

A tissue-conductive acoustic sensor applied in speech recognition for privacy

P. Heracleous, Y. Nakajima, H. Saruwatari, K. Shikano

Nara Institute of Science and Technology, Japan

1

Contents

- Non-Audible Murmur (NAM)
- NAM characteristics
- NAM automatic speech recognition
 - Recognition in clean environments
 - Recognition in simulated noisy environments
 - Recognition in real environments
- Effect of Lombard reflex in NAM recognition
 - Lombard characteristics
 - Experiments using Lombard data
- Conclusions

2

Non-Audible Murmur (NAM)

■ 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.



*Silicon NAM microphone
NAIST, 2003-2005*

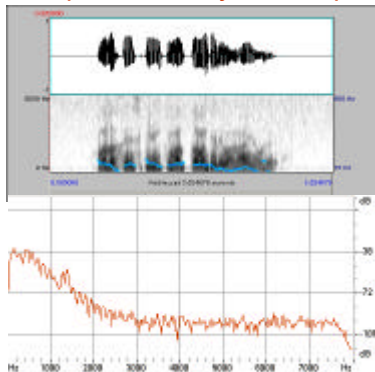


Attachment of NAM microphone

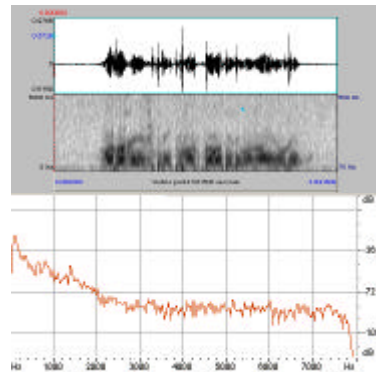
3

Non-audible murmur characteristics

Audible speech received by NAM microphone



Non-audible murmur

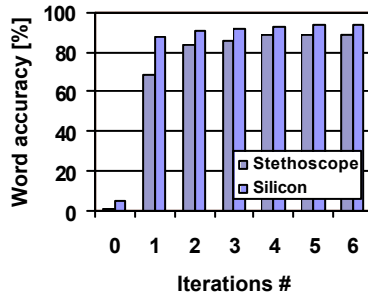


- Body transmission and loss of lip radiation act as low-pass filter
- Lower intelligibility and quality.
- After 1kHz the spectral components are attenuated. After 3kHz remain flat.
- Enough information to distinguish and recognize sounds.

4

NAM recognition in clean environment

- **Privacy in human-machine communication**
- Normal-speed HMMs cannot be used
 - Feature distortions
- HMMs trained with NAM database are necessary
- Using speaker and environment adaptation techniques
 - Maximum Likelihood Linear Regression (MLLR)
 - Maximum A Posteriori (MAP)
 - Iterative approach

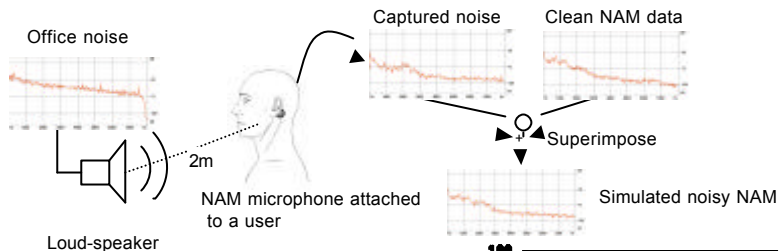


- Clean training NAM data
- Clean test NAM data
- Multi-iterative MLLR
- High performance

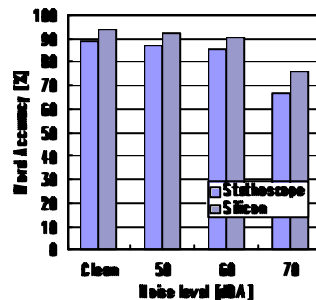
-350 NAM training utterances, 48 NAM test utterances

5

NAM recognition in simulated noisy environments

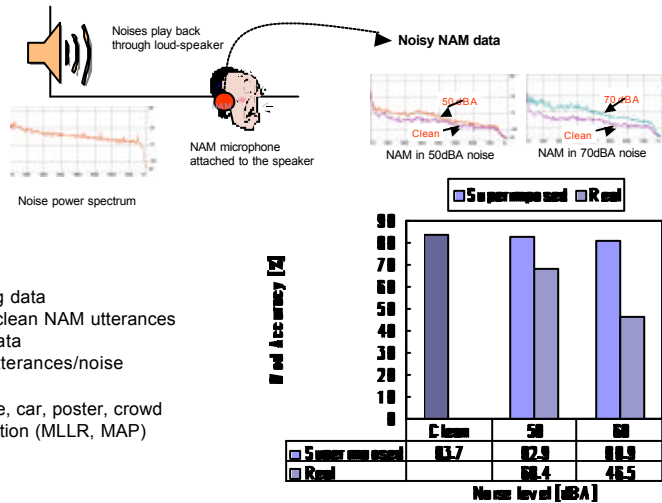


- Training data: 350 clean NAM utterances
- MLLR and MAP adaptation techniques
- 48 simulated noisy test data
- Japanese dictation task
- 20k vocabulary



