Elmar Nöth, Tobis Bocklet, Arnd Gebhard, Tino Haderlein, Andreas Maier, Korbinian Riedhammer, Stefan Steidl, Maria Schuster

Medical Speech Processing - Pathologies, Treatment Assistance, Clinical Trials

Sunday, September 26th, 2010
Das ist Frosch

Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, ...
Outline

- Motivation
- Examples for disorders in speech production
- Speech technology and its use in established tests
- Ongoing and planned projects
  - Linguistic level that needs to be modeled
  - Results (where already available)
- Peaks: a client server architecture
- Outlook: emotional disorders & multimodality
- Summary
Motivation

Disorders in Speech Production

- Dysfunction of the neurological processing → **language** disorders
- Dysfunction of excitation → **voice** disorders
- Dysfunction in articulation → **speech** disorders
Motivation

Disorders in Speech Production

- Dysfunction in articulation → speech disorders
Motivation

Disorders in Speech Production

- Dysfunction of excitation $\rightarrow$ voice disorders
Motivation

Necessity of Evaluation

- **Diagnosis**
  - How intelligible is the patient? (holistic impression)
  - How strongly does the patient nasalize? (distinct aspect)
- **Therapy control**
  - Has the situation of the patient improved during therapy?
- **Comparison of therapy methods**
  - Which therapy method leads to the best results for a group of patients?
- **Screening**
  - Is the quality of a child’s speech according to its age?
- **Computer-assisted therapy**
  - Did the patient perform the exercise correctly?
Motivation

Evaluation of Pathologic Speech

- **Problem:**
  - So far no objective, validated, and simple method to evaluate language, voice, and speech disorders
  - In clinical routine: subjective, perceptive judgment (1 listener)
  - “Objectification” via $n$ judges only possible for research studies due to cost constraints

- **Solution:**
  - Automatic speech analysis as an objective method to evaluate and quantify the disorders (second opinion)
Outline

- Motivation
- **Examples for disorders in speech production**
  - Speech technology and its use in established tests
  - Ongoing and planned projects
    - Linguistic level that needs to be modeled
    - Results (where already available)
  - Peaks: a client server architecture
- Outlook: emotional disorders & multimodality
- Summary
Examples for Disorders in Speech Production

- Learning disorders
- Sigmatism
- Articulation disorders of children with cleft lip and palate
- Stuttering
- Alaryngeal (substitute) voice
- Dysglossia (voice after removal of malign tissue in the oral cavity)
- Dysarthria
- Aphasia and dementia
Examples for Disorders

Learning Disorders

- Dyslexia (reading disability) - most common learning disability
  - Difficulty with phonemic awareness
  - Difficulty with matching letter combinations to specific sounds
- Disorders of speaking and listening
  - (Writing disability)
  - (Math disability)
Examples for Disorders

Sigmatism

- Phonetic disorder (mispronunciation of /s/ & /z/)
- Intelligibility is usually not compromised
- Still normal until the age of 5 years or during second dentition
- Wrong positioning of lips, tongue & teeth

---

interdental

addental

lateral
Examples for Disorders

Cleft Lip and Palate (CLP)

- Structural malformations of
  - Nose
  - Throat
  - Mouth
  - Jaw
- Negative effects on
  - Respiration
  - Nutrition
  - Hearing
  - Speaking
  - Psychosocial competence
- Prevalence: 1 : 500-700
Examples for Disorders

Stuttering

- Repetition type
  - Sequences of fast contractions of the vocal cords
  - Repetitions of phonemes, syllables, words, phrases
  - Example: *t-t-t-t-table*

- Blocking type
  - Long-lasting contractions of the vocal cords
  - Pauses occurring before or within words
  - After the spasmic plosion rising tone pitch
  - Example: *t---able*

- Often combined

- Prevalence: 5% of children, becoming chronic in ca. 20%
Examples for Disorders

Repeating Stuttering

Examples for Disorders

Alaryngeal (Substitute) Voice

- Removal of the larynx due to cancer
- Breathing is detoured through a tracheostoma
Examples for Disorders

Alaryngeal (Substitute) Voice

- Removal of the larynx due to cancer
- Breathing is detoured through the tracheostoma
- Speaking is enabled by a substitute voice
Examples for Disorders

Oral Squamous Cell Carcinoma

- Origin: mucosa of the oral cavity
- Causes: nicotine, alcohol, low standard of oral hygiene
- Therapy: resection of the tumor and reconstruction
- Effect: reduced mobility & sensibility, asymmetry, and change of resonance → reduced speech abilities
Dysarthria

- A speech disorder affecting the coordination of muscles in the vocal tract, face, larynx, and respiratory system (dysarthrophonia).

