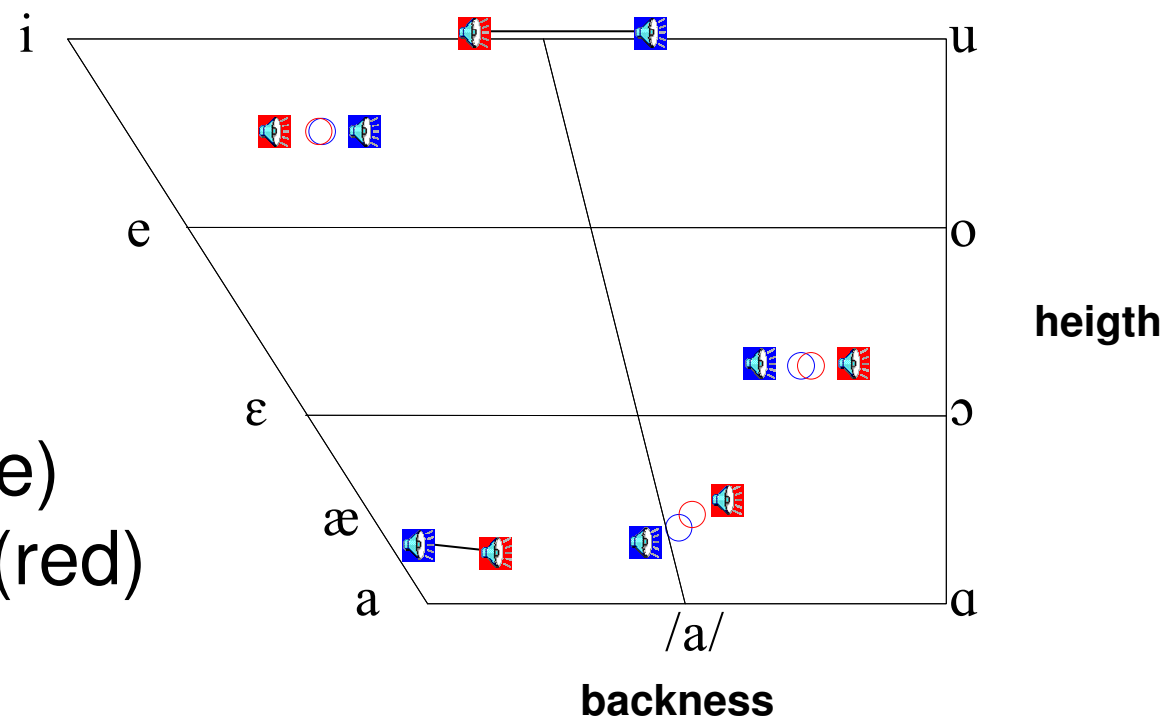
Verification

Vowel quality in violin sounds?

