Lombard Effect in Speech Production by Cochlear Implant Users: Analysis, Assessment and Implications

Ph.D THESIS DEFENSE

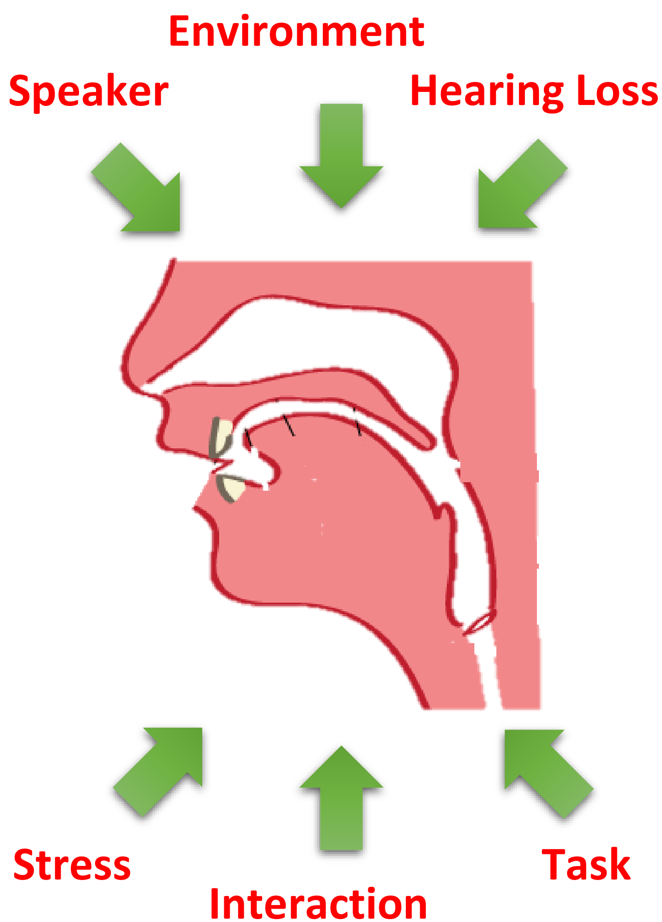
Jaewook Lee

March 20, 2017

Center for Robust Speech Systems -
Cochlear Implant Laboratory (CRSS-CIL)
Department of Electrical Engineering
The University of Texas at Dallas

External Factors which Influence on Speech Production

Factors Influence on Speech:



- Changes in speech production based on *auditory feedback* are an important research domain.
- Factors influence on speech:
  - **Speaker:** *age, gender, nativeness, emotion*
  - **Environment:** *noise, reverberation, distance*
  - **Conversation:** *read, spontaneous*
  - **Task:** *stress, workload, inter-personal interaction*

This study focused on *"Acoustic noisy environment"*

Lombard Effect

- **Lombard effect** - involuntary response a speaker experiences in the _presence of noise_ (Lombard, 1911).
- Cause increased _vocal effort_ including:
  - Vocal intensity
  - Pitch period structure
  - Formant characteristics
  - Glottal spectral slope
  - Speech rate, etc.

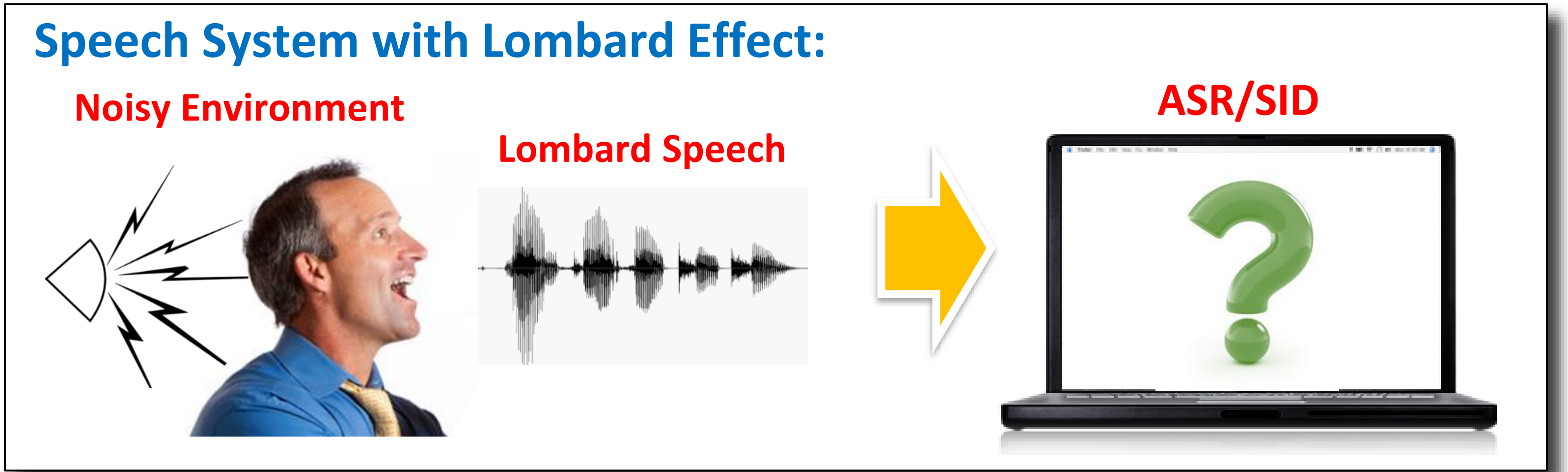
  (Lane & Tranel, 1971; Hansen, 1988; Garnier *et al*., 2010; Junqua, 1992)
- Helps to maintain _intelligible communication_ in challenging listening environments.

Intelligible Communication in Noise Environment

Lombard Effect

- Also cause to degrade *automatic speech system (ASR/SID)* performance when system is modeled with neutral speech.
- Several signal processing technique have been proposed to *compensate for LE* in speech to improve the performance of the speech systems (Hansen and Varadarajan, 2009; Hansen, 1994).

Well studied in *normal hearing (NH)* listeners and *speech systems*, but not in the field of *cochlear implants (CI).*

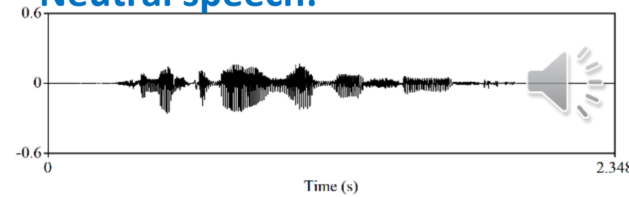


Example "Lombard" Speech

Lombard Speech Collection from NH

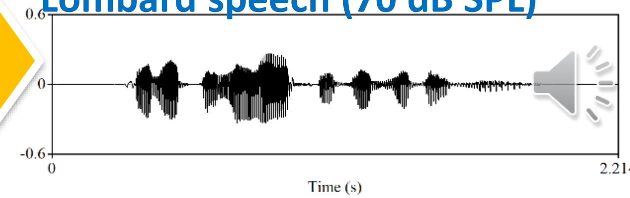
- Produced *AzBio sentences* in a way of 2-way conversation.
- *Large crowd (LCR) noise* was presented at 70dB, 80dB, and 90dB SPL.

Neutral speech:



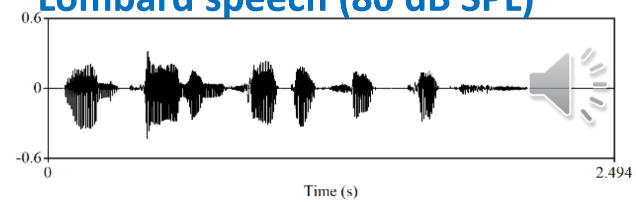
"She missed a very obvious solution."

Lombard speech (70 dB SPL)



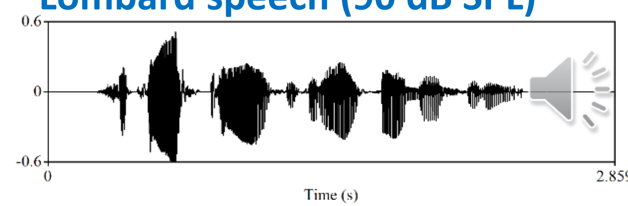
"He bit the head off the gingerbread man."

Lombard speech (80 dB SPL)



"I am craving chocolate chip cookies."

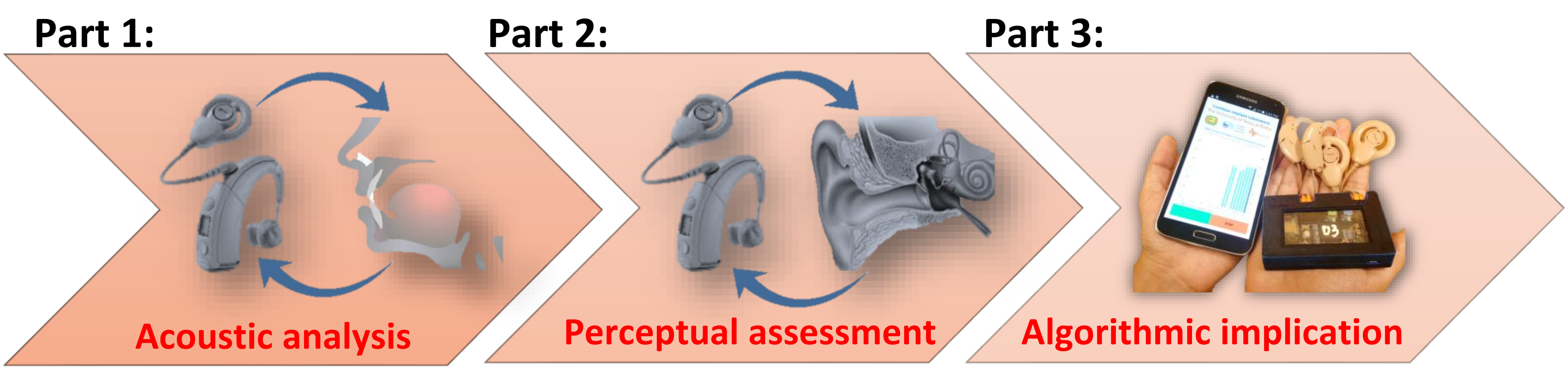
Lombard speech (90 dB SPL)



"She crossed paths with an old friend yesterday."