- Mostly results from a neurological injury, such as a stroke or other kind of brain injury.
Examples for Disorders

Dysarthria

talking with intermittent Dysarthria
Examples for Disorders

Aphasia

- Different forms and combinations
- Variable symptoms:
  - Inability to
    - Comprehend language
    - Speak spontaneously
    - Form words
    - Name objects
    - Read & write
      - Inability to repeat a phrase or persistent repetition of phrases
    - Paraphasia, neologisms
    - Agrammatism
    - Dysprosody
Examples for Disorders

Wernicke’s Aphasia: Answering Interview Questions

Examiner: *I’m going to ask you some questions, and I just want you to answer “yes” or “no”. Okay? “Yes” or “no”. Is your name Smith?*

Patient: *Where would I be, what they’re eating avarment I don’t know.*

Examiner: *Is your name Brown?*

Patient: *Oh mistress triangland while listen you walking well things things this for year for thee.*

Examiner: *Okay, just say “yes” or “no”. Is your name Brown?*
Examples for Disorders

Wernicke’s Aphasia: Naming Objects

Examiner: *Tell me the names of each of these.*

Patient:

[cigarette] *This is a cigarette.*

[comb] *A wongt.*

[fork] *A filt.*

[key] *A wote.*
Wernicke’s Aphasia: Repeating Words

Examiner: *Now I’ll say the name of each one, and then you say it after me. Okay? Toothbrush.*
Patient: *Stocktery.*

Examiner: *Cigarette.*
Patient: *Cigarette.*

Examiner: *Pen.*
Patient: *Tankt.*

Examiner: *Knife.*
Patient: *Nike.*
## Examples for Disorders

### Disorders in Speech Production

<table>
<thead>
<tr>
<th></th>
<th>Voice</th>
<th>Articulation</th>
<th>Semantics Lexicon</th>
<th>Morphology Syntax</th>
<th>Pragmatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developmental Speech Disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sigmatism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleft Lip and Palate</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stuttering</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>Alaryngeal voice</td>
<td></td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer of the Oral Cavity</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dysarthria</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aphasia &amp; Dementia</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Red: affected  
Pink: Can be affected  
* : Presbyphonia
Outline

- Motivation
- Examples for disorders in speech production
- **Speech technology and its use in established tests**
- Ongoing and planned projects
  - Linguistic level that needs to be modeled
  - Results (where already available)
- Peaks: a client server architecture
- Outlook: emotional disorders & multimodality
- Summary
Speech Technology and Tests

Speech Technology

- Automatic speech processing methods
  - Word and phoneme recognition
  - Acoustic speaker modeling
  - Prosodic analysis
  - Evaluation measures
  - Visualization
- Application of methods in different kinds of established tests
- Off-the-shelf technology
- Semi-continuous HMMs
  - Easier to adapt with small amounts of data
  - Comparable results with continuous models
  - 11 Mel cepstrum coefficients + energy + 1. derivative
Speech Technology and Tests

Word and Phoneme Recognition

- **Acoustic modeling** depends on patient population (children need different AM than senior speakers)
- Adaptation of AM by MAP/MLLR can be useful (simulation of an expert listener)
- **Language modeling** depends on kind of spoken text (known/unknown or spontaneous speech)
  - Known text: restriction of vocabulary to words of test
  - Text unknown to raters: medium size vocabulary
  - Unigram language model (emphasis of acoustic information)
  - Spontaneous speech, e.g., description of image sequence: large vocabulary, restriction to scenario, higher-order language models
Acoustic Speaker Modeling

**Idea:**
- Acoustic space of speakers can be modeled
- Space represents the multidimensional characteristics of voice of a speaker
- Degree of pathology varies in acoustic space
- Find characteristics of degree of speech disorder

**Approach:**
- Acoustics modeled by Gaussian Mixture Models (GMMs)
- Train Universal Background Model (UBM) with normal speakers
- Train GMM of path. speakers and transform into vector
- Perform a classification/regression (depends on the task)
- Variations of speakers with different degrees of pathology
- Can be modeled by adaptation from UBM to GMM
Gaussian densities \((i = 1, \ldots, N)\) of speaker model defined by mean values \((m_i)\) and covariance matrices \((K_i)\)

\[
\mathbf{m}^s = \begin{bmatrix}
m_1 \\
m_2 \\
m_3 \\
m_4 \\
m_5 \\
m_6 \\
\end{bmatrix}
\]

\[
\mathbf{K}^s = \begin{bmatrix}
K_1 \\
K_2 \\
K_3 \\
K_4 \\
K_5 \\
K_6 \\
\end{bmatrix}
\]
Acoustic Speaker Modeling