- 45 out of 120 randomly chosen violin sounds revealed vowel quality

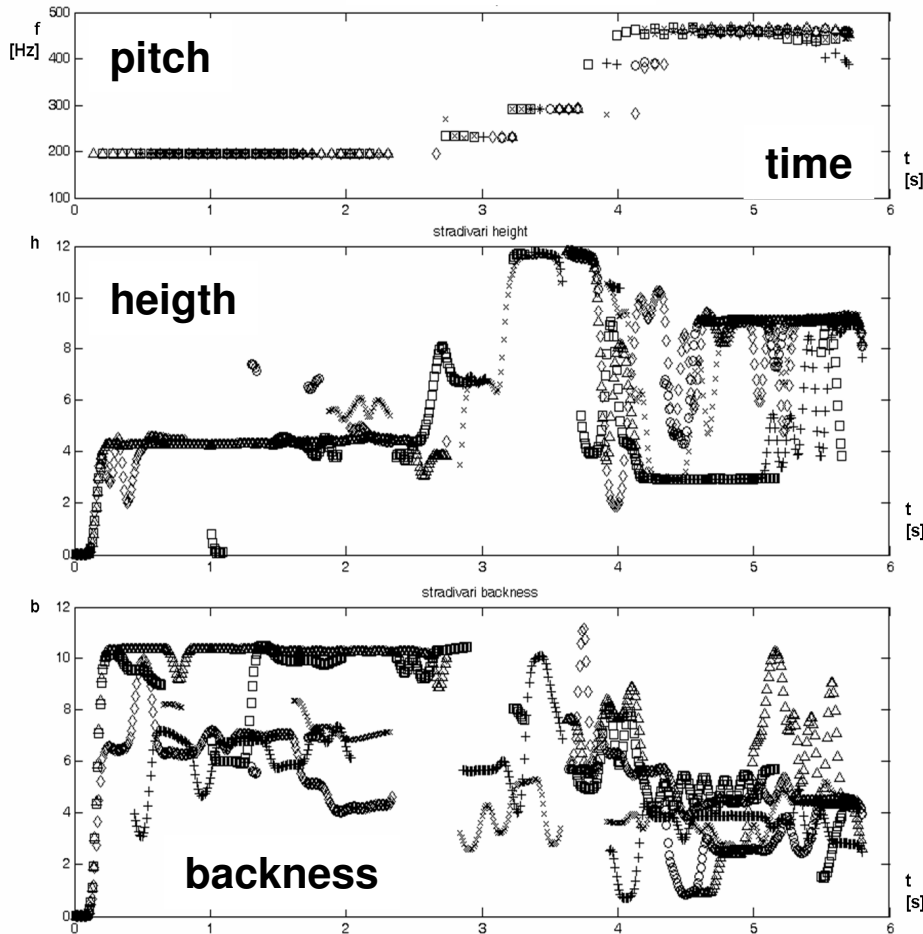
- violin (blue)
vs. voice (red)

