Physics in Phonetic Measurement and Analysis: Illustrations from Electromagnetic Articulography and other Experimental Techniques

Andreas Zierdt

EKN, *Entwicklungsgruppe Klinische Neuropsychologie,*
Dachauerstrasse 164, 80992 München
Klinikum Bogenhausen, Städtischen Klinikums München GmbH

24. November 2008

Is this physics?

- Articulatory phonetics investigate how humans produce speech sounds
- Measurement focus on different structures of the vocal tract, called the articulators (tongue, lips, jaw, palate, teeth)
- While sound production is closely related to aerodynamics, investigating articulators generally means imaging or point tracking and recording of sensor trajectories
- The challenge is to analysis this movements in an noninvasive way and without too much distraction of the speech
- This talk will cover one special measurement system, which uses EM-Fields to track movements inside the mouth
- Development needed some theory on electromagnetism and a lot on numerical analysis.

Overview

- Use of electromagnetic articulography, a timetable
- What means 2D, 3D and 5D-EMA ?
- Measurement principle and system outline
- Position calculation: Field model and some numerical methods
- Problems inherent to the measurement principle
- Outlook

Electromagnetic Articulography

- 1974 Sonoda uses a permanent magnet for lip movement
- 1977 van der Giet introduces EM-System, sagittal (2D) plane
- 1980 Perkell miniaturizes sensors, enabling in mouth EMMA
- 1983 Schönle develops 3T-System → AG100 (1st commercial)
- 1987 Panagos uses Helmholtz pair as transmitter
- 1993 Zierdt invents method for 3D-Measurment
- 1995 Start of development in Munich (IPS, Fa Carstens, NTT)
- 2000 First 3D-Measurment on human subject
- 2002 Fa Carstens sells 1st 3D-Device (AG500)
- 2003/2004 New calibration unit, device development finished
- Since 2005, about 50 AG500 in use at different places
- 2007 12 channels, 200 Hz and real-time display.

2D-EMMA

- 3 transmitter coils forming a triangle around the head
- U ~ 1/r^3 → r ~ U^3
- U ~ 1/r^3 * cos(β)
- Measured in the subjects mid-sagittal plane, where β is equal for all coils
- Errors occurred, when sensor was leaving the plane while twisted
- Subject had to wear the device.

Advantages of 3D-EMA

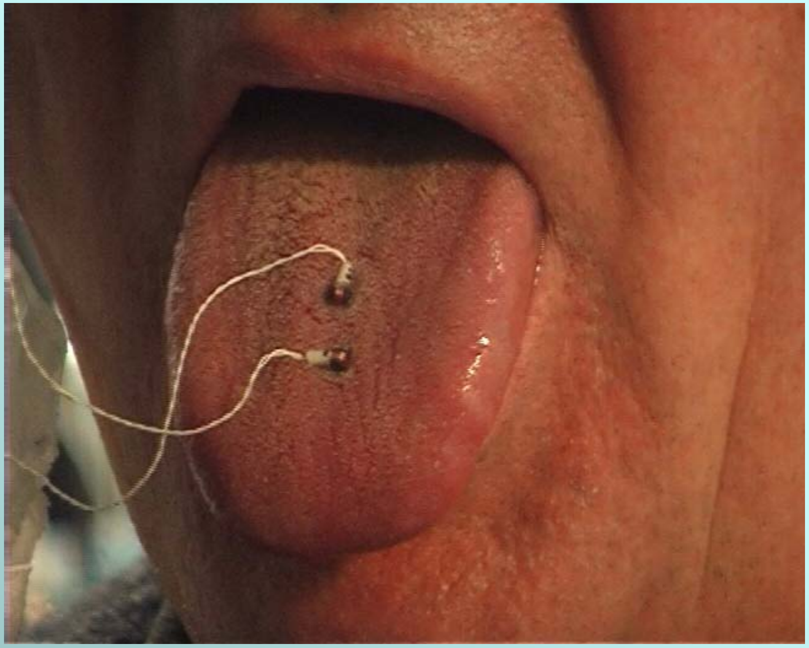
- Fleshpoint-Measurment
  - high temporal resolution (200 Hz)
  - selective analysis of points of interest, i.e. tongue-tip
  - gets trajectory with velocity and acceleration (forces)
  - automatic processing of large data sets possible (Corpora)
- non-invasive, field-strength in the range of earth (µT)
- free head movement
  - more natural position
  - non-fatigue, enables longer recordings (4h)
- nearly soundless, good for acoustic recordings
- spatial measurement
  - lateral tongue movement
  - laterally placed sensors: jaw, corner of the mouth
  - complex jaw movements, facial expression, head movements
- truly it’s 5D, not 3D