- Discriminate between different types of pathology

  - Create SVs of speakers
  - Train some classifier on labeled SVs
  - Create SV of test speaker
  - Classify SV of test speaker

Points correspond to supervectors (SVs)

- Blue points represent speakers with pathology type 1
- Red points represent speakers with pathology type 2
Estimate degree of pathology

Train a regression (linear/SVR)
Create SV for a test speaker
Estimate degree of pathology
Prosodic Analysis

- Prosody: rhythm, intonation, stress, and related attributes
- Computation of prosodic features on word level, across several words or across syllable nuclei
- Computation across several words requires ASR
- Computation across syllable nuclei requires syllable detection
- **Local features:**
  - Pauses before/after segments, signal energy, word duration, and F0
  - Calculation of mean, max., min., and std. dev.
- **Global features:** jitter, shimmer, voiced/unvoiced characteristics
  \[ \approx 100-200 \text{ features per test utterance} \]
Computation of prosodic features

1. onset
2. onset position
3. offset
4. offset position
5. maximum
6. position of maximum
7. minimum
8. position of minimum
9. regression line
10. error of the regression line

voiceless sections

reference point
Speech Technology and Tests

Prosodic Parameters

-3 Einst
-2 stritten
-1 sich
0 Nordwind
1 und
2 Sonne

<feature_name>-2,-1
<feature_name>0,0
<feature_name>1,2

reference point
Word accuracy (WA) and word correctness (WC)
Calculated features
- Features of acoustic speaker models
- Features of prosodic analysis
Correlation (Pearson & Spearman) based on calculated features or WA, WC with human listener
Classification based on calculated features
Interpretation of relevant features after feature selection
Word accuracy WA:

\[ WA = (1 - \frac{S+D+I}{N}) \cdot 100\% \]

Word correctness WC:

\[ WC = (1 - \frac{S+D}{N}) \cdot 100\% \]

N: number of **spoken** words

S: number of **substituted** words

D: number of **deleted** words

I: number of **inserted** words
Speaker can be seen as point in the $\mathbb{R}^N$ based on the produced text.

- Each dimension given by the value of one acoustic or prosodic feature (e.g. mean word duration)
- Projection onto 2-D or 3-D, e.g., *Principal component analysis*, *Kohonen map* or *Sammon transformation*
Speech Technology and Tests

Adding Old Reference Speakers
Adding Laryngectomees
Speech Technology and Tests

Adding Chronically Hoarse Speakers

- alaryngeal voice
- young reference speakers
- hoarse speakers
- old reference speakers
Speech Technology and Tests

What Do We Model?

- Isolated vowels
  - Phoneme recognition
  - Acoustic and prosodic features
- Unknown text (naïve listener)
  - Word recognition
- Known text (expert listener)
  - Word recognition
  - Acoustic speaker modeling
  - Prosodic analysis
- Spontaneous speech
  - Prosodic analysis
  - Syntactic/semantic analysis based on word recognition
Isolated Vowels

- Distinctive aspect (hoarseness, creakiness)
- No intelligibility (holistic impression)

**Example**: Hoarseness Diagram

- Analyzing voice quality w.r.t. hoarseness
- Based on sustained vowels
- Two-dimensional representation of voice quality:
  - X-Axis: degree of “irregularity” (based on jitter, shimmer and short-time cross-correlation of adjacent cyclic periods)
  - Y-Axis: noise, amount of pulselike vs. noisy excitation (based on Glottal to Noise Excitation Ratio (GNE))
Speech Technology and Tests

Isolated Vowels – Hoarseness Diagram

- Aspirated voice
- Rough voice
- Normal voice

The diagram plots noise levels against irregularity, with points scattered across the graph indicating different voice qualities.

- Aspirated voice is located in the upper right quadrant, suggesting high irregularity and high noise.
- Rough voice is found in the lower right quadrant, showing high irregularity but lower noise.
- Normal voice is in the lower left quadrant, indicating low noise and low irregularity.