Thesis Outline

- **Background:**
  - Fundamentals of cochlear implants (CI)
- **Proposed research:**
  - **Part 1***:* *Acoustic* ***analysis*** of Lombard effect by CI users
  - **Part 2**: *Perceptual* ***assessment*** of Lombard speech by CI users
  - **Part 3**: *Algorithmic* ***implication*** of Lombard effect for CI users
- **Thesis contributions**

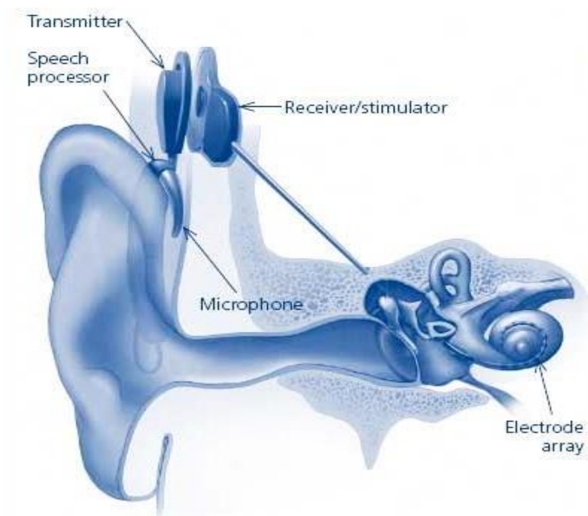


Cochlear Implants

- **Cochlear implant (CI)** - electronic device that is surgically implanted in *the inner ear*.
- Directly stimulates the basilar membrane to provide *partial sense of sound* (Loizou, 1999; Zeng *et al*.,2008).
- Children and adults who are *deaf* or *severely hard-of-hearing* can be fitted for this device.
- The field of cochlear implants has experienced a considerable growth over the past few decades (*324,000 users worldwide* (NIDCD, 2011)).

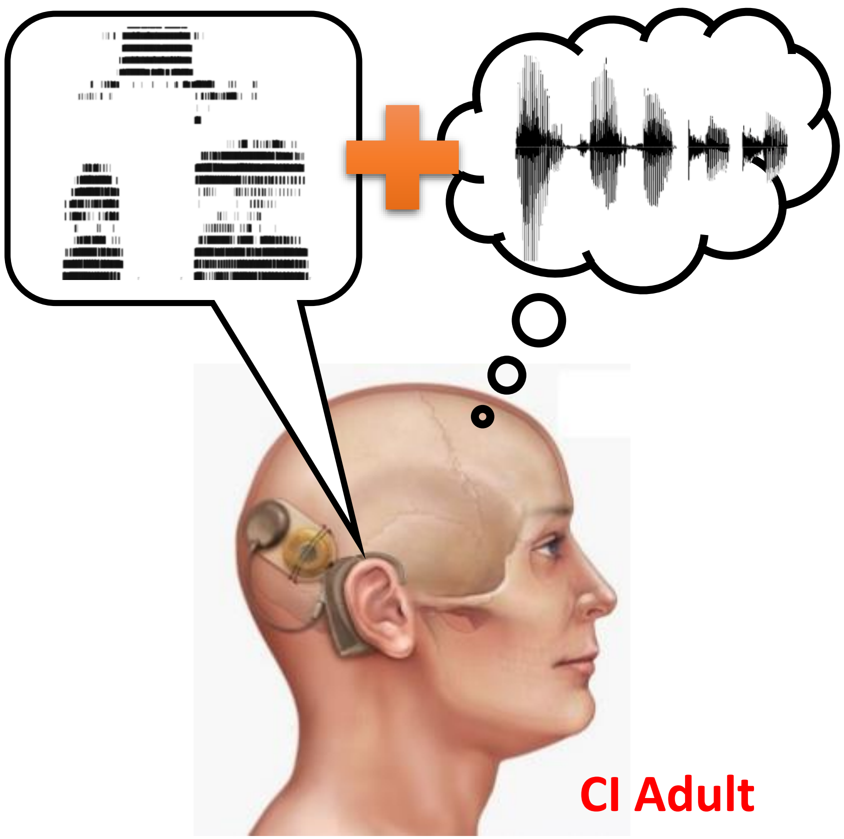
This study focused on *post-lingually deaf CI users*.

CI Device & User:



Cochlear Implants – Post-lingual Deaf Users

Speech Perception of Post-lingual CI Adult:



- **Post-lingual deaf -** adults who lost hearing after the age of 18.
- Able to learn *how to associate* the signal provided by an implant with the sound they remember.
- Allows to *identify speech* without lip-reading or sign language.

➔ Only limited studies have been performed in the area of *speech production of CI adults*.

Cochlear Implants – Speech Production of CI Adults

- **Long-term longitudinal study:**
  - Restored auditory feedback after CI is a crucial factor for *increased adult speech perception* (Bilger *et al*.,1977).
  - *Improved quality of speech production* when compared to corresponding quality before single channel implantation (Hochmair-Desoyer *et al*., 1981).

- **Short-term immediate study:**
  - Rapid change in *formant frequencies* when turning speech processor either on or off in short-time (Svirsky and Tobey, 1991).
  - Immediate response for many speech parameters (*F0, vowel duration, etc.*) to short-time constraints (within few seconds or less) (Svirsky *et al*., 1992).

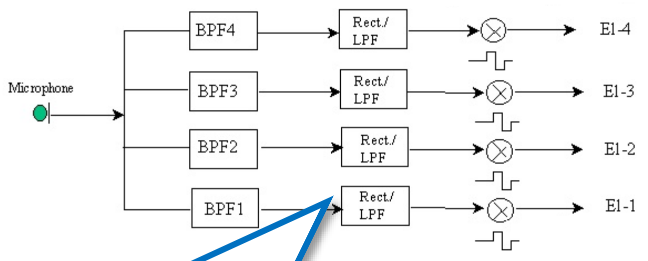
These findings established only when auditory feedback was artificially distorted (modified manually).

Long-term longitudinal effect: Pre-Implantation → Post-Implantation

Short-term immediate effect: Processor Turned-On → Processor Turned-Off

Cochlear Implants – Performance in Noisy Environments

Electric Hearing under Noise:



- Few electrodes
- Limited pulse rate
- Neural survival
- Electrode position

- CI listener perform well in quiet environments (>95% WRR); but their *performance degrades significantly by background noise* (limited temporal/spectral resolution).
- Proposed numerous *speech enhancement algorithms*:
  - Single- and multi-channel (Dolco, 2005; Loizou, 2005)
  - Channel selection (Kim, 2010; Hazrati, 2013)
  - Environmental optimization (Hu, 2010; Goehring, 2015)

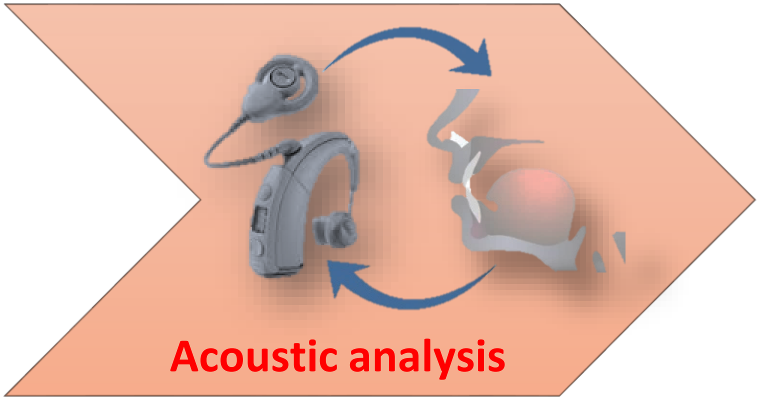
Mostly focused on *noise suppression strategies*.

Thesis Contributions

- **Contribution #1 -** Acoustic analysis of speech production by CI users with respect to noisy environmental changes.
- **Contribution #2** – Pairwise comparison of Lombard speech between CI adults and NH age-mates.
- **Contribution #3 -** Perceptual analysis of Lombard effect by CI users in noisy environments.
- **Contribution #4 -** Development of a Lombard effect-based speech enhancement algorithm for CI users.

Part 1

Effect of environmental noise on speech production of CI users: a naturalistic study



Objectives

- **Problem statement:**
  - Not all of the Lombard research goals have been achieved to date.
  - It is still unknown if CI users employ ***Lombard effect*** in noisy conditions.

- **Goal 1** - Analysis of *speech production of CI users* with respect to diverse environment conditions.

- **Goal 2** - Investigate the effect of auditory feedback on speech production in *everyday naturalistic environments*.

Real-world Environments/Scenarios:

UTD-CI-LENA Corpus Development

Setup for Naturalistic Data Acquisition:

- **Personal audio recording -** *<u>A long single-session</u>* audio stream (8-16 hrs.).
- Collected over an individual's *<u>daily environments/scenarios</u>*.
- **LENA unit** – A portable digital audio recorder capable of up to *<u>16 hrs. of continuous audio</u>* (LENA Foundation, 2014).

Prior advancement in this domain include the "***<u>Prof-Life-Log</u>***" *<u>longitudinal study</u>* at UT-Dallas (Ziaei *et al*., 2012; Ziaei *et al*., 2013)

Subjects

- Six *post-lingual deaf CI users* (mean age: 65 yrs.) participated and produced read and spontaneous speech.
- The same number of *NH speakers* (mean age: 37 yrs.) were participated as a conversational partner.
- IEEE sentences and a list of general topics to converse were given to participants (*e.g., sport, news, food, etc.).*

Biographical Information for CI Participants:

Speaker	Gender	Age (yrs)	Years of hearing loss	Years implanted	Etiology of hearing loss	Implant ear	Sound coding strategy
SPK1	Female	61	56	11	Hereditary	Bilateral	ACE
SPK2	Female	52	48	5	Hereditary	Bilateral	ACE
SPK3	Female	61	14	4	Hereditary	Bilateral	ACE
SPK4	Male	67	12	6	Hereditary	Left only	ACE
SPK5	Male	81	55	9	Hereditary	Bilateral	ACE
SPK6	Male	71	18	4	Unknown	Bilateral	ACE

Controlled On-Campus Environments

Six Locations on College Campus:

1. Office/Lab

Stationary (PC Fan)
1-5 People
1/10 Stationarity

2. Hallway

Impulsive (foot, door)
5-10 People
3/10 Stationarity

3. Lobby

Reverb + Impulsive
10-50 People
5/10 Stationarity

4. Outside

Babble + Wind
30-50 People
5/10 Stationarity

5. Cafeteria

Babble + Competing Talker
50-200 People
9/10 Stationarity

6. Gameroom

Babble + Music + Impulse
35-75 People
9/10 Stationarity

- Selected 6 *naturalistic environments* on UT-Dallas college campus.
- Noise conditions (*type, mixture, levels*) vary greatly across the conditions.

