Investigating the role of the Lombard reflex in Non-Audible Murmur (NAM) recognition

Panikos Heracleous

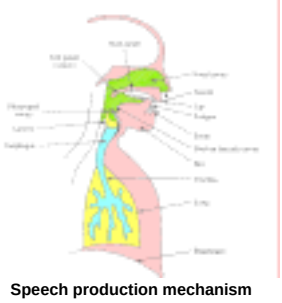
Speech and Acoustics Processing Laboratory, NAIST

Preface

- Non-Audible Murmur (NAM)
  - Definition
  - Characteristics
- NAM Recognition
  - Experiments in clean environments
  - Experiments in noisy environments
- Role of Lombard reflex in NAM recognition
  - Illustration
  - Experiments
- Conclusions & Future work

Non-Audible Murmur (NAM) Definition

- ***Definition***
  - A terminology, which describes unvoiced speech received by a NAM microphone through body tissue.
  - It produced with the vocal cords not vibrating.
  - It originates from a turbulent noise generated in and above the larynx.
  - It can be considered as soft whisper, or small voice. Sound level about 32-35 dB SPL.
  - Optimal capturing position
    - Behind talker's ear
  - **Applications in speech recognition for privacy**



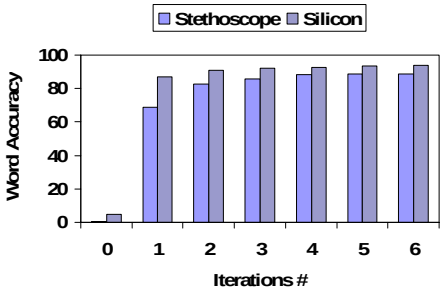
Speech production mechanism

Vocal cords
- Tensed: Air causes then to vibrate. Voiced sounds
- Relaxed: Turbulent air. Unvoiced sounds

Non-Audible Murmur Characteristics (1/2)

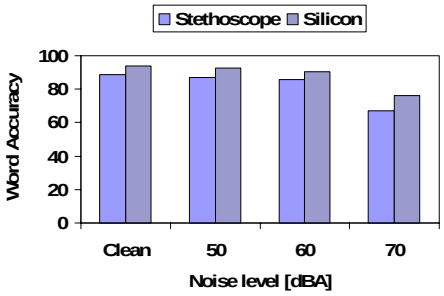
- Characteristics
  - It does not corporate fundamental frequency (F0).
  - Vowels and consonants can be discriminated.
  - Body tissue acts as low-pass filter. High frequency components are attenuated. But, are still present with lower amplitude, and sufficient information to discriminate sounds.
  - Speech intelligibility is affected. Especially, when differences are only in the upper frequency bands.
  - Speech recognition is impossible with normal-speech acoustic models. Models should be adapted to NAM characteristics.
  - Speech recognition using NAM data is possible with high accuracy
    - Speaker-dependent experiments: Over 90%
    - Speaker-independent with small amount of training data: ~75%

Non-Audible murmur recognition using clean data



-Maximum Likelihood Linear Regression (MLLR) adaptation method
-350 training utterances, 48 test utterances

Non-Audible murmur recognition using simulated noisy data

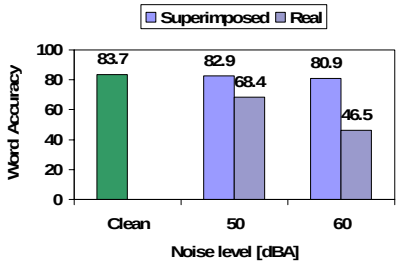


High robustness against noise

-Maximum Likelihood Linear Regression (MLLR) adaptation method
-350 clean training utterances, 48 test utterances
-Noise played back and captured by a NAM microphone attached on a speaker

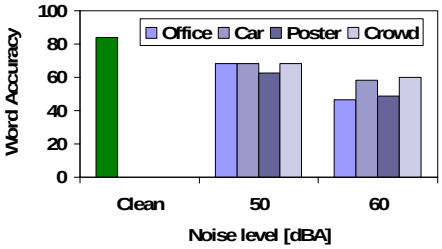
Non-Audible murmur recognition using real noisy data (1/2)

Office noise played back while speaker utters the test set



-100 training utterances recorded in clean environment by a female speaker
-24 test utterances
-Adaptation (MLLR)

Non-Audible murmur recognition using real noisy data (2/2)

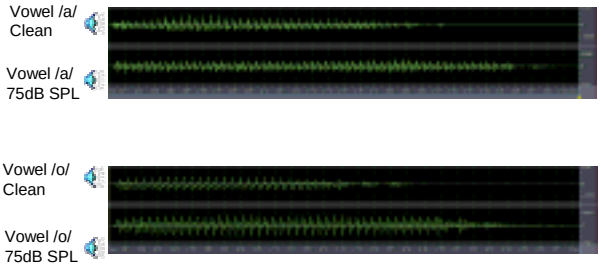


-Clean hidden Markov models – 100 training utterances
-24 test utterances in noisy environment
-Adaptation (MLLR)