The dashed ellipses provide a visual representation of the typical regions for each voice quality.
- Only intelligibility (holistic impression)
- Naïve listener writes down what he heard
- Carefully selected sub-corpus (e.g. 10 out of 400 sentences or isolated monosyllabic words)
- Low predictability for human listener
- Listener can only perform test until he has heard each sub-corpus once
- Board of listeners transcribes utterances
- Human evaluation criteria: word correctness
- Can be modeled directly with ASR
- Word test: monosyllabic words
  - nine
  - team
  - rock
- Sentence test
  - He has played very well.
  - The typewriter needs to be repaired.
  - This is a good place, though small.
  - She doesn’t like to have it hanging around.
  - He dashed across the parking lot and disappeared inside.
Post-Laryngectomy Telephone Test (PLTT)

- German test for measuring intelligibility after laryngectomy
- One speaker and one naïve listener, contact via telephone
- 20 (of 400) monosyllabic words, 5 (of 100) sentences
- Word intelligibility (5*number of understood words, 0 … 100)
- Sentence intelligibility: 2 points if understood completely correct, 1 point if one word is missing or incorrect → 10*sentence points (0 … 100)
- Total intelligibility: 0.5*word intell. + 0.5*sentence intell.
PLTT (Human Raters)

11 naïve raters; correlation to average of other 10 raters: $0.83 \leq \rho \leq 0.94$
Speech Technology and Tests

Measuring Intelligibility with Unknown Text

- Models communication situation in everyday life best
- Needs a large panel of naïve listeners
- Needs continuous supply of naïve listeners
- Automation: same as for known text
- Creation of ground truth is different: naïve vs. expert

- Direct comparability and evaluation of distinct aspects is more difficult than with known text
- Intelligibility and distinctive aspects (e.g. nasalization)
- Expert listener judges with Likert or analog scale
- Can be used for longitudinal/follow-up study
- Direct comparability of speakers
- Automatic system needs ground truth, i.e., intelligibility estimation of expert listeners
- System may need large corpus of age-, gender-, …-matched patients
“Representative” texts (phoneme distribution similar to that of the language per se)

- The North wind and the Sun (44 languages)
  http://www.aesoplanguagebank.com/
- The Rainbow Passage
- Comma Gets a Cure

Texts for one pathology, e.g., high portion of nasal sounds

- Isolated words
- Age matched texts, e.g., vocabulary for 1\textsuperscript{st} grade
- Pictograms, e.g., for children who cannot read
Evaluation of the audio data by 5 speech experts
- For holistic (intelligibility) and distinct aspects (e.g. nasality)
- On a scale from 1 to 5
- For each turn (sentence/pictograms)
- Averaging for each speaker leads to a continuous scale from 1 to 5
Perceptive evaluation by 5 experts

41 laryngectomees, correlation according to Spearman:

<table>
<thead>
<tr>
<th>expert $i$ vs.</th>
<th>$E{1}$</th>
<th>$E{2}$</th>
<th>$E{3}$</th>
<th>$E{4}$</th>
<th>$E{5}$</th>
<th>$\emptyset$</th>
<th>$E = {1,\ldots,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.80</td>
<td>.87</td>
<td>.80</td>
<td>.84</td>
<td>.81</td>
<td>.82</td>
<td></td>
</tr>
</tbody>
</table>
Perceptive evaluation by 5 experts

31 CLP children, correlation according to Spearman:

<table>
<thead>
<tr>
<th>expert $i$ vs.</th>
<th>$E{1}$</th>
<th>$E{2}$</th>
<th>$E{3}$</th>
<th>$E{4}$</th>
<th>$E{5}$</th>
<th>$\emptyset$</th>
<th>$E = {1,\ldots,5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E{1}$</td>
<td>$.92$</td>
<td>$.93$</td>
<td>$.93$</td>
<td>$.92$</td>
<td>$.92$</td>
<td>$.92$</td>
<td>.92</td>
</tr>
</tbody>
</table>

- No significant difference between Pearson’s ($r$) and Spearman’s ($\rho$) correlation coefficient
- Typical correlation values: $.80 \leq r, \rho \leq .95$
Speech Technology and Tests

Use of Speech Technology with Known Text

- Intelligibility
  - Word accuracy / word correctness
  - Prosodic analysis
  - Acoustic speaker modeling
- Distinctive aspects
  - Effort – Prosodic analysis, e.g., pauses between words
  - Stuttering – Phoneme accuracy
- Screening
  - Comparison with control group
Spontaneous Speech

- Topic is given: restricted, yet open vocabulary
  - Examples:
    - „Name all animals you can think of“
    - „Describe the content of these pictures“
- Needs transcriptions (based on ASR)

- Syntactic structure (e.g. use of three-word sentences, POS sequences)
- Semantic coherence, e.g., of the named animals
  (cat, mouse vs. cat, gorilla)

So far little research based on automatic transcription
Outline

- Motivation
- Examples for disorders in speech production
- Speech technology and its use in established tests
- Ongoing and planned projects
  - Linguistic level that needs to be modeled
  - Results (where already available)
- Peaks: a client server architecture
- Outlook: emotional disorders & multimodality
- Summary
Pathologic reduction in reading ability

Reading ability depends on many factors:

- Reading speed (fluency)
- Reading errors (accuracy)
- Intelligence
- Age

All factors have to be considered in order to diagnose a reading disorder.

