Impact of Abnormal Acoustic Properties on the Perceived Quality of Electrolaryngeal Speech

Geoffrey Meltzner
ALPHATECH Inc., Burlington MA

Robert E. Hillman
Voice and Speech Lab, MEEI, Boston MA

August 28, 2003

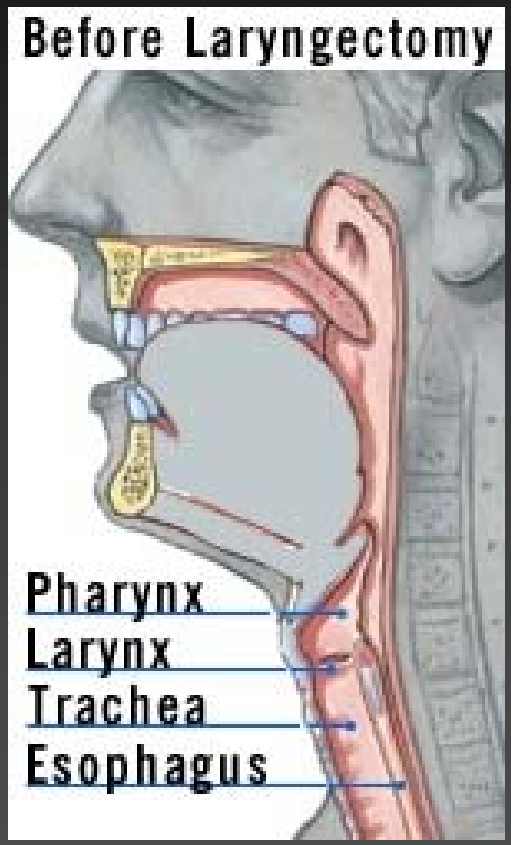
Outline of Talk

- The Electrolarynx (EL) and EL speech
  - Deficiencies of EL speech
- Perceptual Experiments
  - Methods
  - Analysis
  - Interpretation
- Summary and future work.

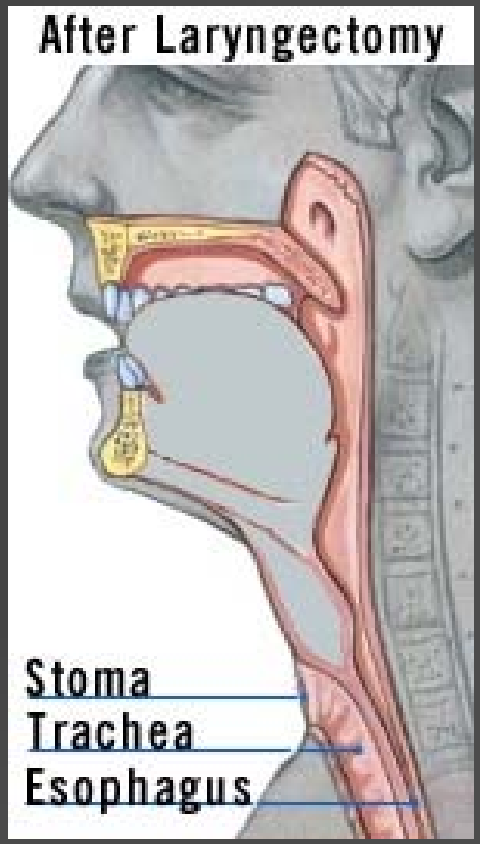
The need for Electrolaryngeal (EL) Speech

- Each year thousands of people in the US alone undergo laryngectomy surgery to treat laryngeal cancer.
- These people need to rely on an alternative means of communication: alaryngeal speech
- Most common form of alaryngeal speech is EL speech (Hillman *et al*.1998).