Note! - *Office environment* will be used as the quiet baseline (≤ 45 dB SPL).

Analysis of Noise & Listening Environments

- **Long-term Averaged Spectra (LTAS)**
- **SNR with Neutral speech (SNRN):**
  - Energy ratio of *neutral speech* in quiet baseline to *noise energy* in each environment, assumed to be without Lombard effect.

$$SNRN = 10\log_{10}\left(\frac{E_{Neutral}}{E_{noise}}\right) \quad (1)$$

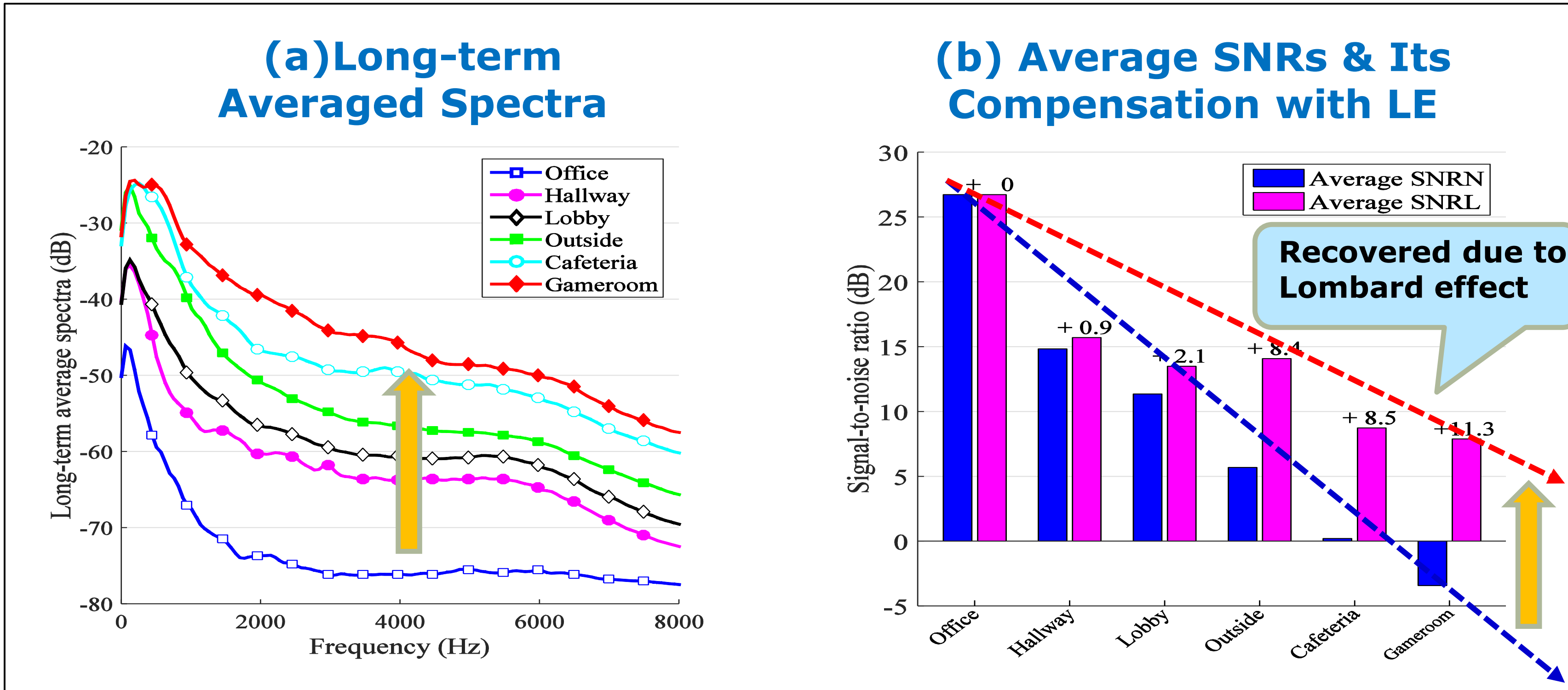
Fixed for every conditions

- **SNR with Lombard speech (SNRL):**
  - Energy ratio between *Lombard speech* in each environment and *corresponding background noise*, assumed to be with Lombard effect.

$$SNRL = 10\log_{10}\left(\frac{E_{Lombard}}{E_{noise}}\right) \quad (2)$$

- Speech and background noise *boundary detections* were used to separate speech from noise.

Results – Analysis of Noise & Listening Environments

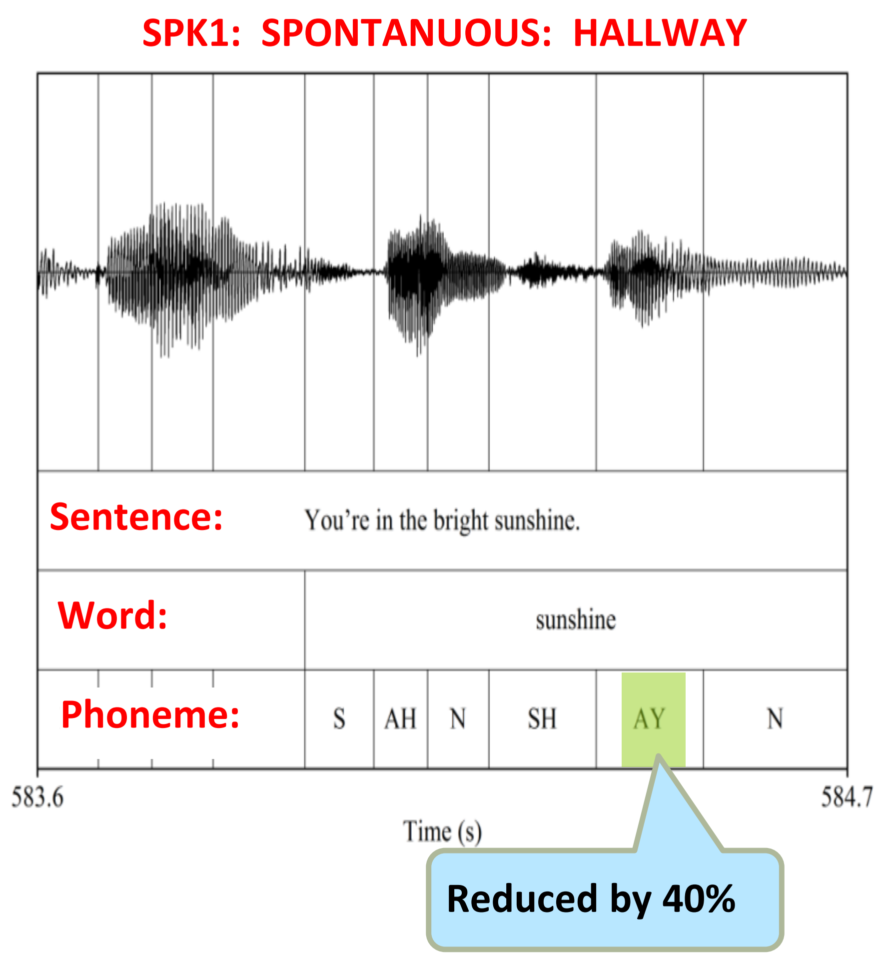


- Average spectral energy *gradually increased*
- SNRN also *decreased* from 27 dB down to -3dB (Gameroom)
- SNRN *recovered due to LE* by up to +11.3 dB (Gameroom).

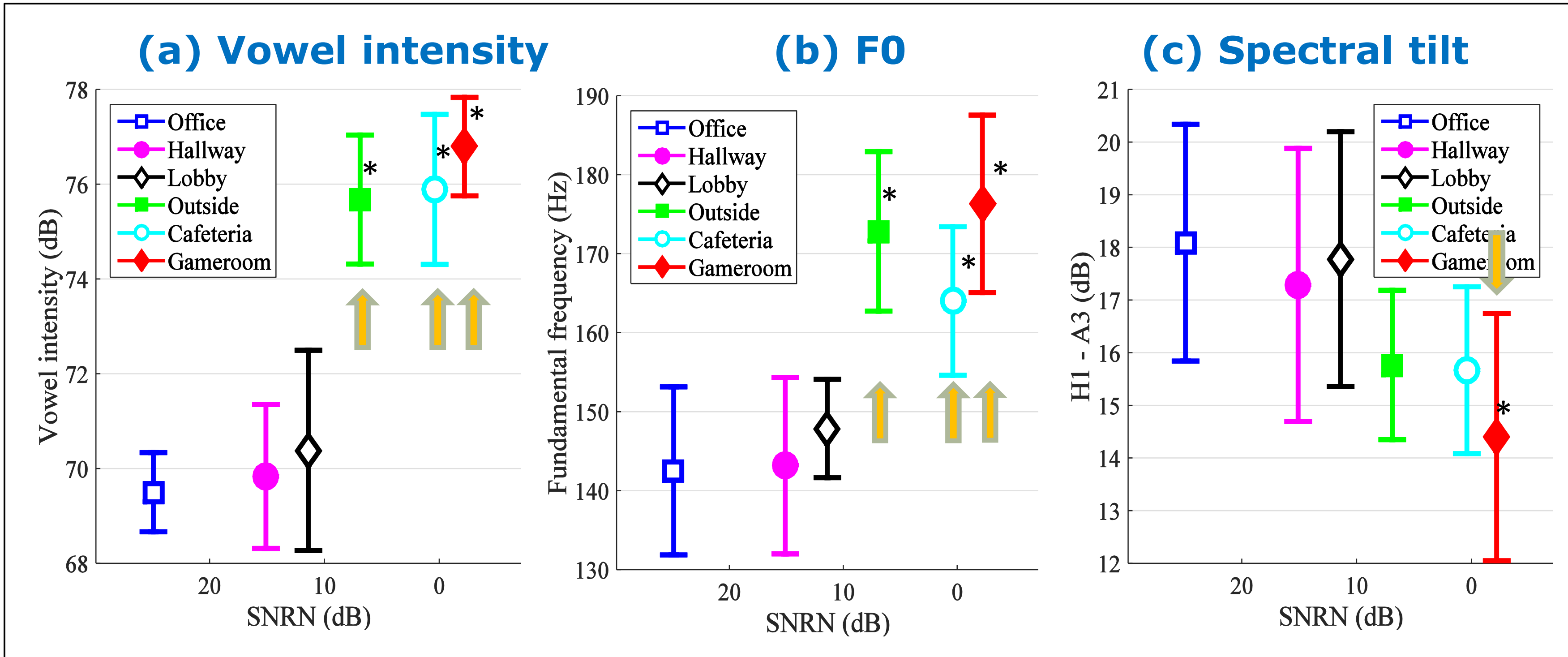
Analysis of Speech

- Measured *speech production parameters* via PRAAT (Boersma, 2002):
  - **Vocal intensity (I)**
  - **Fundamental frequency (F0)**
  - **Overall spectral tilt (ST)**
  - **C-V Intensity Ratio (CVIR)**
  - **C-V Duration Ratio (CVDR)**
  - **First formant frequency (F1)**
  - **Second formant frequency (F2)**
- *Phoneme-level transcription* labels were automatically assigned by forced alignment tool (Yuan, 2008).
- The beginning and ending of each phoneme were reduced by 20% to eliminate any *transitional effect.*