All factors are inter-related.
- Children are confronted with an unknown text which has to be read:
  - Usually a test consists of different sub-tests
  - Each sub-test is used to test for different aspects of the reading ability
  - Sub-tests are designed for different age groups
- Evaluation procedure:
  - Therapist marks every erroneous word
  - Measurement of reading time
  - IQ of the child is determined
The IQ is crucial for the diagnosis:

- Based on the IQ an age-dependent percentage rank of the child’s IQ is determined.
- Depending on the percentage rank age-dependent time and errors limits are obtained.
- If one of both limits is exceeded the child is diagnosed as pathologic.

<table>
<thead>
<tr>
<th>IQ Ranking</th>
<th>Limit Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>IQ = 85</td>
<td>5 reading errors</td>
</tr>
<tr>
<td>35%</td>
<td>82 seconds</td>
</tr>
<tr>
<td>35%</td>
<td></td>
</tr>
</tbody>
</table>
38 children recorded (12 girls and 26 boys)
All children were speculated to have a reading disorder
Age distribution:

<table>
<thead>
<tr>
<th>group</th>
<th>#</th>
<th>mean</th>
<th>std. dev.</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>38</td>
<td>9.7</td>
<td>0.9</td>
<td>7.8</td>
<td>11.3</td>
</tr>
<tr>
<td>girls</td>
<td>12</td>
<td>10.2</td>
<td>0.7</td>
<td>9.0</td>
<td>11.3</td>
</tr>
<tr>
<td>boys</td>
<td>26</td>
<td>9.5</td>
<td>0.9</td>
<td>7.8</td>
<td>11.3</td>
</tr>
</tbody>
</table>
Ongoing Projects

Corpus (Reading Disorders)

- Recognition accuracy from a speech recognition system
  - Word Accuracy (WA)
  - Word Correctness (WC)
- Reading duration
- Age (date of birth / date of recording)
- Age-dependent error limits
- Prosodic features
### Ongoing Projects

**Salzburger-Lese-Rechtschreib-Test (SLRT)**

<table>
<thead>
<tr>
<th>sub-test</th>
<th>content</th>
<th># of words</th>
<th>grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLRT1</td>
<td>A short list of bisyllabic, single, real words to introduce the test. This part is not analyzed according to the protocol of the test.</td>
<td>8</td>
<td>1–4</td>
</tr>
<tr>
<td>SLRT2</td>
<td>A list of mono- and bisyllabic real words</td>
<td>30</td>
<td>1–4</td>
</tr>
<tr>
<td>SLRT3</td>
<td>A list of compound words with two to three compounds each</td>
<td>11</td>
<td>3–4</td>
</tr>
<tr>
<td><strong>SLRT4</strong></td>
<td>A short story with only mono- and bisyllabic words</td>
<td>30</td>
<td>1–2</td>
</tr>
<tr>
<td>SLRT5</td>
<td>A longer story with mainly mono- and bisyllabic words but also a few compound words</td>
<td>57</td>
<td>3–4</td>
</tr>
<tr>
<td>SLRT6</td>
<td>A short list of pseudo words with two to three syllables to introduce the pseudo words. This part is not analyzed according to the protocol of the test.</td>
<td>6</td>
<td>3–4</td>
</tr>
<tr>
<td>SLRT7</td>
<td>A list of pseudo words with two to three syllables</td>
<td>24</td>
<td>1–4</td>
</tr>
<tr>
<td>SLRT8</td>
<td>A list of mono- and bisyllabic pseudo words which resemble real words</td>
<td>30</td>
<td>2–4</td>
</tr>
</tbody>
</table>
- LDA Classifier in combination with Ada-Boost
- Two-class problem: reading pathology
- Leave-one-speaker-out evaluation
- Evaluation measures
  - Area Under the Receiver-Operated-Characteristic (ROC) Curve (AUC)
  - Recognition Rate (RR)

- Speech Technology: word recognition, prosodic analysis
### Ongoing Projects

**Classification System (Reading Disorders)**

<table>
<thead>
<tr>
<th></th>
<th>SLRT 4 &amp; 5</th>
<th>SLRT 7</th>
<th>SLRT 8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Classification task</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Reading pathology&quot;</td>
<td>0.96 94.7</td>
<td>0.99 94.7</td>
<td>0.83 92.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>AUC</th>
<th>RR [%]</th>
<th>AUC</th>
<th>RR [%]</th>
<th>AUC</th>
<th>RR [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Reading pathology&quot;</td>
<td>0.96</td>
<td>94.7</td>
<td>0.99</td>
<td>94.7</td>
<td>0.83</td>
<td>92.1</td>
</tr>
</tbody>
</table>
Ongoing Projects

Classification Results (Reading Disorders)

38 children + 82 age matched control children

AUC = .96
Can speech recognition algorithms be used to detect problems in speech development?