The Laryngectomy



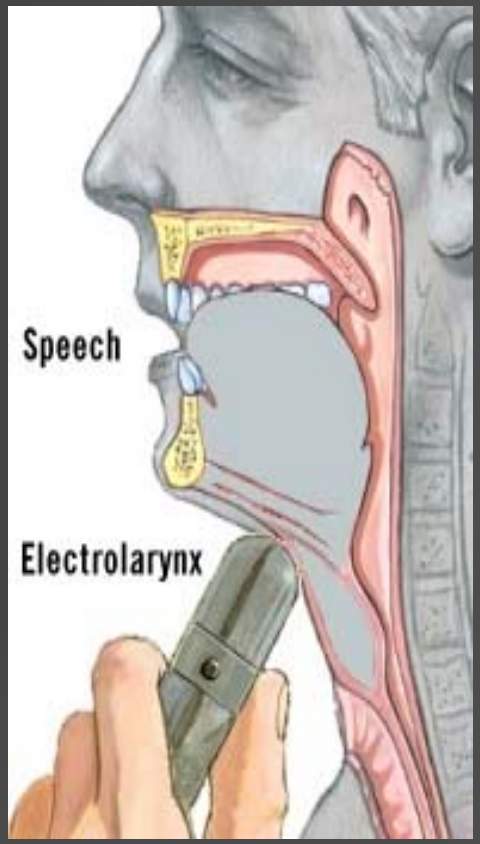
- The larynx and surrounding structures are removed
- Trachea is sewn into the neck to make breathing hole (stoma).
- Vocal tract decoupled from lower airway
- Articulators remain intact



Figures taken from http://www.inhealth.com/

How is an Electrolarynx used?

- User holds device against the neck (or just under the chin)
- Vibrations are transmitted through the neck wall into the vocal tract.
- User articulates normally



Graphic taken from http://www.inhealth.co

Inadequacies of EL Speech

- Artificial, mechanical quality
  - often draws unwanted attention to EL user
- Reduced loudness (relative to normal speech), especially in noisy environments
- Reduced intelligibility
  - Confusion between voiced and unvoiced consonants (Weiss *et al.*1979; Weiss & Basili 1985)
  - Vowel intelligibility ~ 80 % (Weiss & Basili 1985)

Reasons for poor EL speech quality

- Improper source spectrum. (Qi and Weinberg 1991, Weiss *et al.*1979)
- Interference of directly radiated sound
- Lack of fine control over pitch, amplitude and voice on/offset.
- Others?
  - e.g. distortions introduced to the vocal tract transfer function

Attempts at Improving EL Speech

- Source
  - Qi & Weinberg (1991): attempted to reduce the “low frequency deficit” in EL speech.
  - Ma (1999): replaced EL excitation signal with normal excitation signal
- Direct noise
  - Verdolini *et al.* (1979): amplified signal at lips to increase “SNR”
  - Cole *et al.*(1997): used a combination of noise reduction techniques to remove direct noise
  - Espy-Wilson *et al.* (1998): used adaptive filtering to reduce the directly radiated noise
- Pitch Control
  - Uemi *et al.* (1994) used air pressure from stoma to control the fundamental frequency of EL

Motivation for research

- Previous studies all reported some improvement, but:
  - All were done in isolation
  - No indication of their relative improvements
  - No measure of how much closer improved speech is to normal speech.

Research Goals

- Determine the relative contributions of the properties of EL speech to its artificial quality.
  - (Perceptual Experiments)

Perceptual Experiments

- **Goals:**
  - Establish the relative contributions of different EL speech properties to its unnatural quality.
- **Experimental Procedure:**
  - Create different enhanced versions of EL speech to be compared against each other.
  - Method of Paired Comparisons and Law of Comparative Judgment
    - Establish a ranking for differently enhanced versions of EL speech and normal speech based on how normal they are perceived to be.
  - Visual Analog Scale (VAS)
    - Listeners rate each speech token based on how they differ from normal natural speech

Perceptual Experiments - Methods

- **Experimental Procedure details:**
  - Two speakers (normal male and female)
  - Generate 10 speech tokens per sentence
    - One sentence consisting of all voiced phonemes
      - We were away a year ago where I wore Rollerblades.
    - One sentence with voiced and unvoiced phonemes
      - She tried the cap and fleece so she could pet the puck.
    - Spoken using normal voice and Servox EL
      - Subjects held breaths while using Servox
    - Three enhancements added to EL speech:
      - Low frequency enhancement (L)
      - Removal of direct noise (N)
      - Added pitch modulation (P)
    - Normal speech made monotonous.