Example of Acoustic/Orthographic Transcription:

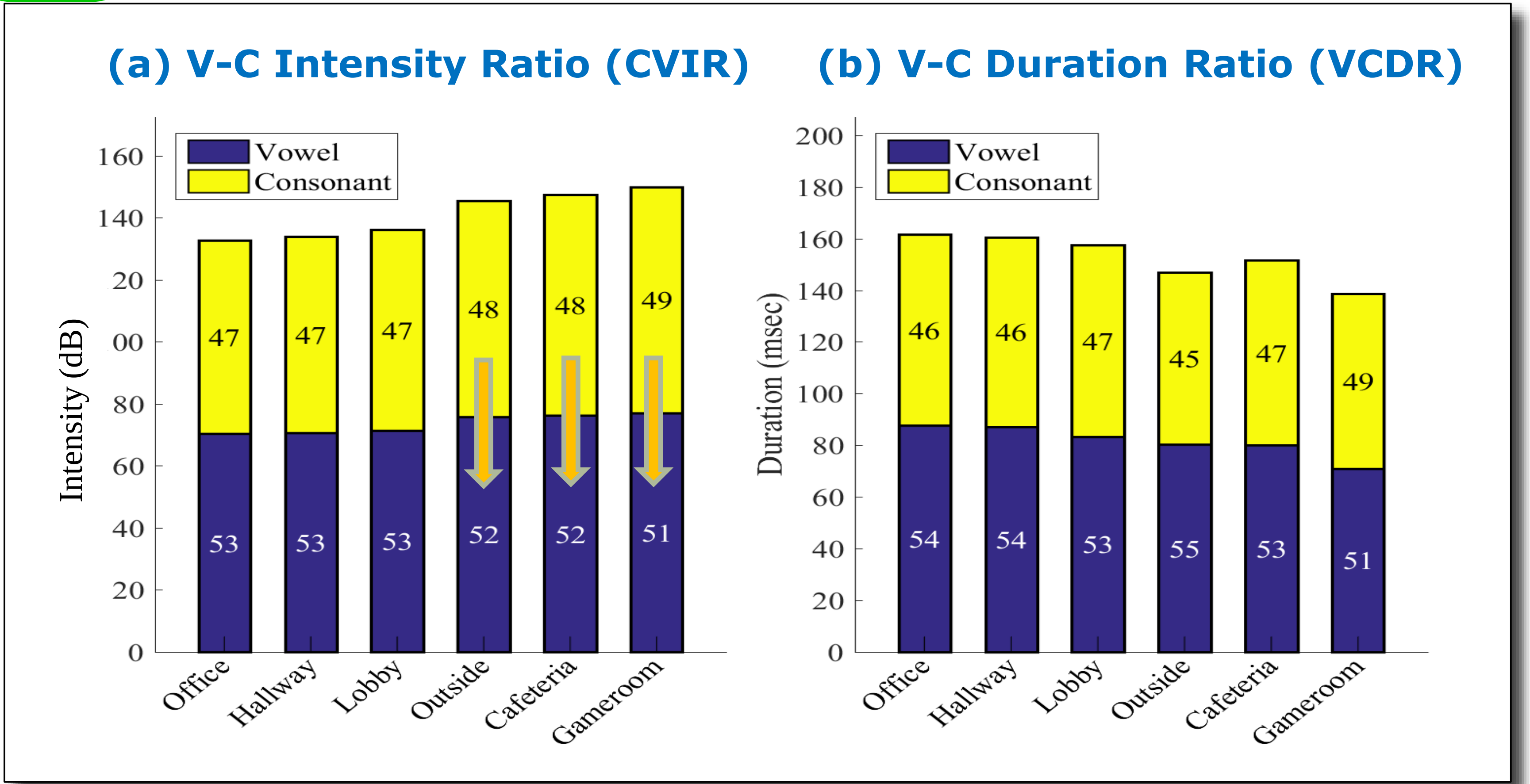


Results – Vowel Production



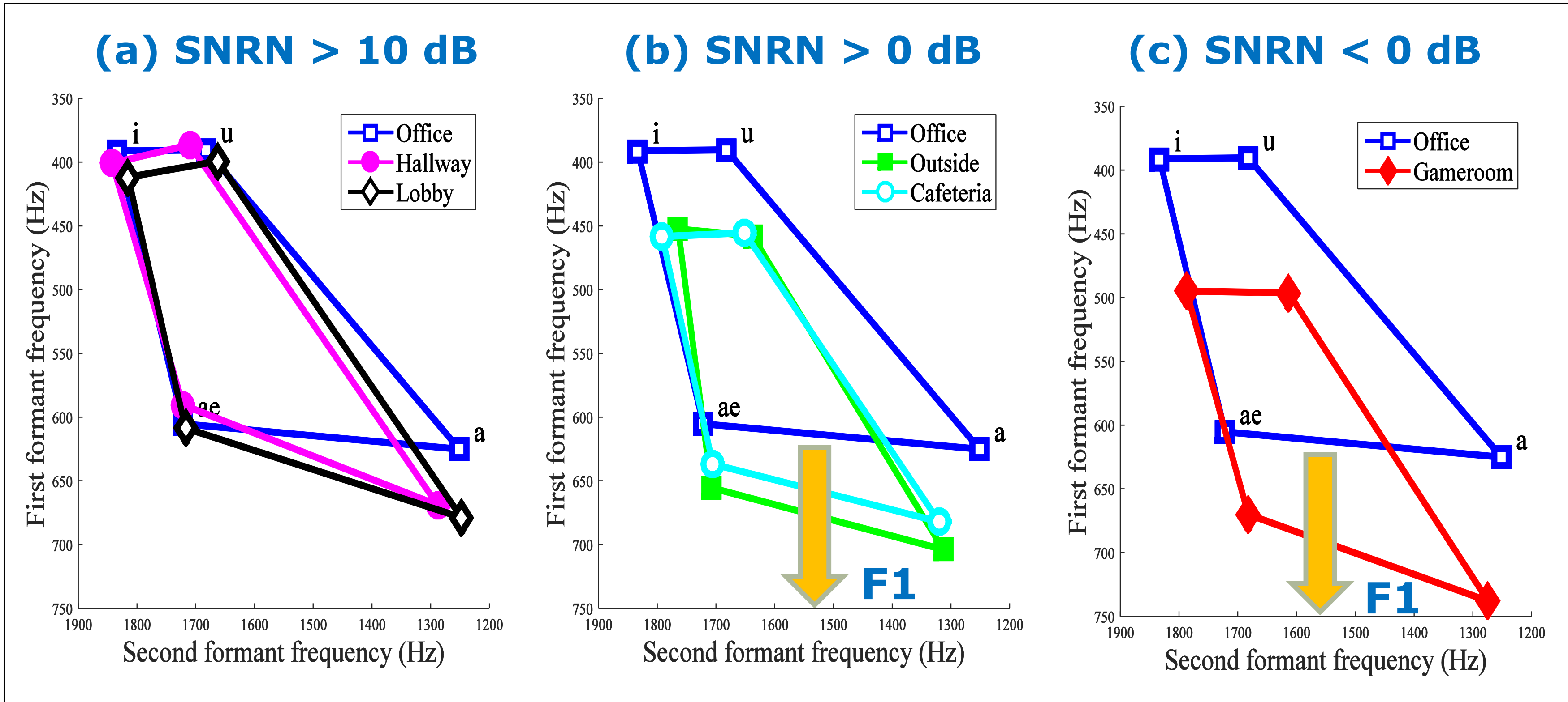
- I, F0 *increased significantly* (Out, Cafe, Game)
- ST *decreased significantly* (Game) - Higher spectral energy emphasized

Results – V-C Ratios



- Consonant intensity *significantly increased* (Out, Cafe, Game).
- Known that consonant is more important to intelligibility.
- *No significant shift* in duration between classes.

Results – Vowel Space



- Significant changes in F1 (/a/,/ae/,/i/,/u/) (<10 dB SNR)
- Only little changes in F2 (/a/,/ae/,/i/,/u/)

Results – Summary of Lombard Parameter Analysis

- Grouped into two areas: *low* (SNRN >10dB) and *high* (SNRN ≤10dB) noise conditions.
- Provide summary of important *Lombard effect parameters* by fixing two conditions.
- Employed Analysis of Variance (*ANOVA*) for significance test.

- **I, F0, CVIR ($p < 0.001$)**
- **ST, F1 ($p < 0.002$)**
- **CVDR, F2 ($p < 0.05$)**

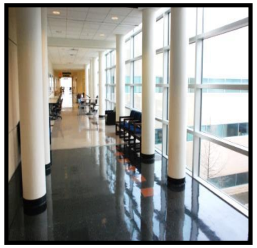
In line with the result of NH studies (*Lane, 1971; Jansen, 1988; Junqua, 1992; Garnier, 2010*).

Multiple Naturalistic Environments:

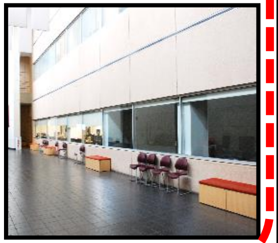
Low noise group (> 10 dB SNR)

1. Office/Lab

2. Hallway



3. Lobby



High noise group (≤ 10 dB SNR)

4. Outside

5. Cafeteria

6. Gameroom

Pairwise Comparison of CI vs NH

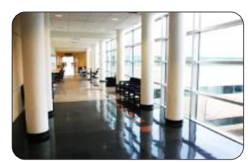
Still unknown how CI speech is different from those of NH in noisy conditions.

- Repeated the *same data collection* with NH listeners for pairwise comparison.
- *Six NH age-mates* (mean age: 47 yrs) participated and produced speech at the *same locations* on UT-Dallas campus.
- Measured the *same speech parameters* used in the CI analysis.
- Performed *ANOVA test* to determine statistical difference between two speaker groups.

