LNTS - LiNguiStTech Solutions

True Phonemic-Based Speech Recognition

Gad Oren – CEO/CTO, LNTS
True Phonemic-Based Speech Recognition

- **Speech Recognition Issues**
- **LNTS Segment Approach**
- **LNTS Algorithm Set**
- **LNTS Phoneme Recognizer Approach**
- **Spectravocs Technology**
Speech Recognition Issues

• Run-on words – spontaneous, continuous speech

• Large variance between speakers
  ◦ Variance in speaker vs. self
  ◦ Variance between different speakers

• Current SR engines require:
  ◦ New phonetic dictionary per language
  ◦ New language model per language variation:
    • US English
    • Indian English
    • UK English
    • Australian English
    • others…
Dictionary Size vs. CL Trade-Off

• Key SR parameter is **Confidence Level (CL)**

• **CL decreases** as dictionary size **increases**
SR Engines: General Characteristics

Current Engines

Discrete Speech

Full Language Dictionary Approach
[using either HMM or NN, or combination]

Spectravocs Engine

Continuous Speech

Segment Database Approach

Phoneme Database Approach
True Phonemic-Based Speech Recognition

- Speech Recognition Issues
- LNTS Segment Approach
- LNTS Algorithm Set
- LNTS Phoneme Recognizer Approach
- Spectravocs Technology
LNTS Segment Approach

(Segment ≈ syllable. e.g. – segment has 2 segments – seg and ment)

- Compares segments to segment DB, not to phonetic dictionary
  - Dictionary comparison – \( x \) million words
  - Segment comparison – \( y \) thousand segments
    - 3 orders of magnitude less

- LNTS has defined thousands of segments

- Segments are not language specific; same segment can be reused in other languages
Segment Spectra:

- Temporal & spectral analysis, presented in 3D
- Color = amplitude
- Lines = sum of amplitudes of each time period

Feb 2008
Time/Frequency/Amplitude View
LNTS SR Process

Input sentence WAV file

Split to Words → Transform to 3D fingerprint → Normalization → Split to Segments

3D print

Parameter Extraction for DB Sorting → DB Sorting

Full Segment Database → Dedicated Matching Function

Recognized Segments → Segment Correction

Recognized & Corrected Segments

Off Line DB Creation
LNTS Segment Method

1. Segment Database built offline
2. DB includes thousands of representations of all segments
3. Voice data (sentences) split to words, then to segments
4. Each segment recognized using matching algorithm with the segment DB

No phonetic dictionary used.
True Phonemic-Based Speech Recognition

- Speech Recognition Issues
- LNTS Segment Approach
- **LNTS Algorithm Set**
- LNTS Phoneme Recognizer Approach
- Spectravocs Technology
LNTS Algorithm Set

Dedicated set of algorithms to enable SR:

A. Offline Segment DB Creation
B. Split Sentence to Words
C. Split Words to Segments & Parameter Extraction
D. Normalize to DB parameters
E. Matching with Segment DB
F. Segment Correction
A. Offline Segment DB Creation

1. Records created of segments for all speaker types

2. DB of 3D representations of segments built in multiple sub-DBs

3. Sub DBs optimized for faster reference
B. Split Sentence to Words

Intra-word Intervals  Run-on Words

Feb 2008
Noise Filtration

1. Background noise, e.g. - inside a car

2. Mobile networks, e.g. - compression, bandwidth limitation etc.

3. Inferior microphone quality

- Background noise pre/post filtering

- Compounded noise, affects original signal
Split Sentence To Words

1. Input: WAV vector sampled at 8 kHz, encoded in 16 bits

2. Output: start & end time tags of each word

3. Uses powerful heuristics on energy curves

Energy Curves

Feb 2008
C. Split Words to Segments & Parameter Extraction

- Takes “Split to Words” results to find beginning / end of each segment
- Extracts parameters to define Sub DB to optimize matching process (E below)
D. Normalize to Reference Frequency (F0)

• Special alignment of frequency
E. Matching with Segments DB

• Finds best match between the spoken segment with the segment DB

• Uses extracted parameters (from C above):
  ▪ to reduce DB size
  ▪ to shorten matching process
F. Segment Correction Algorithm

- Minor fine-tuning to increase recognition accuracy
- Increases Confidence Level
True Phonemic-Based Speech Recognition

• Speech Recognition Issues
• LNTS Segment Approach
• LNTS Algorithm Set
• **LNTS Phoneme Recognizer Approach**
• Spectravocs Technology
LNTS Phoneme Recognizer Approach

• Goal – reduce the comparison DB even further e.g. – only 41 phonemes for US English
  (gain of another 2 orders of magnitude: smaller = faster)

• Unique Vocal 3D "fingerprint" for each phoneme, same concept as for Segment Approach

• Slightly different algorithms applied for phonemes
LNTS Phoneme Spectra

T phoneme

A phoneme

Feb 2008
## Segments vs. Phonemes Tradeoff

<table>
<thead>
<tr>
<th>Issue</th>
<th>Segments</th>
<th>Phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DB Size</strong></td>
<td>Thousands of segments</td>
<td>Tens of phonemes – faster matching</td>
</tr>
<tr>
<td><strong>Sound Energy</strong></td>
<td>More – higher CL for recognition</td>
<td>Less</td>
</tr>
<tr>
<td><strong>Complexity</strong></td>
<td>More combinations</td>
<td>Fewer combinations – fewer complexity-caused errors, easier integration</td>
</tr>
</tbody>
</table>

*Ideal solution is a hybrid system, with advantages of both.*
True Phonemic-Based Speech Recognition

- Speech Recognition Issues
- LNTS Segment Approach
- LNTS Algorithm Set
- LNTS Phoneme Recognizer Approach
  - Spectravocs Technology
Spectravocs Technology Advantages

- No phonetic dictionary, faster and more accurate recognition
- Accurately recognizes continuous speech
- Extracts more data from even less sound energy e.g. - alphanumerics
- Extensible in near future to pure phonemes, and phoneme/segment combination, per SAMPA
- Same segments for multiple languages e.g. mar – in French mardi and English market
LNTS Technology Applications

- **Voice Search**  *mobile & desktop search - internet, GPS, 411, ...*
- **Speech to SMS**  *freestyle speech converted to SMS or IM*
- **Call Centers**  *accurate free language input*
- **Speech-activated transactions**  *telebanking, financial services, travel, entertainment ...*
- **Dictation to text**  *medical/legal, meeting minutes, etc.*
- **Speech Authentication**  *caller and gender identification, verification for security purposes*
Thank You!

LNTS

gad.o@lntstech.co.il  www.lntstech.com

(571) 426-9804