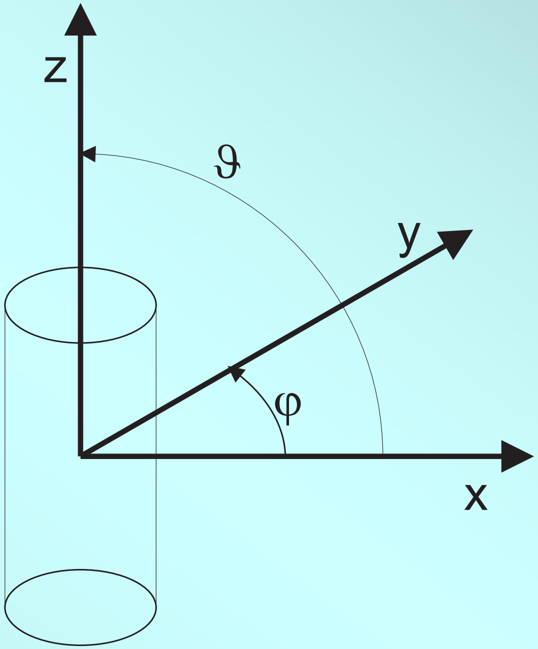
Advantages of 5D-EMA

- spatial sensor orientation (two angles)
  - was not requested in the beginning, but necessary to compute
  - other systems are using much larger sensors
  - same reasonable priced sensors as AG100 (disposable)
- more information per sensor
  - number of sensors cannot be arbitrarily high since
    - articulation disturbance
    - cross-interference of the sensors
  - less sensors necessary for head movement
- sensor directly gains tangential gradient
  - yaw opening
  - curvature of the tongue .

EMA-Sensors

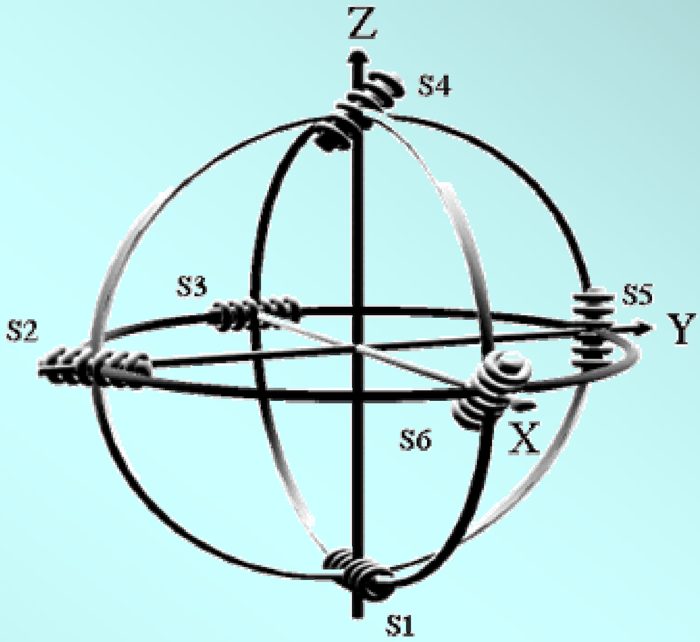
- each sensor has 5 DOF (x, y, z, φ, θ)



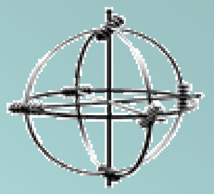


5D EMA, Measurement principle

- a spherical arrangement of 6 transmitter coils, radiating VLF fields with different frequencies
- sensor coils on the tongue receiving field strength as function of their position
- signal-demodulation separates signal-components
- sensor position is calculated from these 6 signals.



Calculation the position



- sensors are independently processed
- 5-D (x, y, z, φ, θ) state-vector $\underline{s}$
- 6-D value-vector $\underline{v}$ per state (6 Transmitters, signalamps)
- field model f describes relation: $\underline{f}(\underline{s}) = \underline{v}$
- $\underline{f}$ is a system of equations, one per transmitter
- distribution of magnetic field → non-linear equations
- finding sensor position is an *inverse problem*: $\underline{s} = \underline{f}^{-1}(\underline{v})$
- numerical solution of the non-linear equation set
- similar problems: geophysics, medical imaging, astronomy.