Naturalistic Data Collection with NH Subjects

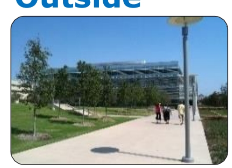
Office/Lab

Hallway



Lobby

Outside



Cafeteria

Gameroom

Results – Pairwise Comparison of CI vs NH

PAIRWISE COMPARISON OF CI VS NH

Acou. Param.		ANOVA
Vowel	Intensity	
	F0	*↑
	Spec. tilt	
	Duration	

Phon. Param.		ANOVA
F1	/a/	**↑
	/æ/	
	/i/	
	/u/	

Phon. Param.		AVOVA
F2	/a/	
	/æ/	
	/i/	**↑
	/u/	

Significance levels: ***<0.001, **<0.01, and *<0.05.

- Most CI parameters have *similar pattern* changes with NH results.
- Some CI parameters (F0, F1 for /a/, F2 for /i/) are *significantly different* from NH.
- These differences may be due to *partial restoration* of auditory feedback by CI device (*e.g.,* poor in pitch and formant perception)

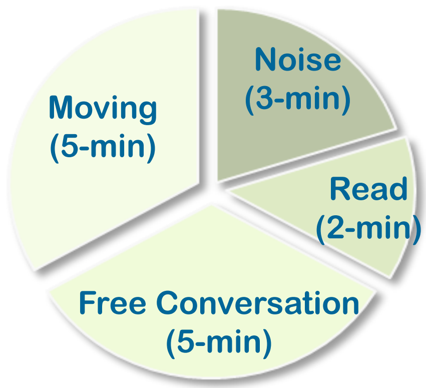
Summary & Discussion

- ***Lombard effect*** has been found in the speech production of CI users who are post-lingually deaf adults.
- Auditory feedback interacts with both *suprasegmental* (I, F0, ST) and *segmental properties* (CVIR, CVDR, F1, F2) in naturalistic context.
- Demonstrated acoustic and phonetic feature patterns in Lombard speech which are *similar with NH listeners*.
- These consistency mainly due to the *presence of auditory feedback* provided by a CI device, allowing more nearly typical pattern of speech in response to noise.

Research Opportunity – Part 1

UTD-CI-LENA Corpus:

Controlled (24 hrs):



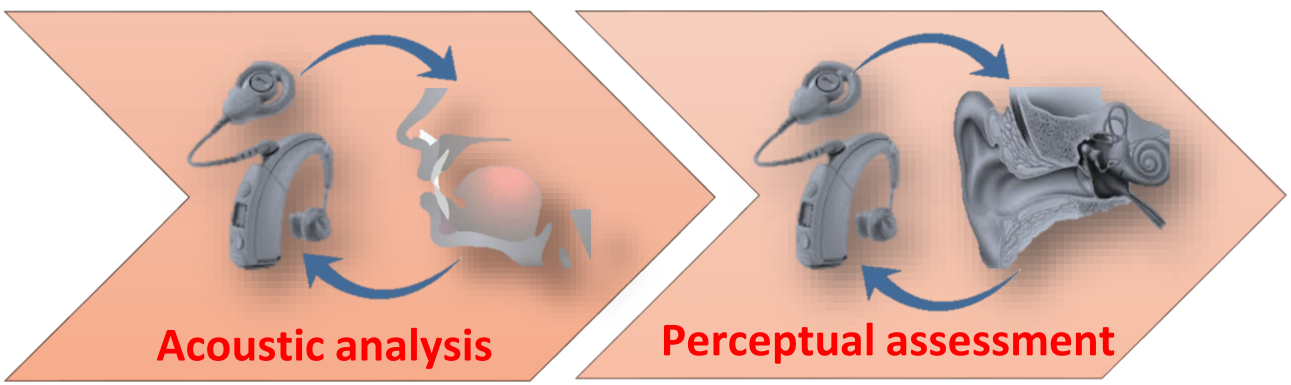
Take Home (75 hrs):

- A total of 24 hours of *personal audio recordings* were collected from 24 speakers (6 CI and 18 NH) users while participated in college campus.
- Additional 75 hours of *naturalistic audio* were collected from everyday situation (e.g., home, restaurant, store, etc.).
- A set of *acoustic and orthographic transcription labels* were assigned by a human annotator.
- Potential application areas for CI users:
  - Language acquisition/development
  - Speech-related disorder
  - User/environmental customized coding algorithm

  (Gilkerson and Richards, 2008)

Part 2

Influences of Lombard effect on speech intelligibility in CI users



Objectives

- **Problem Statement:**
  - No study has been performed to if LE influences *speech perception performance* for CI users.
  - There are very limited data concerning how CI users respond to the *different speaking styles* (*e.g.*, whisper, soft, loud, etc.)
- **Goal 1 -** Examine the influence of *Lombard effect* on speech perception of post-lingually deaf CI users.
- **Goal 2 -** Investigate how *perception performance differs* from speech produced in various noisy environments.

Perceptual Difference of Speaking Styles:

Corpus Formulation with NH Speakers

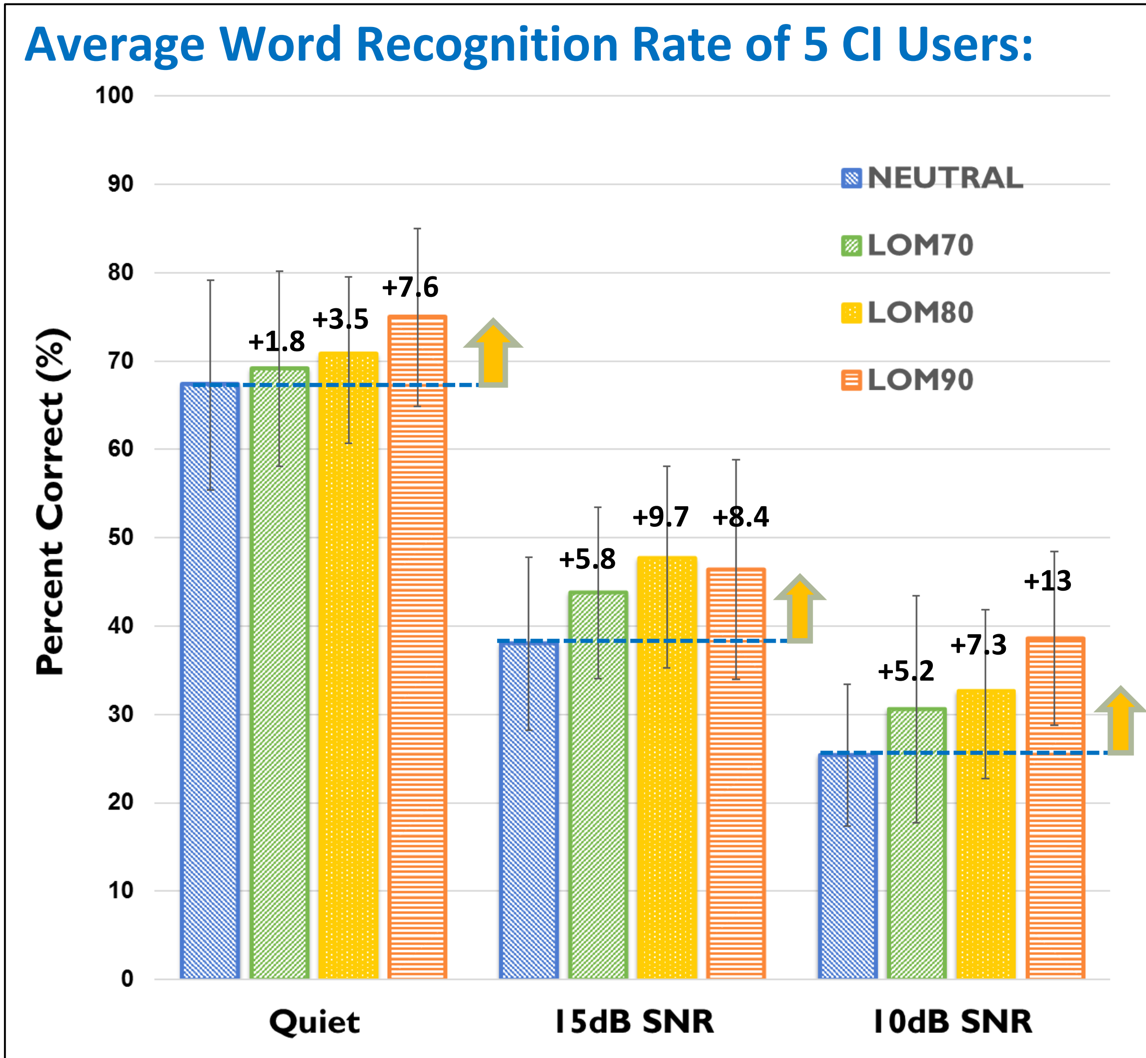
- Four *NH speakers* participated to read AzBio sentences (Spahr, 2012) in sound recording booth.
- *Speech partners* were seated 1m in front to speakers and listen to the speech.
- *Large crowd (LCR) noise* were presented via open-air headphone at 70dB, 80dB and 90 dB SPL.

Subjective Evaluation with CI Listeners

- Five *post-lingual CI users (mean age: 64 yrs.)* participated.
- Original clean stimuli were *corrupted by large crowd noise* at 10 dB and 15 dB SNR.
- CI subjects listened to the stimuli presented via *loud speakers*.
- *Recognition score* was calculated based on the number of words identified.