Different speech/language skills tested with established tests
- Auditory memory: ability to remember words/sentences
- Ability to discriminate speech sounds
- Ability to construct sentences/short story

Speech of about 150 children collected so far; will be evaluated by speech therapists
Auditory Memory:
- Units of speech have to be represented in memory
- Problems in repeating units of speech indicate deficiencies

Two different aspects of auditory memory are tested:
- Ability to remember and repeat complex sentences
- Ability to repeat nonsense words

Ability to remember sentences:
- Grammatical knowledge is essential to repeat sentences
- Nonsense sentences test for short-term auditory memory
- Example: Der Kindergarten wird von den roten Bären geschüttelt. (The nursery is shaken by the red bears)

Speech Technology: word recognition, prosodic analysis
Ongoing Projects

SAD: Auditory Memory

- Ability to repeat nonsense words:
  - Ability to use known phones to create nonsense words
  - How accurate are phones represented in the memory?
- Examples
  - Ronterklabe
  - Seregropist
  - Glösterkeit

- Speech Technology: phoneme & word recognition
Ongoing Projects

SAD: Discrimination Between Phones

- Children have to repeat pairs of words/syllables and discriminate “same” from “different”
- Discrimination and articulation
- Fortis/lenis consonants:
  - pronounced with or without tenseness
- Place of articulation:
  - where in the oral cavity is the phone built

- Children with dyslexia often have difficulties to discriminate similar phones
Ongoing Projects

SAD: Discrimination Between Phones

- Test has 3 different categories
  - Words with fortis/lenis consonants
    - Kuss  Guss
    - Seide  Seite
  - Syllables with fortis/lenis consonants
    - kra  gra
    - tra  tra
  - Words with different places of articulation
    - Draht  Grad
    - Kirche  Kirsche

- Speech Technology: phoneme & word recognition
Ongoing Projects

SAD: Grammatical Usage (Short Story)

- Child has to tell a continuous story
- Is the grammatical usage appropriate for the age?

- Speech Technology: Open word recognition, prosodic analysis, syntactic analysis
Ongoing Projects

Computer Assisted Sigmatism Therapy

TRAIN YOUR S !!!

Word Test

Record

Zebra

Back

Next

Main Menu
Ongoing Projects

Computer Assisted Sigmatism Therapy

• Child appropriate response

• Speech Technology: phoneme & word recognition
Detection of sigmatism by acoustic modeling
- Test contains 16 words with /s/ and /z/ sounds
- Higher frequencies are more important
- MFCCs with filter bank from 1 kHz to 16 kHz
- 2 different datasets:
  - Sigmatism simulated by 37 speech therapists students
  - 6 children with sigmatism in therapy
Ongoing Projects

Computer Assisted Sigmatism – First Results

points correspond to supervectors

- blue: speakers without sigmatism
- red: speakers with sigmatism
Ongoing Projects

Computer Assisted Sigmatism – First Results

- Two different experiments:
  - LOO cross-validation on simulated data
  - Training on simulating speakers and evaluation on 6 children in therapy
- Recognition results on phoneme level:
  (recognition on segmented /s/ sounds)
  - Simulated data: 85 %
  - Real-life data: 84 %
- Recognition results on word level
  - Simulated data: 80 %
  - Real data: 70 %
Ongoing Projects

Articulation Diagnosis of Cleft Lip and Palate

- PLAKSS-Test:
  - 99 words on 33 slides
  - Contains all German phonemes
  - All phonemes appear in different context (beginning, center, end)
  - Standard speech test which is commonly used by speech therapists

- 31 children with CLP:
  - 10.1 ± 3.8 years old

- Speech Technology: phoneme & word recognition
Ongoing Projects

Speech Intelligibility (CLP)
Ongoing Projects

Speech Intelligibility (CLP)

[Graph showing data points and regression line]

[Images of anatomical structures]
Ongoing Projects

Speech Intelligibility (CLP)

[Graph showing correlation between standard accuracy and receptive evaluation, with a regression line indicated.]
Ongoing Projects

Diagnosis Based on a Stutter Model

- Automatically generated pronunciation dictionary and grammar
- Optional breaks/insertions after every phoneme
  - To predecessor phoneme and beginning of word
  - To predecessor and beginning of syllable
- Hot spots: beginning of clauses (high probability for stuttering)
- Repetitions of more than one word
- “Well, said the old rat angrily, we can't wait all day for you [...]”