Perceptual Experiments - Methods

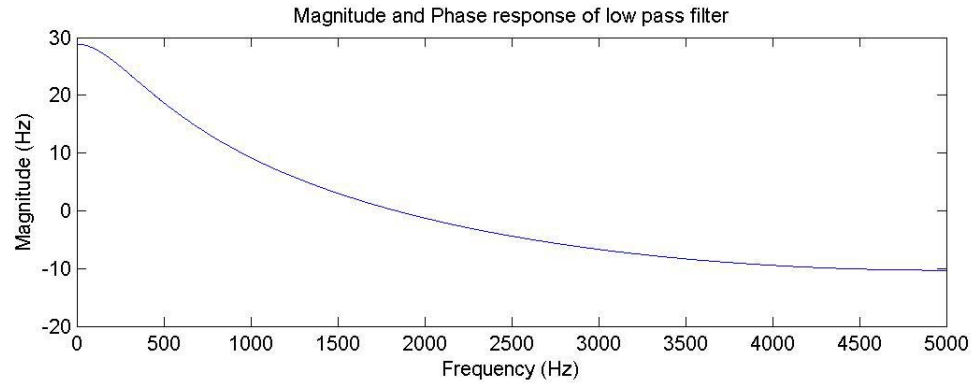
- **<u>Experimental Procedure details (cont):</u>**
  - Speech Tokens to be generated
    - (1) Raw EL Speech
    - (2) EL-L, (3) EL-P, (4) EL-N
    - (5) EL-LP, (6) EL-LN, (7) EL-NP
    - (8) EL-LNP
    - (9) Monotonous normal speech
    - (10) Normal speech

      - L – (Low frequency enhancement)
      - N – (Removal of direct noise)
      - P – (Added pitch modulation)

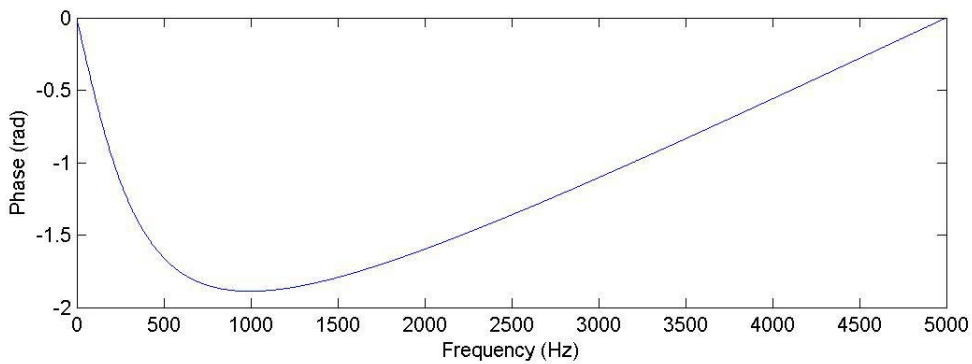
Implementing Enhancements - L

- Low frequency (L):
  - low pass filter speech as in Qi and Weinberg (1991)

$$H(z) = \frac{1}{\left[1 - az^{-1}\right]^2}$$



Before



After

Implementing Enhancements: The “Door”

- Remove direct noise (N):

- Before After

Implementing Enhancements - P

- Added pitch modulation (P):
  - Pitch contour extracted from normal speech and used to modulate EL speech using MELP vocoder
    - Before After
  - Shift monotone EL pitch
  - Flatten pitch contour in normal speech to make monotonous
    - Before After

MELP

- **M**ixed **E**xcitation **L**inear **P**redictive Vocoder
- Based on US Federal standard at 2.4 kbs
  - Modified to make it more perceptually invisible
- Allows for pitch modification without affecting speech quality

Perceptual Experiments

- Example of experimental sentences:
  - Normal
  - Normal – mono
  - EL-LNP
  - EL-NP
  - EL-LP
  - EL-LN
  - EL-P
  - EL-N
  - EL-L
  - EL-raw

Perceptual Experiments

- **Experimental Procedure details (cont):**
  - 45 total pairs presented
  - Pairs presented twice to establish intra-subject reliability
  - Repeated for each speaker/sentence condition
  - Prior to presentation of pairs:
    - The normal speech token presented as a standard.
    - All sentences presented to listener one by one.

