Telephone Based Automatic Voice Pathology Assessment.

Rosalyn Moran¹, R. B. Reilly¹, P. de Chazal¹, P.D. Lacy²

¹ Department of Electronic and Electrical Engineering, University College Dublin, Ireland

² Royal Victoria Eye and Ear Hospital (RVEEH), Dublin, Ireland.
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

- Currently, the accurate diagnosis requires visualisation of the larynx.
  - Videostroboscopy is the current gold standard
    - costly, time consuming and labour intensive.
Anatomy of the Vocal Folds

- 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

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

Acquisition → Feature Extraction → Classifier → Normal / Abnormal

- Sustained Phonation of vowel sound /a/
- High Quality Speech: 25kHz Sound Proof Chamber
- Measures of vocalisation constancy
  - Pitch
  - Amplitude
  - Noise
- Various, HMMs, ANNs, LDA
Where have we come to date?

- Successful Automatic classifiers:
  - Classification rate in excess of 90% for separating normal from pathology voice

- Motivation: Can we make this more useful?

  .................. Could you use a telephone ...

(D.S.P. Research Group, University College Dublin)
New method of acquisition employing IVRs to allow transfer of data across telephone networks and the internet.

- Remote
- Secure
- Identifiable

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

D.S.P. Research Group, University College Dublin
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.

*CORPUSS 5*
Feature Groups

…… 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

- Clean 10kHz: 89.104%
- Clean 8kHz: 83.35%
- Filtered: 82.49%
- Noisy: 79.14%
- Telephone: 74.15%

D.S.P. Research Group, University College Dublin
Composite Feature Breakdown

Clean 10kHz

Jitter: 68.93%
Shimmer: 77.1%
HNR: 77.8%

Bandlimited (Sampling Rate 8kHz)

Jitter: 66.09%
Shimmer: 77.97%
HNR: 79.79%

Filtered 100Hz-3200Hz

Jitter: 66%
Shimmer: 75.63%
HNR: 63.66%

Noise Corrupted (30dB SNR)

Jitter: 75.7%
Shimmer: 74.85%
HNR: 57.16%
Composite Feature Breakdown

Shimmer Group proving most robust.
HNR accuracies fall significantly.

Telephone Corpus

- Jitter: 64.7%
- Shimmer: 73.03%
- HNR: 57.85%
Conclusions

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

Further Developments
- Extend pilot study with local Hospital Speech and Language Departments.
- Extend study to more pathology classes
- Extend system of automatic grading
Thank you

How is your voice?
Dial 00353 1 4364288 and leave a sample.....

rosalyn.moran@ee.ucd.ie