- Problem: Some stutterers don’t stutter while reading
Ongoing Projects

Test Data for Adults (Who Can Read)

- “The North wind and the Sun”:
  - 107 words (71 disjoint)
  - Contains all German phonemes
  - Commonly used by speech therapists and phoneticians
- 52 stutterers
- Verification of the read text with the modified word and syntax model

- Speech Technology: phoneme & word recognition, precompiled syntactic analysis
Ongoing Projects

Evaluation Criterion for Stuttering

\[ \text{lexical model without optional stutter events} \]

\[ \text{forced alignment with stutter model} \]

“Phoneme accuracy” (not counting substitution \([i] \rightarrow [y]):\)

\[
(7-3) \times 100 / 7 = 57% 
\]
Ongoing Projects

Automatic Classification of Stutter Degree

Stutter events per word (perceptual evaluation)
Ongoing Projects

Degree of Stuttering vs. “Phoneme Accuracy”

Graph showing the relationship between stutter events per word (perceptual evaluation) and phoneme accuracy.
Ongoing Projects

Intelligibility (Laryngectomees)

- “The North wind and the Sun”:
  - 107 words (71 disjoint)
  - Contains all German phonemes
  - Commonly used by speech therapists and phoneticians

- 41 laryngectomees:
  - 62.0 ± 7.7 years old

- Word recognition with a unigram language model

- Speech Technology: phoneme & word recognition, prosody
Ongoing Projects

Intelligibility (Laryngectomees)
Ongoing Projects

Measuring Intelligibility with PLTT

- Two Recognizers:
  - Semicontinuous HMMs
  - Continuous HMMs
- Different vocabulary sizes:
  - All the words from all PLTT words and sentences (738)
  - Addition of acoustically similar words (1017)
- Word models based on monophones or triphones
- Unigram language model
- Comparison with average of 11 naïve listeners

-.81 \leq r, \rho \leq .93, \text{ improvement by combination of recognizers}
“The North Wind and the Sun”:
- Word accuracy corresponds to intelligibility
- Intelligibility is influenced by different factors
  - CLP: problems with individual phonemes (nasalization)
  - Laryngectomized: disturbing noises, hoarseness, match of breath and sense units
- Evaluate distinct aspects with acoustic and prosodic parameters
### Ongoing Projects

**Evaluation of Distinct Aspects**

<table>
<thead>
<tr>
<th>criterion</th>
<th>prosodic feature type</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hoarseness</td>
<td>energy</td>
<td>-.74</td>
</tr>
<tr>
<td>disturbing noises (stoma)</td>
<td>energy</td>
<td>-.76</td>
</tr>
<tr>
<td>agreement of breath and sense units</td>
<td>pause word duration</td>
<td>.84</td>
</tr>
<tr>
<td>effort</td>
<td>pause word duration</td>
<td>.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.76</td>
</tr>
</tbody>
</table>
Ongoing Projects

Intelligibility (Oral Cancer)

- “The North wind and the Sun”:
  - 107 words (71 disjoint)
  - Contains all German phonemes
  - Commonly used by speech therapists and phoneticians

- 62 patients:
  - 59 ± 12 years old

- Word recognition with a unigram language model

- Speech Technology: word recognition, prosodic analysis
Intelligibility (Oral Cancer)

![Graph showing correlation between word correctness and experts' scores with correlation coefficient r = -0.93]
Ongoing Projects

Intelligibility vs. Size of Tumor (TNM)

Mann-Whitney-U-Test: $\alpha = 0.05$
Ongoing Projects

Intelligibility vs. Anatomical Structure

Mann-Whitney-U-Test: $\alpha = 0.05$
Ongoing Projects

Intelligibility vs. Surgical Method

Word correctness

Surgical reconstruction

Mann-Whitney-U-Test: $\alpha = 0.05$
Ongoing Projects

Medical Use

- Speech competence is an important outcome parameter for cancer treatment
- Prognosis (how severe will the dysfunction be after therapy?)
- Planning of therapy (which method is best?)
### Ongoing Projects

#### Speech Intelligibility (Summary)

<table>
<thead>
<tr>
<th>Pathology</th>
<th>Correlation Perceptual vs. ASR evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLP children</td>
<td>0.89</td>
</tr>
<tr>
<td>Alaryngeal voice</td>
<td>0.88</td>
</tr>
<tr>
<td>Oral Cancer</td>
<td>0.93</td>
</tr>
<tr>
<td>Dysarthric Speakers</td>
<td>0.90</td>
</tr>
</tbody>
</table>
Ongoing Projects