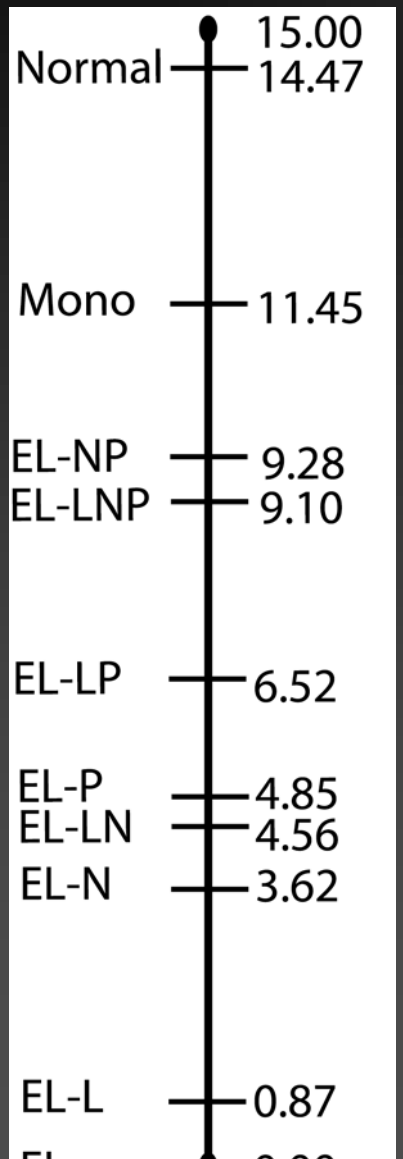
Perceptual Experiments - Analysis

- Law of Comparative Judgment
  - Compute the proportion of times each stimulus is judged "better" than the other stimuli
  - Convert proportions to scale vales on a psychological continuum

Perceptual Experiments – Analysis (cont.)

- Visual Analog Scale
  - The distance from the end of the VAS marked “Not at all different” used as estimate of how different speech token is from normal natural speech.
  - Mean and standard deviation of distances were computed
  - ANOVAs followed by post-hoc Bonferroni corrected *t*-tests were computed to test for significant differences

Perceptual Experiments: Ranking Results

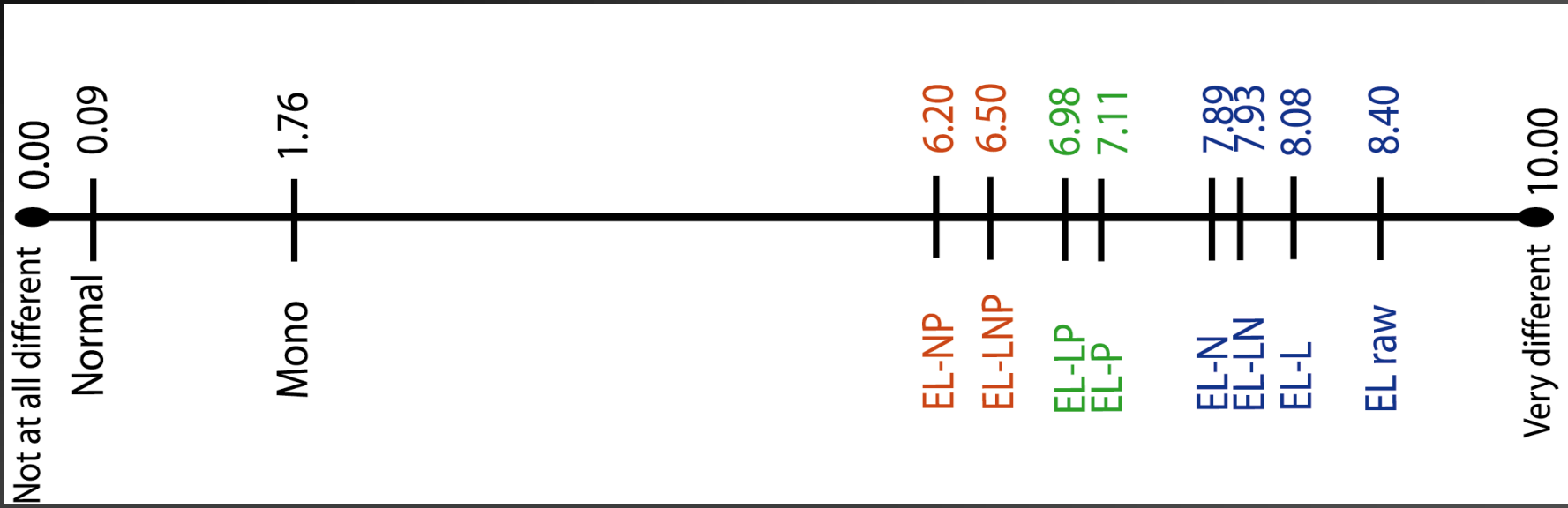


- Monotonous normal speech is the most normal (other than normal speech).
- EL-NP and EL-LNP are the most natural sounding versions of EL speech.
- Much can be done to improve EL-LN speech to make it equal to monotonous normal speech.