Method

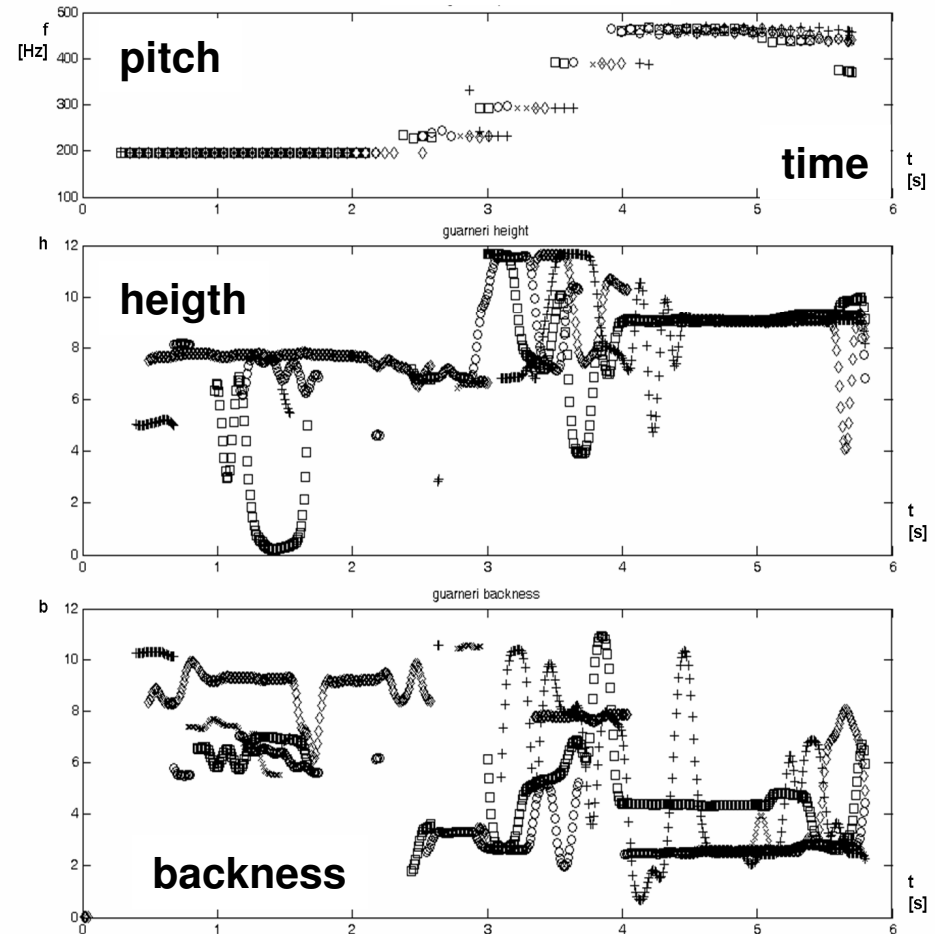


Vowel quality over time

6 x Stradivari

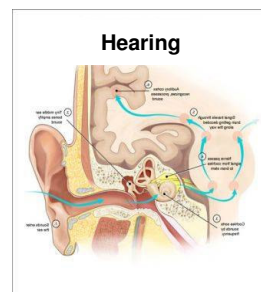
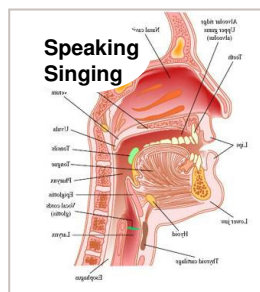


5 x Guarneri



Nasality

learning base



several physical ingredients

cognition model is a huge task

Nasality – medical perspective

- looks at: formants, anti-formants, bandw., power
- Baken, Orlikoff 2002: clinical measurements
Kersten thesis 2008: none of the 7 key ingredients triggered „nasality“ in subjects
- Zecevic PhD thesis 2002: > 3000 sounds, > 20 technical features extracted, HMM, 70% correct classification on [0123]-scale for open nasality, women, [a]

Nasality – speech processing community

- Chen 2000 or Pruthi 2007 look at extra poles P0 and P1, usually hidden behind F0, F1, F2
- Malhotra thesis 2009: concentration on time-independent features (from warped LPC):
bandwidth F1, P0, P1
amplitude differences of P0, P1 vs. F1
frequency relations of P0, P1 vs. F1
- 75% correct classification with a single feature on [01]-scale for open nasality, women, [a]

Sparseness

- vowel quality code entropy

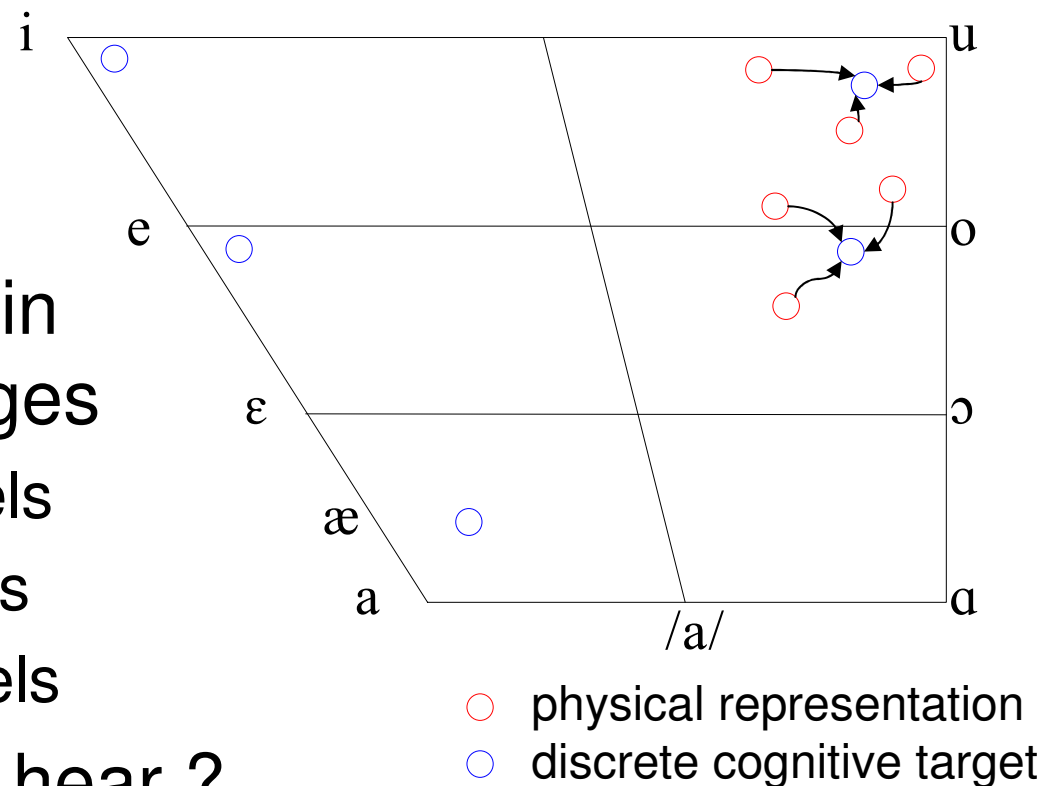
some 10 kbyte of half a second sound
condensed to 1 byte