Lombard Reflex (or effect)

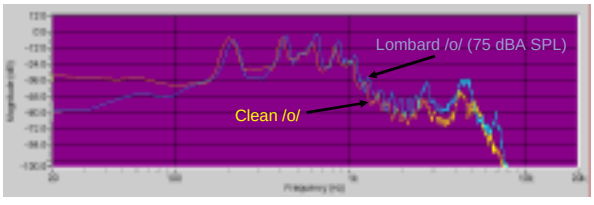
- **When speech is produced in the presence of noise, speech characteristics are changed due to Lombard reflex [Junqua, 1996].**
  - Intensity (or power, or loudness) is increased
  - F0 (or pitch) contour is changed
  - Formants are shifted
  - Phrase durations are increased
  - Spectral tilt is changed
- **Lombard reflex affects speech recognition**
  - Performance decreases not only by the presence of noise, but also due to speech characteristics changes.
  - 60dBA level Lombard speech recognition shows significant decreases
- **The effect of Lombard reflex is strongly individual**
  - How talker cope with the noise
  - Different noise causes different changes in speech characteristics
- **Lombard speech:** Clean speech uttered while listening to noise through headphones or earphones

Lombard reflex illustration for normal speech Duration changes



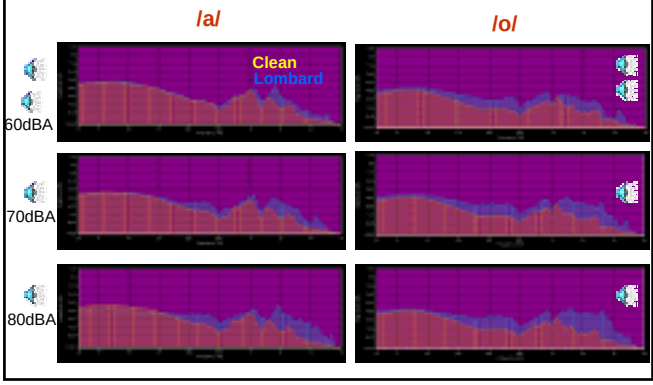
Lombard speech: Listener while talking listens to noise using ear-phone
The changes in duration can be seen and be heard.

Lombard reflex illustration for normal speech Spectral changes



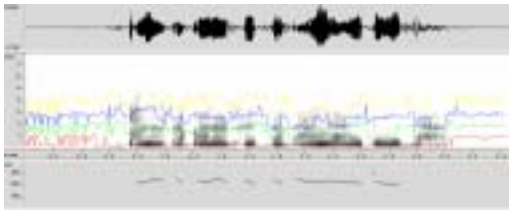
It can be seen in Lombard speech:
-Intensity is increased. Talker's vocal tract efforts are increased in order to increase his speech intelligibility. Intelligibility in noise is higher.
-Formants are shifted
-Spectral tilt is changed
-The changes might be different due to noise type and level
-Mismatch in speech recognition. Performance is decreased

Lombard Reflex illustration for Non-Audible Murmur



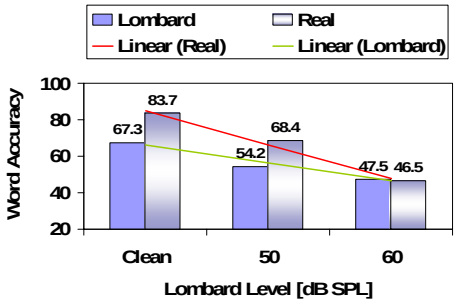
Lombard reflex in NAM recognition

- Clean speech - The well known NAM.
- 50 dB SPL - Intensity is increased. Still NAM.
- 60 dB SPL - Intensity is further increased. Still NAM??
- 70 dB SPL - Already differs from NAM. Vocal cords vibrations.
- 80 dB SPL - This is not NAM!!! Vocal cords are vibrating. F0, formants are detectable.



Lombard NAM speech at 80 dBA SPL noise level. Normal speech characteristics

Experiment using Lombard data



- 50 training utterances, 24 test utterances from a female speaker
-However, Lombard reflex affect NAM recognition dramatically.
-**Analogy between real noisy data and Lombard data**

Conclusions – Future Work

- Non-audible murmur recognition in clean and noisy environments
- High robustness against noise using simulated noisy data
- Performance decreases using real data: **Lombard reflex**
- Lombard reflex in NAM recognition
  - A negative impact effect. Performance decreases.
  - Using Lombard speech to train HMMs
  - Effect of various noise types, levels on Lombard reflex
  - Lombard reflex-robust NAM HMMs using Lombard training speech
    - NAM, small voices, whisper speech can be used to train a set of HMMS
    - Apply in NAM recognition

Thank you!