Subjective Listening Test with CI:

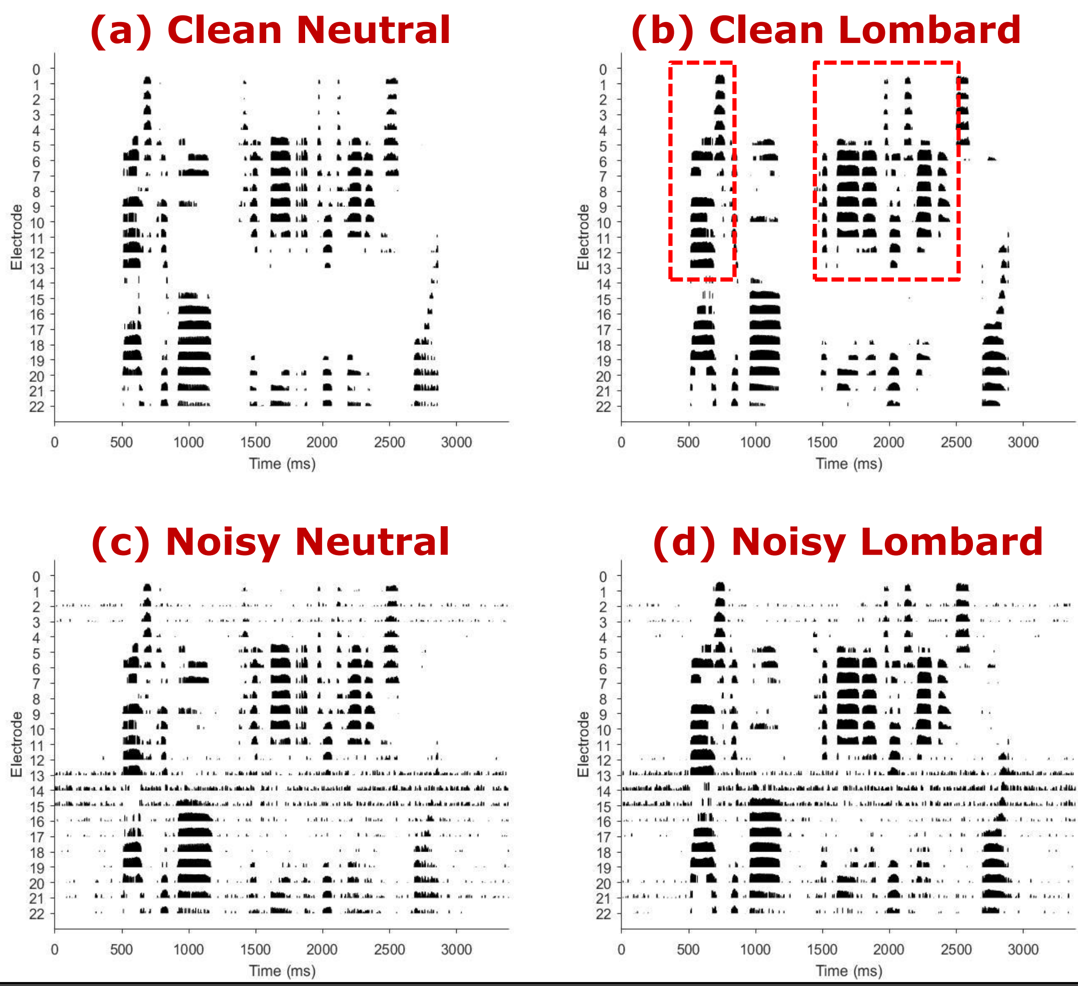
Results – Intelligibility of Lombard Speech



- *Improvement in intelligibility* when Lombard speech was presented.
- Larger improvement in *challenging listening environment* (10 dB vs 15 dB SNR)
- *Higher vocal effort speech* (Lom 90) are more intelligible than lower vocal effort speech (Lom 70).

Results – Example Stimuli Output Pattern

Clean Speech Mixed with SSN at 10 dB SNR:



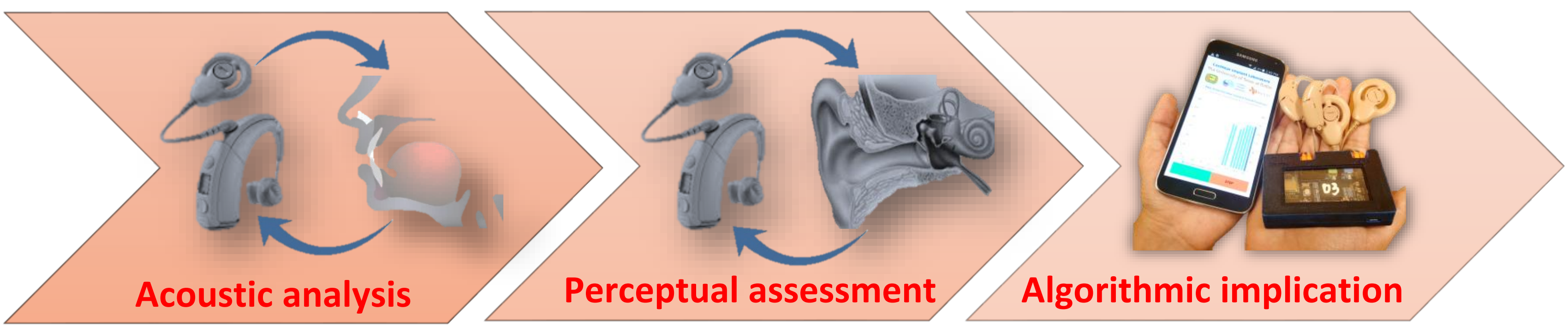
- LE speech provided increase in electrical activity on *high frequency channels* (1-16).
- Cause *consonant and formant signal* tend to be more enhanced.
- More evident in noisy sentences (noise *distorts the lower frequency* channels more).

Summary & Discussion

- *Acoustic and perceptual characteristics* of speech under LE were analyzed.
- Lombard speech had a *higher intelligibility* than neutral speech for electric hearing in both quiet and noisy environments.
- The improvement was larger in *challenging listening environments*.
- The modification of speech *using LE* might contribute to the higher intelligibility.

Part 3

Development of an intelligibility enhancement algorithm based on Lombard effect properties



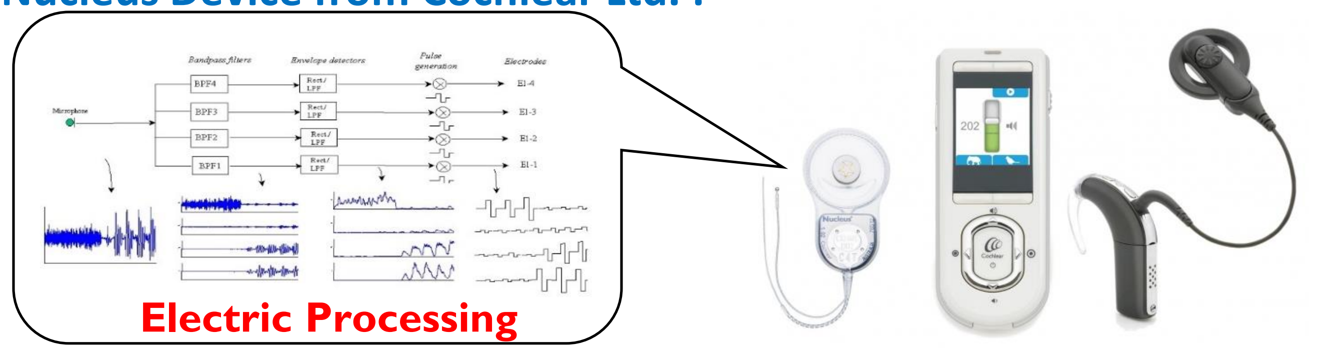
Objectives

- **Problem Statement:**
  - Conventional signal processing for CI have focused on *noise suppression strategies* to improve performance in noisy conditions (not perceptually motivated).
  - It would be desirable to have an algorithmic modification of neutral speech based on *speech production constraints* (*e.g.*, LE, Clear).

- **Goal 1 –** Develop a novel *intelligibility enhancing speech modification algorithm* based on LE properties.

- **Goal 2 –** Examine how CI users perceive the *artificially modified Lombard speech* in noisy environments.

Nucleus Device from Cochlear Ltd. :