Intellegibility

- understandable representation
- with / without computer

Link to Ethnological Musicology

- lock-in
- population of discrete vowel qualities in various languages
 - most: 3-4 vowels
 - few: >10 vowels
 - Sedan 55 vowels
- so what will we hear ?



Summary

- encouragement to use human voice features for representations of short steady-state sounds
- examples demonstrate the translation and construction capabilities of human voice
- these capabilities directly link to cognitive targets and add value to the sparsely encoded physical representation

- Aures, W.:** Berechnungsverfahren für den sensorischen Wohlklang beliebiger Schallsignale, *Acustica* 59, S.130, 1985.
- Baken, R. J., Orlikoff, R. V.:** *Clinical measurements of Voice and Speech*, Singular Publications, 2002.
- Beauchamp J., (1982):** Synthesis by spectral amplitude and "Brightness" matching of analyzed musical instrument tones. *J. Acoust. Eng. Soc.*, 30(6): 396-406.
- Bismarck, G. von (1972):** Extraktion und Messung von Merkmalen der Klangfarbenwahrnehmung stationärer Schalle, Diss., In: Mitteilung a. d. Sonderforschungsbereich 50 "Kybernetik", München.
- Gordon, J., and Grey, J. M. (1978):** Perceptual Effects of Spectral Modifications on Orchestral Instrument Tones." *Computer Music Journal*, Vol. 2, N° 1, pp. 24-31.
- Kersten, J., Sprechen versus Singen - eine Klanganalyse an Musikinstrumenten, Diplome Thesis, faculty DMI, HAW, Hamburg, Apr. 2008.**
- Liebe, A., : Die Leistung der deutschen Sprache zur Wesensbestimmung des Tones, Habilitationsschrift, Berlin, 1958.**
- Moore, B. C. J., Glasberg, B. R. and Baer, T. (1997).** "A Model for the Prediction of Thresholds, Loudness, and Partial Loudness." *Journal of the Audio Engineering Society* 45(4): 224-240.
- Müller, S.:** Vokale in Klängen – eine LPC-basierte Extraktion der Vokalqualität zur Darstellung von Violinenklängen im Vokaldiagramm, Diplome Thesis, faculty DMI, HAW, Hamburg, Nov. 2007.
- Daniel, P.:** Berechnung und kategoriale Beurteilung der Rauhigkeit und Unangenehmheit von synthetischen und technischen Schallen, Dissertation (C.v.O. Universität Oldenburg 1995)
- Ricci, R.:** *The Glory of Cremona*, Compac Disc, MCA Records, 1989.
- Sottek, R.** Gehörgerechte Rauhigkeitsberechnung DAGA 1994, Dresden
- Zwicker, E. and H. Fastl (1999).** *Psychoacoustics: Facts and Models*. Berlin, Springer-Verlag.