Solving non-linear equations

Example:

$(x-2)^3 + x = 3$

Newton method:

$x_0 = 0$

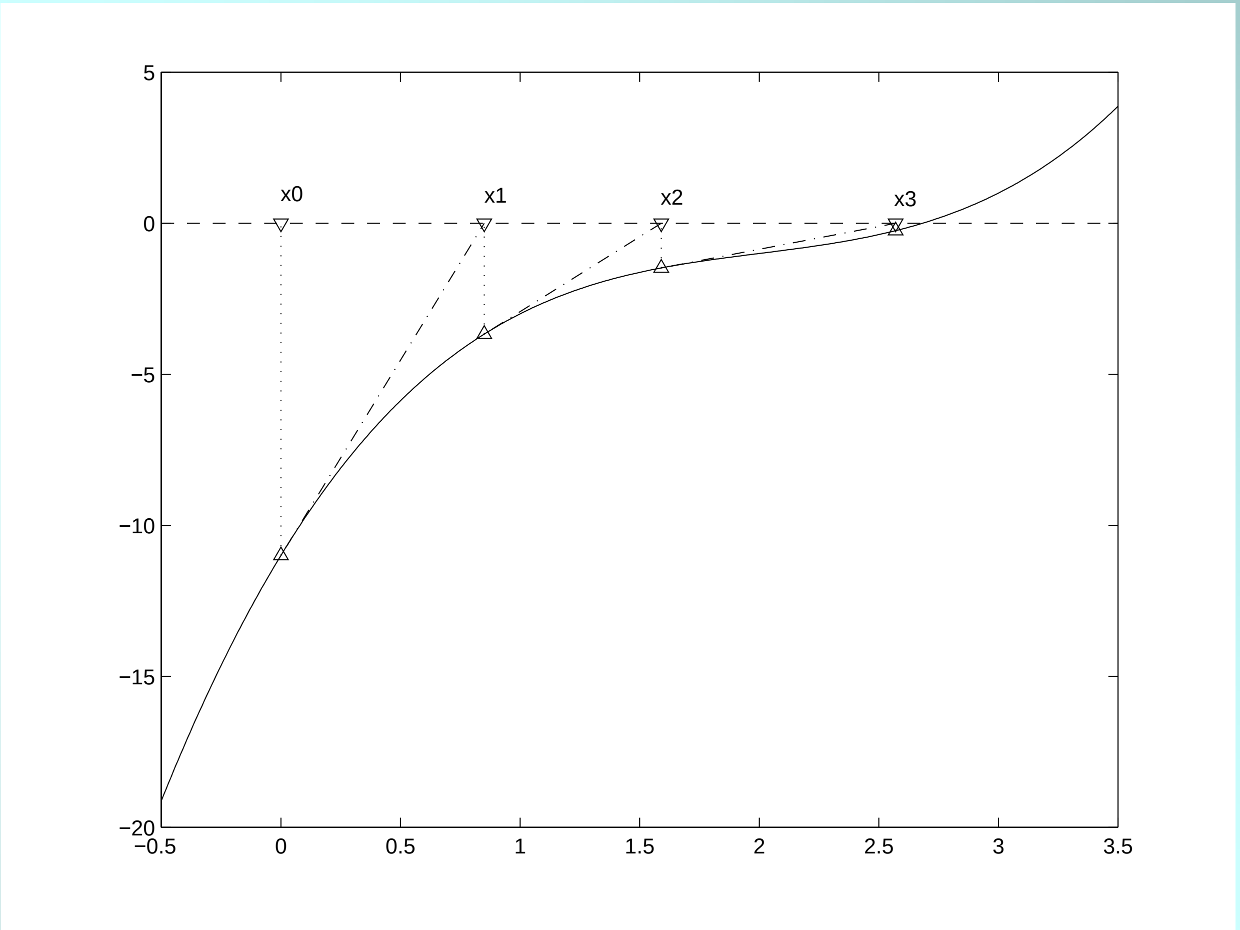
$x_1 = 0{,}9$

$x_2 = 1{,}6$

$x_3 = 2{,}6$

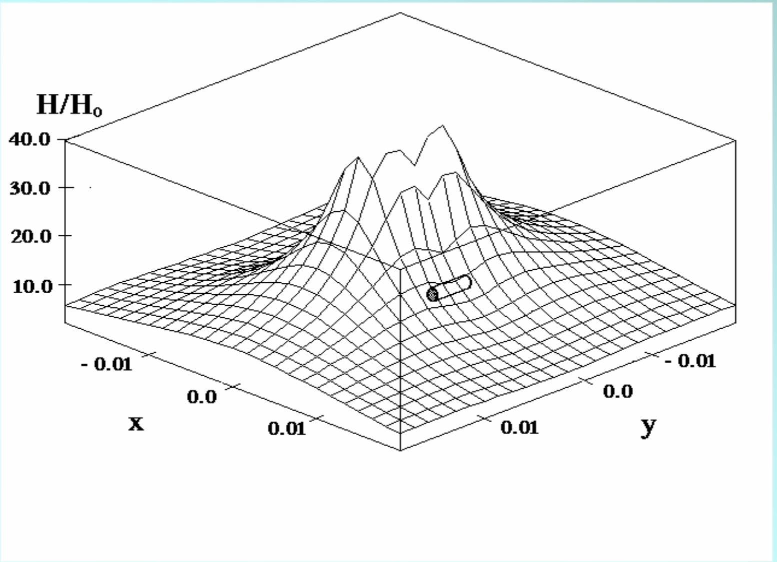
Solution:

$x = 2{,}682$



Field model (1)

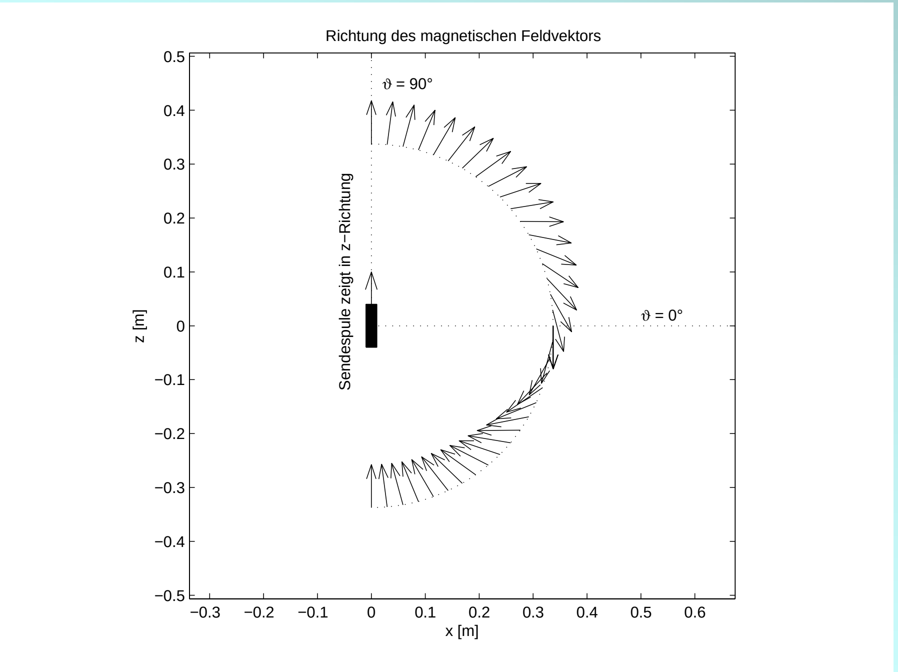
- spatial distribution follows field of a dipole
- cubic function of distance r
- large dynamic range
- field near poles twice as strong as between (Gaußsche Hauptlagen)
- function is highly non-linear.



$$f(\varphi, \vartheta, r) = \frac{1}{r^3} \cdot \sqrt{1 + 3 \cdot \sin^2(\vartheta)}$$

Field model (2)

- field has vector character
- signal depends on the orientation of the sensor
- field is symmetrical around the coil axis
- signal becomes zero when sensor is perpendicular to field vector
- sensor orientation defines signal sign
- v ~ cos(β).



Position calculation with Matlab

- overdetermined system of equations $\underline{f}$
- optimization problem → roots of the derivative
- Gauß-Newton-Method with error square
- Levenberg-Marquardt slightly more stable
- only local convergence is assured ("stability")
  - estimation of most stable sample of a trial (start-position)
  - tongue moves max. 1.5mm between samples (300mm/s), increasing Fs makes position calculation more easy
  - Method might find secondary minima, solution not necessarily true position, thus verification required (residuals, iteration depth)
  - start-position can be found by means of specialized (slow) methods.

Problems inherent to the system

- dipole fields are symmetrical in terms of φ, orientation can only partly be determined (5 out of 6 DOF)
- direction of current is arbitrary, “hot” and “cold” end of sensor not distinguishable → “4.5D”-EMA
- dynamic range more than 80 dB
  - thermal noise of the pre-amplifiers, electrical interspersion
  - clipping can occur, 16 Bit ADC, 16 significant bits are a good deal
- field model $\underline{f}(\underline{s}) = \underline{v}$ is highly non-linear
  - functional space is like “high mountain range”, steep and pseudo-discontinuous
  - hard to estimate accuracy
- only local linearization(Jacobi-Matrix) possible
  - numerical noise by round-off errors, ill-conditioned problem
- 5 dimensions and high resolution → huge co-domain
  - tables will not help here
- links Euclidean geometry with spherical geometry
  - projections are never distortion-free (like maps).

Outlook

- Increased sample rate (400 Hz) will improve convergence behavior
- Real-time processing of the data
- Use of alternative algorithms like the Kalman filter or sequential Monte Carlo methods (EKF, SMC). Such particle filters could help to overcome the limitations of the noisy and ill-conditioned system
- Another approach is the utilization of physical, physiological and even phonetic constraints, e.g. spline-filtering
- Standardization of data representation and analysis
- Wireless measurement.