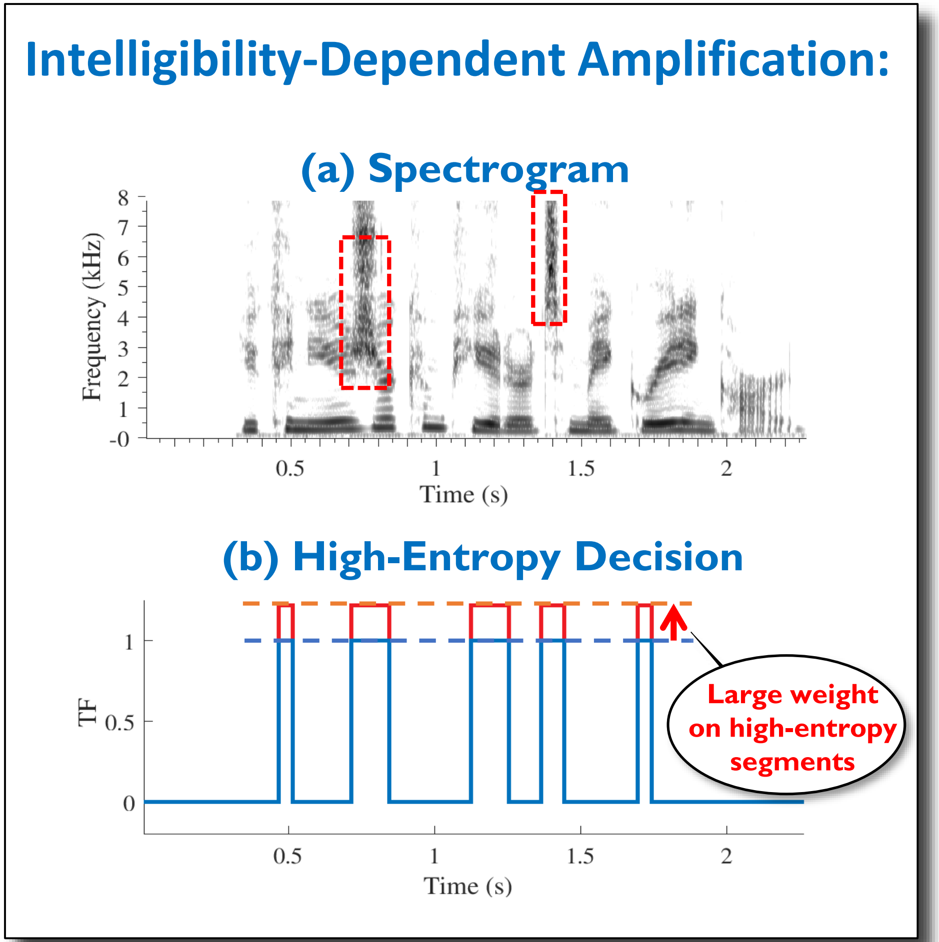


Proposed Algorithm – System Description

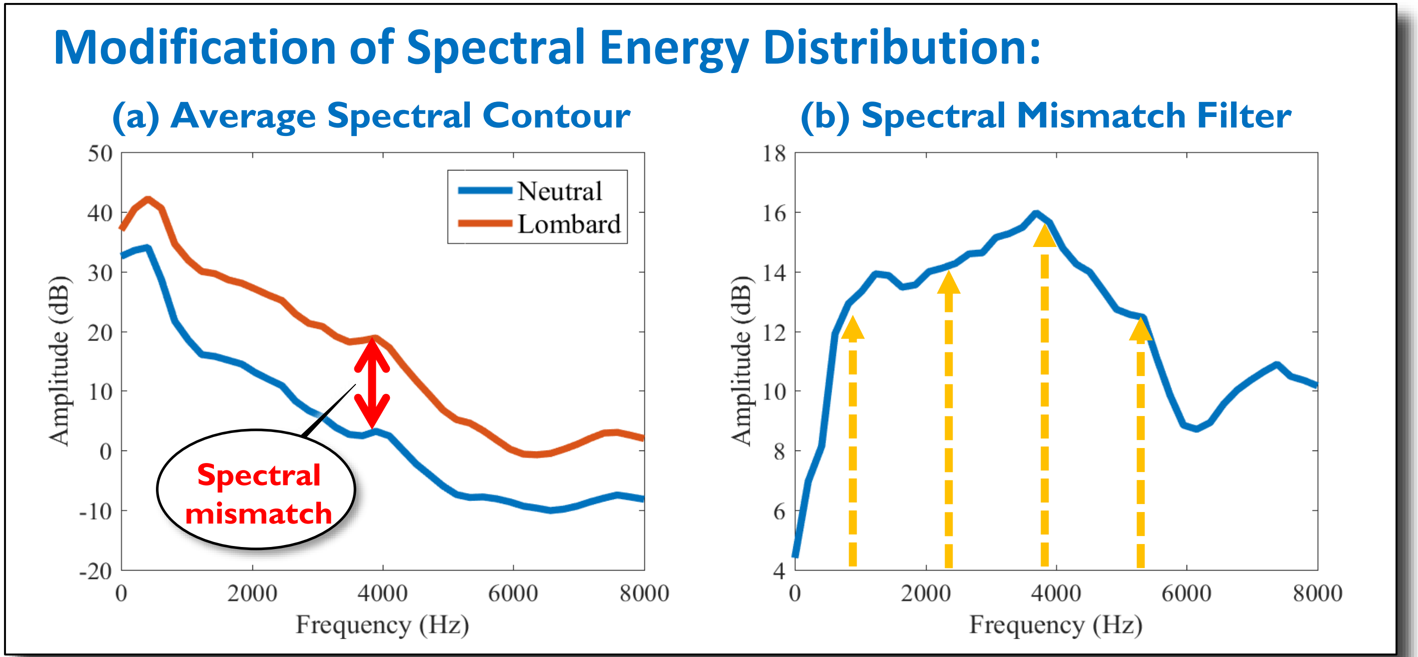
- Controlled the acoustic features of *input neutral speech* to present *Lombard speaking-style output*.
- The *modification areas* considered here were:
  1. ***Temporal amplification***
  2. ***Overall spectral contour***
  3. ***Sentence duration***
- Modeled *acoustic variation* for neutral and Lombard speech.
- Dataset for parameter modeling was derived from *UT-Scope stressed speech corpus* (Ikeno, 2007).

Feature 1 - Entropy-Based Temporal Amplification

- Aim to increase amplitude for high intelligibility segments (*e.g., consonants, V-C boundaries*).
- *A cochlear-scaled entropy* (Stilp and Kluender, 2010) was used to estimate the information bearing segments.
- Place a large weight on *high-entropy segments*.



Feature 2 – Spectral Contour Transformation

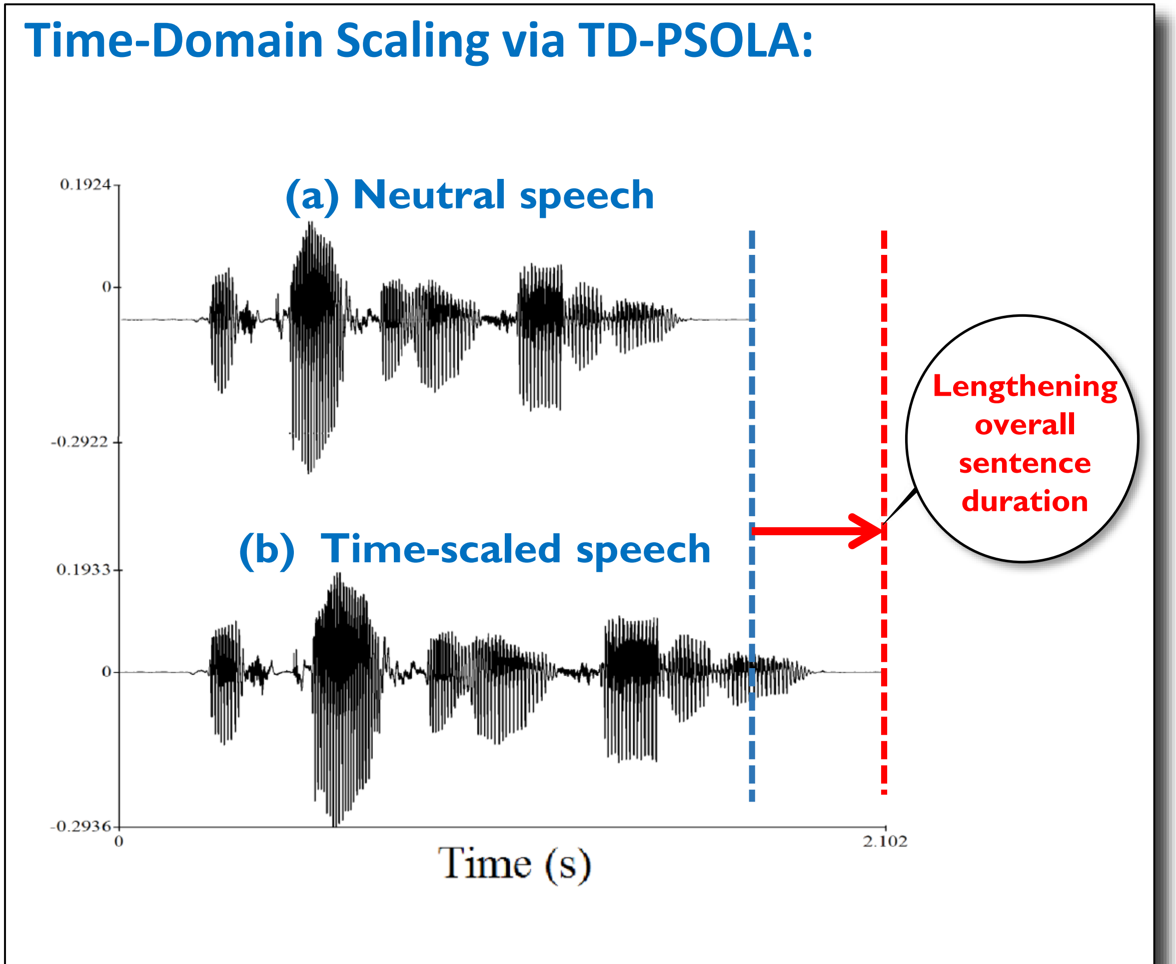


❖ Designed to increase *high frequency structure*; important for *channel selection* by CI platform.

❖ A time-invariant filter was estimated by computing *spectral difference* between neutral and Lombard speech.

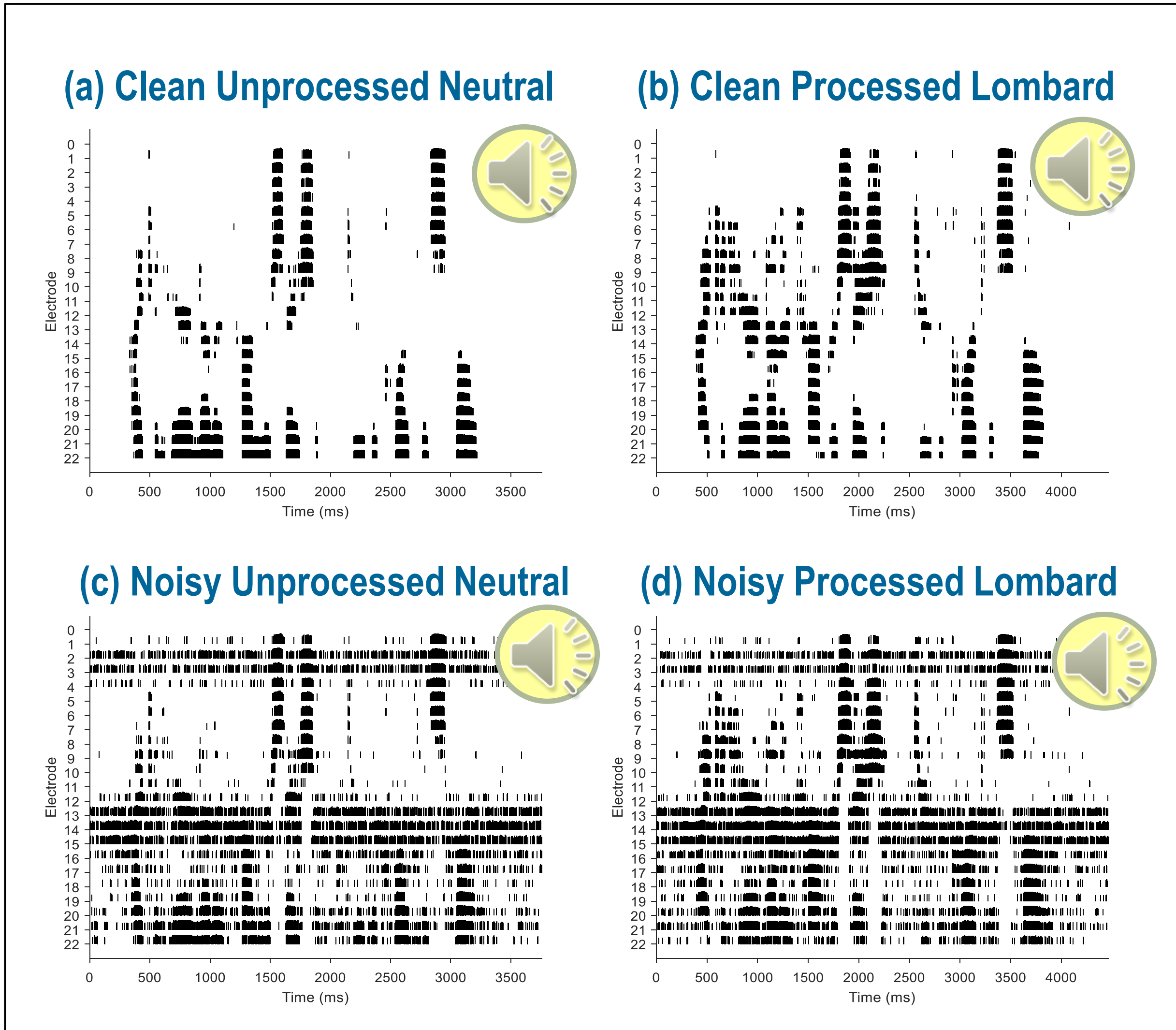
❖ Used to adjust *overall spectral balance* of the input signal.

