Telephone Based Automatic Voice Pathology Assessment.

Rosalyn Moran¹, R. B. Reilly¹, P.D. Lacy²

¹ Department of Electronic and Electrical Engineering, University College Dublin, Ireland
² Royal Victoria Eye and Ear Hospital (RVEEH), Dublin, Ireland.


........ To Follow

• Why a system for voice disease?
• The Anatomy of Voice
• Generic Voice Classifiers
• The next obvious step
• System Design
• Results
• Refinement
Why is this area Important?

- Voice disorders are relatively common in the general population
  - 5% suffering abnormalities requiring medical intervention.
  - Cancerous tumors of vocal folds account for 40% of all head and neck carcinomas
  - At risk professionals: logistically difficult to monitor

- Currently, the accurate diagnosis requires visualisation of the larynx.
  - Videostroboscopy is the current gold standard
e  - costly, time consuming, often subjective and labour intensive.

Anatomy of the Vocal Folds: *Healthy Waves*

- Vagus Nerve activates fold closure
- Air Pressure from lungs force folds apart
- For short “Voiced” phonation (long vowels) folds move periodically; The Mucosal Wave
Vocal Fold Pathologies, *Unhealthy Waves*

- Structural (growths),
- Neurological (Loss of effective nerve action),
- *Lack of constancy*
- *Escaping Air due to incomplete fold closure*

Generic voice pathology classifier

- **Acquisition**
  - Sustained Phonation of vowel sound /a/
  - Language Independent!

- **Feature Extraction**
  - Measures of vocalisation constancy
  - Pitch
  - Amplitude
  - Noise

- **Classifier**
  - Various,
  - HMMs
  - ANNs
  - LDA

- **Normal / Abnormal**
Where have we come to date?

- **Successful Automatic classifiers:**
  - Acquisition → Feature Extraction → Classifier → Normal / Abnormal
  - **Classification rate in excess of 90% for separating normal from pathology voice**
    - C. Maguire, P. de Chazal, R. Reilly, P.D. Lacy, World Congress on Medical Physics and Biomedical Engineering.

  **Motivation:** Can we make this more useful?
  ............... Could you use a telephone ...

Under Investigation.....

- **New method of acquisition employing IVRs to allow transfer of data across telephone networks and the internet.**
  - **Remote**
  - **Secure**
  - **Identifiable**

  **System Infrastructure:** VoiceXML
An intelligent dialogue system

Incorporation of digital signal processing algorithms

Transmission Characteristics

DSP

Comms

Voice XML Acquisition

- VoiceXML Scripts held on a web server.
- Transferred to VoiceXML Gateway Voxpilot for TTS and speaker recognition.
- Dial up applications using any telephone.
Database

- Disordered Voice Database Model 4337
- Massachusetts Eye and Ear Infirmary
- 631 valid patient samples of sustained phonation of /a/
  - Wide variety of pathologies condensed to normal / abnormal in this study
  - Prelabelled by panel of experts
  - Recorded in soundproof environment using a high quality microphone

....Corrupting The Corpus

- To Identify Causes of Information loss
  - Imitate telephone conditions by progressively degrading the quality of the database.
  - Examine feature accuracies at each stage
Creating 5 Test Corpii

1. Begin: 631 High Quality Speech Files @ 10kHz
2. Degrade: Resample to 8kHz
3. Degrade: Bandpass filtered from 100Hz-3.2kHz
4. Degrade: Add Noise
5. Transmit: Original Database

Transmission Channels:
Analog and Digital Long Distance Links

1. VoiceXML Calling Application: Plays 30 Speech files,
2. VoiceXML Application: “answer” and save transmitted speech files.

*CORPUS 5
Features Features Features!

...... Of Medical Relevance
    in conjunction with our Medical Consultants

- Pitch Perturbation Features, Jitter (12)
- Amplitude Perturbation Features, Shimmer (12)
- Energy Measures, Harmonic to Noise Ratio HNR (11)

Classifier / Performance Estimation

- Classifier
  - Linear Discriminant Analysis

- Performance Estimation
  - Normal recordings duplicated to balance classes
  - 10 runs of 10 fold cross-validation
    • independent training and testing sets
  - Specificity, sensitivity, predictivities, accuracy
Classification Performance

Composite Feature Breakdown
Composite Feature Breakdown

- Shimmer Group proving most robust.
- HNR accuracies fall significantly.

Moving On

- Classification rate of 74% for separating normal from pathology voice......over the telephone.

- Further Refinement
  - Homogenous Data Sets * Physical * Neuromuscular * Mixed
Physical Pathology

- Occurs when the function of the larynx has been affected by a physical change in the anatomy of the larynx. For example an arytenoid granuloma.

Neuromuscular Pathology

- Occurs when the nerves that control the movement of the muscles in the larynx have been altered in some way. An instance of this is Vocal Fold paralysis.
Telephone Based Results: Improved Accuracy

<table>
<thead>
<tr>
<th></th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuromuscular</td>
<td>87%</td>
</tr>
<tr>
<td>Physical</td>
<td>78%</td>
</tr>
<tr>
<td>Mixed</td>
<td>61%</td>
</tr>
</tbody>
</table>

Opportunities for providing related health care information by voice applications:

- Voice assessment
- Speech training
- Improving literacy

Wider Impact for Healthcare
Thank you

How is your voice?

rosalyn.moran@ee.ucd.ie