Dementia

- **SKT**: a test to assess memory and attention deficits
  - 9 Subtests: memory (3), attention (6)
  - Five parallel forms to avoid learning effects
- **Main areas of application**
  - Moderate degrees of dementia
  - Mild Cognitive Impairment (MCI)

- Speech Technology: Open word recognition, prosodic analysis, syntactic analysis, semantic analysis
Ongoing Projects

Vision: Self-Test to Detect Changes

Integration into assisted living
- Continuous screening through periodic tests
- Use existing equipment (TV,…)
- Human-Machine Interface:
  - Speech
  - Gesture-based navigation
  - Wii-Mote
- Analysis of test results by the supervising general practitioner
Ongoing Projects

Use of Evaluation in Given Examples

- Diagnosis
  - Stutterers, CLP, laryngectomees, dysarthria, dementia

- Therapy control
  - Sigmatism, stutterers, CLP, laryngectomees, dysarthria, dementia

- Comparison of therapy methods
  - Oral squamous cell carcinoma
  - Planned multi-center study on operation methods for CLP

- Screening
  - Learning disorder

- Computer-assisted therapy
  - Sigmatism, dementia
Outline

- Motivation
- Examples for Disorders in Speech Production
- Speech technology and its use in established tests
- Ongoing and planned projects
  - Linguistic level that needs to be modeled
  - Results (where already available)
- **Peaks: a client server architecture**
- Outlook: emotional disorders & multimodality
- Summary
The Peaks System

Peaks Architecture

Client

- recording (16 KHz)
- transmission (SSL)
- transmission (SSL)
- report

Server

- audio-data
- features
- speech processing
- prosodic features
- recognized word chain
- report
- evaluation

Features:
- audio-data transmission (SSL)
- speech processing
- prosodic features
- recognized word chain
- report evaluation

Transmission Features:
- Chain transmission
- SSL encryption

Data Processing:
- Audio-data transmission
- Prosodic features
- Recognized word chain
- Speech processing

Evaluation:
- Report evaluation
Internet based recording and analysis
Offline version for mobile applications
Motivation
Examples for Disorders in Speech Production
Speech technology and its use in established tests
Ongoing and planned projects
  - Linguistic level that needs to be modeled
  - Results (where already available)
Peaks: a client server architecture
Outlook: emotional disorders & multimodality
Summary
Outlook

Need for Multimodality: Emotional Disorders
Outlook

Need for Multimodality: Emotional Disorders
Outlook

Need for Multimodality: Emotional Disorders
Outlook

Need for Multimodality: Emotional Disorders
Outlook

Need for Multimodality: Emotional Disorders
Outlook

Need for Multimodality: Emotional Disorders
Outlook

Need for Multimodality: Emotional Disorders
Clinical diagnosis of reading disorders is only based on

- IQ
- Time limit
- Reading error limit

- Ignores information about, e.g., eye movements
Outlook

Need for Multimodality: Facial Paresis

- Dysarthric speech often accompanied by other physical impairments
  - Facial paresis
  - Motor handicaps
- Reduced mobility requires therapist to come to patient
  - High costs
  - Waste of therapist’s time
→ Telemedical therapy
Outlook

Telemedical System
Need for Multimodality: Facial Paresis

- Unstressed look
- Lip pursing
- Closing of eyes
- Showing the teeth
- Up to 50 Hz
- More than 25k 3D points (176*144 pixels)
- Eye-safe infrared light / no exposure
Outlook

Need for Multimodality: Facial Paresis
Outlook

Need for Multimodality: Facial Paresis
Outlook

Need for Multimodality: Facial Paresis
Outlook

Need for Multimodality: Facial Paresis

- Illumination
- Stereo microphones
- TOF camera
- Webcam
- Control image for the patient
Outlook

Need for Multimodality: Facial Paresis
Motivation: high demand for objective evaluation

Examples for disorders in speech production

Speech technology and its use in established tests

Ongoing and planned projects
  - Linguistic level that needs to be modeled
  - Use in treatment
  - High correlation with human experts (above 0.9)

Peaks: a client server architecture

Outlook: emotional disorders & multimodality
Thank you for your attention

Supported by
Deutsche Forschungsgemeinschaft (DFG)
Deutsche Krebshilfe (German Cancer Aid)
Wilhelm-Sander-Stiftung

noeth@informatik.uni-erlangen.de