Feature 3 – Uniform Time Stretching



- Lengthening speech duration allowed listeners have *more chances at hearing*; improve CI decoding opportunity
- *TD-PSOLA technique* (Moulines and Charpentier, 1990) was employed to account for the duration variation.

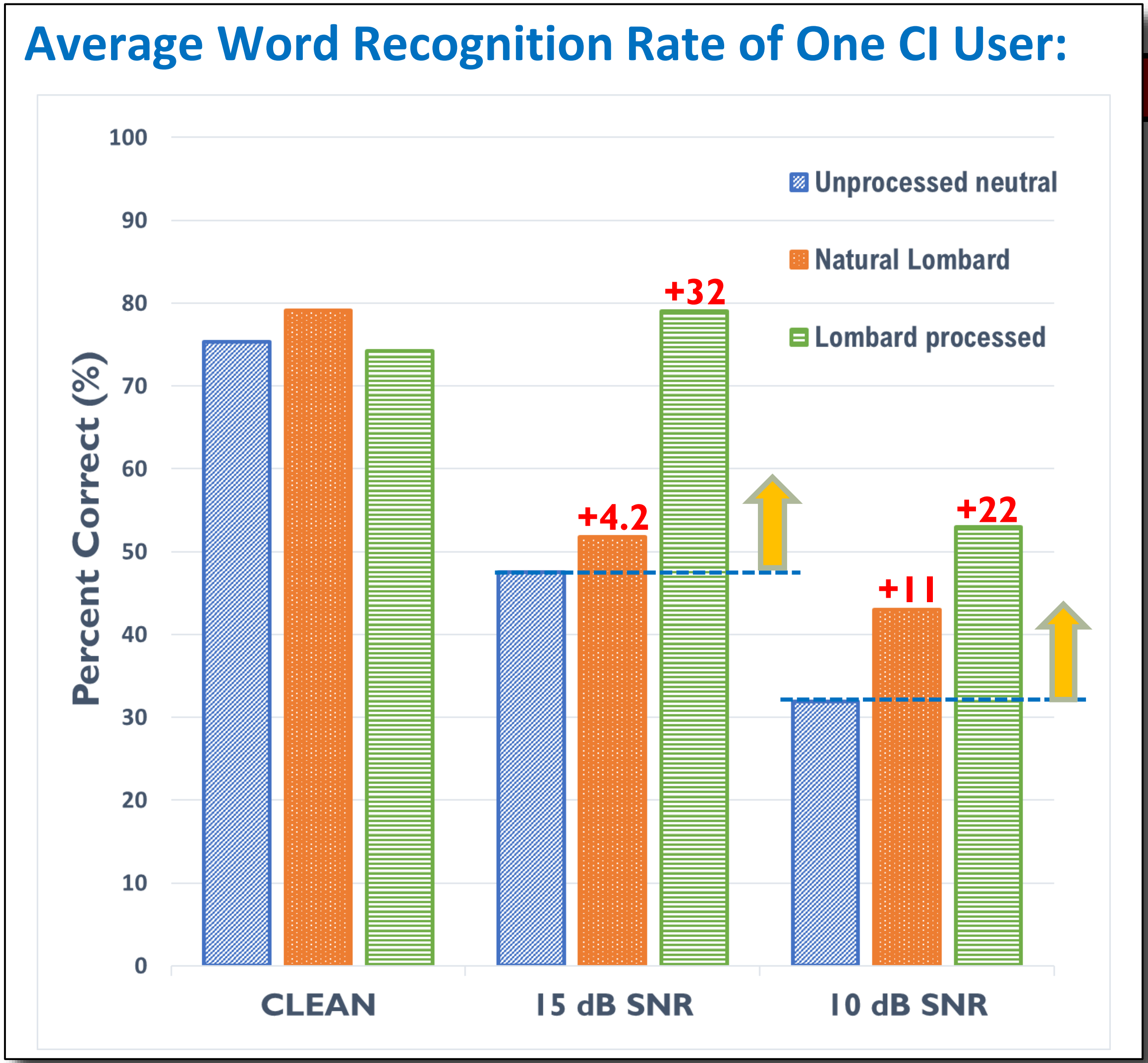
Example "Artificial Lombard" Sentences



Evaluated by CI user:

- A CI subject (age of 64) listened to the *original clean stimuli* corrupted by large crowd noise.
- *Natural Lombard sentence* were also presented for comparison.

Results – Subjective Evaluation with CI User



- Increases in intelligibility was found for Lombard processed speech, *particularly in noisy environments*.
- Larger increases were achieved for *Lombard processed speech* when compared to Natural Lombard speech.
- Largest benefit, + *32%*, was measured in 15 dB SNR.

- Will increase the number of subject.

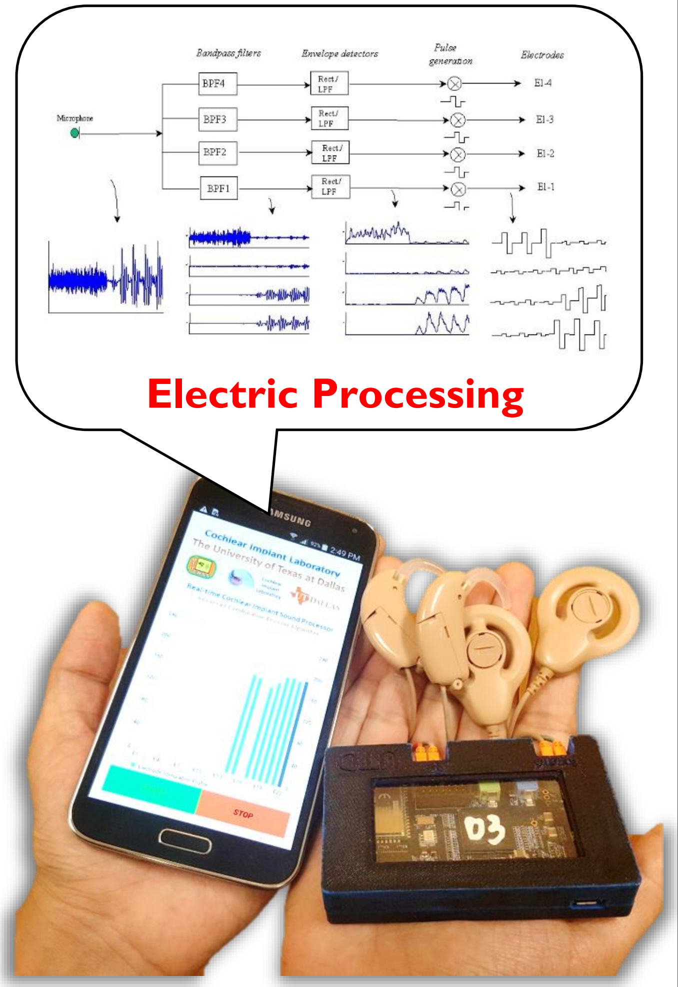
Summary & Discussion

- A new speech modification criterion using *Lombard effect characteristics* was proposed.
- *Improvement in intelligibility* was found with Lombard processed speech in noisy environments.
- The improvement in intelligibility was attributed to the modification of speech *via the proposed algorithm*.
- The result provided potential of the Lombard effect based *speech enhancement algorithm* for CI users.

Research Opportunity – Part 2, 3

- In contrast to the traditional methods for noise suppression, the proposed approach *motivated by natural human speech physiology*.
- This paradigm is more natural, thus, *desirable for hearing impaired individuals*.
- Algorithmic advancement proposed here offers a unique opportunity to *improve the listening/decoding experience* of CI users.

Mobile Research Platform for CIs:



Thesis Contributions

- **Contribution #1 -** Acoustic analysis of speech production by CI users with respect to noisy environmental changes.
- **Contribution #2** – Pairwise comparison of Lombard speech between CI adults and NH age-mates.
- **Contribution #3 -** Perceptual analysis of Lombard effect by CI users in noisy environments.
- **Contribution #4 -** Development of a Lombard effect-based speech enhancement algorithm for CI users.

Publications

- **Journal papers**
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, Emily A. Tobey, John H.L. Hansen, "Acoustic-phonetic analysis of speech communication with cochlear implant users under noisy Lombard environments: a naturalistic study," *Journal of the Acoustical Society of America* (Accepted).
  - Oldooz Hazrati, **Jaewook Lee**, Philipos C. Loizou, "Blind binary masking for reverberation suppression in cochlear implants," *Journal of the Acoustical Society of America, Vol. 133 (3), pp. 1607~1614, March, 2013.*
- **Conference papers**
  - **Jaewook Lee**, Hussnain Ali, John H.L. Hansen, "Intelligibility enhancement of neutral speech based on Lombard effect modification with application to cochlear implant users," *in Proc. Annual Midwinter Meeting of Association for Research in Otolaryngology (ARO 2017)*, Baltimore, MD, February, 2017
  - **Jaewook Lee**, Hussnain Ali, John H.L. Hansen, "The Lombard reflex and its influences on speech perception in adult cochlear implant users," in Proc. CI 2016, Toronto, Canada, May, 2016.
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, Emily A. Tobey, John H.L. Hansen, "Impact analysis of naturalistic environmental noise type on speech production for cochlear implant users versus normal hearing listeners," *in Proc. CIAP 2015*, Lake Tahoe, CA, USA, July, 2015.
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, John H.L. Hansen, "Analysis of speech and language communication for cochlear implant users in noisy Lombard conditions," *in Proc. IEEE ICASSP 2015*, Brisbane, Australia, April, 2015.
  - **Jaewook Lee**, Hussnain Ali, Ali Ziaei, John H.L. Hansen, "Lombard effect based speech analysis across noisy environments for voice communications with cochlear implant subjects," *in Proc. 168th Meeting of Acoustical Society of America*, Indianapolis, USA, October, 2014.

Acknowledgments

- COMMITTEE
  - Dr. John H. L. Hansen, Chair
  - Dr. Peter F. Assmann
  - Dr. P. K. Rajasekaran
  - Dr. Carlos Busso
- CRSS-CIL Members
- CI & NH Subjects who participated in the studies
- NIH for their project support