Perceptual Experiments: Ranking Results

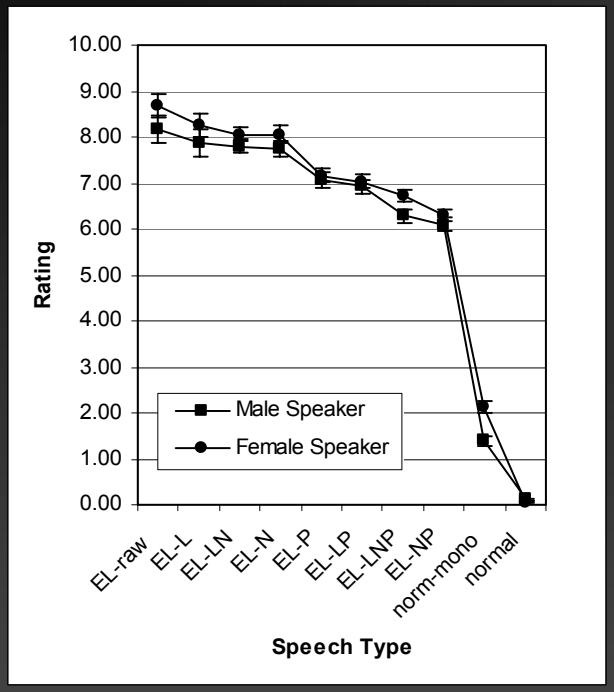
- Listener reliability was found to be 88.3 ± 8.9%
- Rankings of male and female sentences were similar
  - Ranking of EL-LN and EL-P were reversed
- Rankings of EL versions of voiced/voiceless sentence (Sen. 2) were consistently lower than those of their all-voiced counterparts

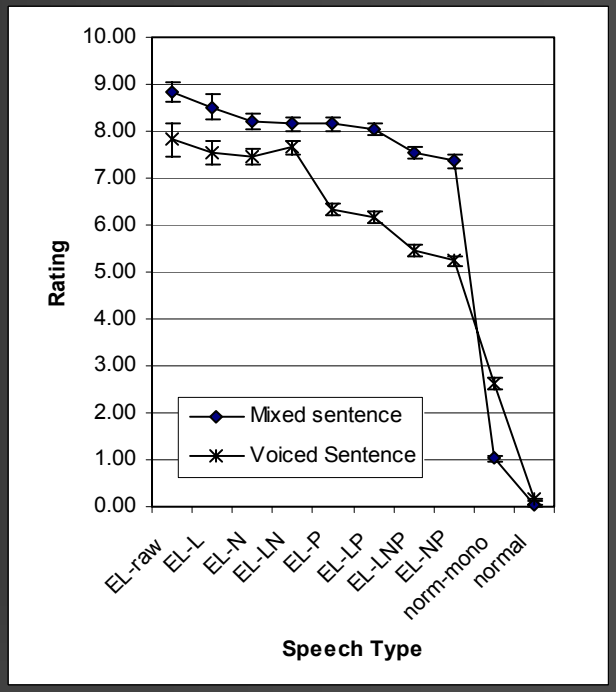
Perceptual Experiments: VAS Results



- Rank order similar to paired-comparison data
  - EL-N and EL-LN are reversed
- EL-NP and EL-LNP are again the best versions of EL speech
- Large gap between normal-mono speech and all EL speech versions
- Ratings of 4 highest rated speech types not significantly different from each other

Perceptual Experiments: VAS Results





- Little difference between ratings of male and femal speakers
- Ratings of voiced/voiceless sentences significantly higher than those of all-voiced sentences

Perceptual Experiments: Summary

- EL-NP and EL-LNP are the most natural sounding versions of EL speech
- Normal-monotonous speech was found to be more natural than any version of EL speech.
  - VAS results indicate a large gap between normal-monotonous and best rated EL speech
- Enhancements not as effective for voiced/voiceless sentences

Perceptual Experiments: Interpretation

- While pitch is important, large improvements can be accomplished without it.
- There are problems with EL speech other than pitch, low frequency deficit and direct noise corruption.
  - Source is still incorrect.
  - Vocal tract transfer function is distorted.

Future Work:

- Look for other potential abnormal properties of EL speech
  - Distortions of vocal tract acoustics
  - EL source has no noise component to it
- Improve EL speech based on findings
  - Develop EL speech enhancement scheme