6

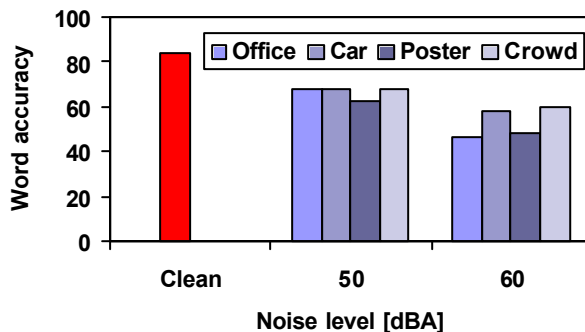
NAM recognition in real noisy environments



- Training data
 - 100 clean NAM utterances
- Test data
 - 24 utterances/noise
- Noise
 - Office, car, poster, crowd
- Adaptation (MLLR, MAP)

7

NAM recognition Real environments



- Adaptation approaches
- 100 training utterances recorded in clean environment
- 24 test utterances recorded in real environment

In real environments, with increasing noise level word accuracy decreases

8

Lombard Reflex (or effect)

- **When speech is produced in the presence of noise, speech characteristics change [Lombard, 1921]**

- Intensity (or power, or loudness) increases
- F0 (or pitch) contour changes
- Formants shift
- Phrase durations increase
- Spectral tilt changes

Lombard NAM speech



Talker is listening to noise while NAM is uttered

- **Lombard reflex affects speech recognition**

- Performance decreases not only by the presence of noise, but also due to speech characteristics changes.

- **The effect of Lombard reflex is strongly individual**

- How talker cope with the noise
- Different noise causes different changes in speech characteristics

9

Lombard reflex Duration changes

Vowel /a/

Clean

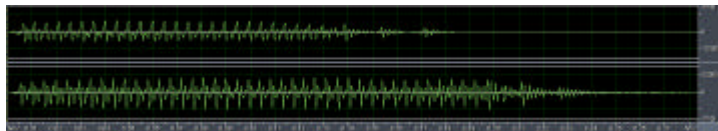


Vowel /a/

75dB SPL

Vowel /o/

Clean



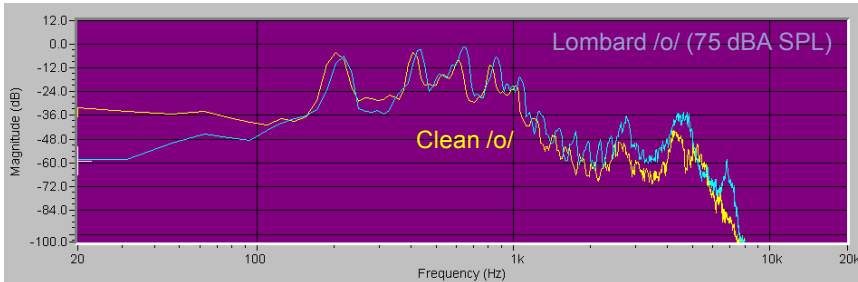
Vowel /o/

75dB SPL

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

10

Lombard reflex Spectral changes



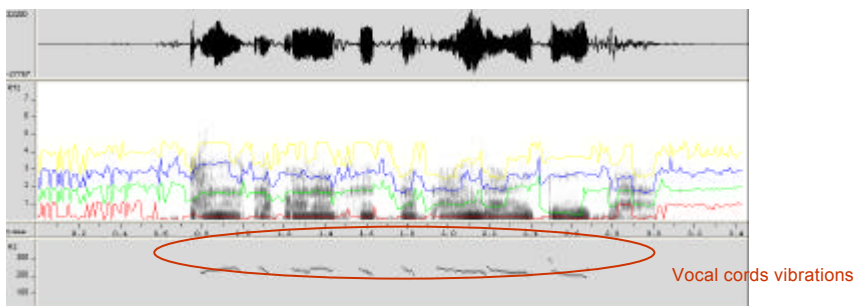
It can be seen in Lombard speech:

- Intensity increases. Talker's vocal tract efforts increase in order to increase the speech intelligibility
- Mismatch in speech recognition. Performance decreases

11

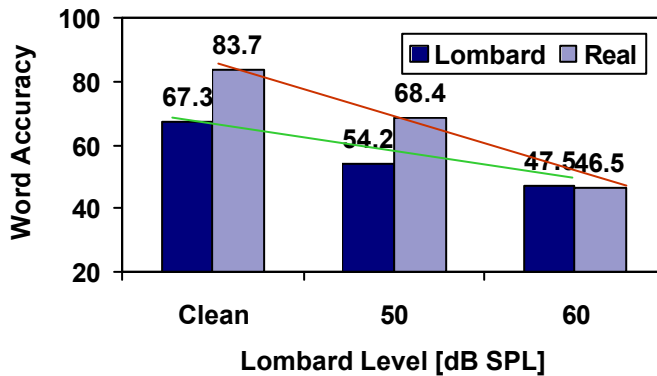
Lombard reflex in NAM recognition

- 50 dB SPL - Intensity is increased.
- 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 ¹²

Lombard baseline experiment



- 50 training utterances, 24 test utterances from a female speaker
- MLLR adaptation using 3000-states PTM HMMs
- Analogy between real noisy experiment and Lombard experiment

13

Conclusions - Future Work

- Introduction to Non-Audible Murmur
 - NAM microphones
 - NAM frequency characteristics
- Non-Audible Murmur recognition
 - Clean environment: **93.9%** word accuracy
 - Simulated noisy test data: Shows robustness
 - Real noisy test data: Performance decreases
- Investigating the effect of the Lombard reflex
 - Markedly affects non-audible murmur recognition

14