5PSC20: EM sensor measurements of glottal
structure versus time
I. The Good News is:

- GHz Electro-Magnetic waves easily penetrate human tissue:
- They reflect from all of the speech articulator tissues:
- CHz Electro-Magnetic waves easily penetrate human tissue:

- are very small.
- can be very low in cost, > $5 each, in quantity
- use very low power EM waves, < 1 mw
- are unaffected by acoustic noise
- provide real-time information
- EM radar-like sensors are important because they...

- They reflect from all of the speech articulator tissues:
The Not So Good News:

- EM waves reflect, refract, scatter, and are partly absorbed by all dielectric and conductivity interfaces in their path.
- Interferometric sensor signals may be ambiguous regarding the origins of their reflected signals due to longitudinal location ambiguity, and the product of Target Area x Movement can be ambiguous longitudinally.

Transverse resolution is typically 2-3 cm, but longitudinal resolution is a few microns.
III. Good News: EM sensors robustly sense Voiced speech signals

- Direct measurements of vocal fold movements, soft palate, etc. are being quantified
- Macro-movements of jaw, tongue, lips, etc.
- Pressure induced trachea wall and surface motions
- Pressure induced vocal resonator wall and surface motions

- E.g., 5-10 micron movements
  - e.g., cheek, tongue, lips, pharynx, sinus surfaces
- 10-20 micron movements
  - Anterior wall
- Pressure induced trachea wall movements (primarily the
  without contact: > 1 mm to 1 cm movements
- Direct measurements of vocal fold movements with
What are the Sources of Radar-like Sensor Signals?
Subject having a "stoma" in the neck are compared

Laser doppler and EM sensor signals from a

Inner Cheek Wall

Laser Doppler from

4 cm below vocal folds

Gems on sub-glottal region

Rear Trachea Wall

Pressure Sensor

Velocimeter Beams

Laser Doppler from

4 cm location

EM Sensor

4 cm location

4 cm location

4 cm location

3 cm location

Stoma

V. Folds

Microphone

Acoustic

Gems on cheek

Gems oncheek

Gems on sub-glottal region
What are the Sources of Radar-like Sensor Signals?
An EM sensor signal reflected from the glottal area, is correlated to an EM sensor signal using high-speed video (3 per ms, 0.03 ms exposure).
Subject was treated for laryngeal paresis following thyroplasty with silicon prosthesis implant, and implant of #6 Montgomery speaking valve.

The EM sensor laryngeal-prominence signal is from vocal fold opening and closing, typ. 1 volt signals.

Typical EM sensor signals:
- Laser velocity
- Sub-glottal pressure
- Laser position
- IEEG vocal fold contact
- Acoustics microphone

The EM sensor laryngeal-prominence signal is from vocal fold opening and closing, typ. 1 volt signals.
The neck model is an infinite dielectric \( \varepsilon = 25 \), with an internal "soup can" air space divided by a 4 mm membrane.

Incoming Plane Wave 2.3 GHz

Example of parabolic slot in membrane

Circular slot in membrane with Vocal Fold

Opening of Vocal Fold

Typical observation point of reflected waves

Simulation mesh of trachea and vocal fold membrane

4 cm

1.5 cm

1.0 cm
Extensive EM simulations were employed to obtain amplitude and phase of Reflecting EM waves.
Simulated EM wave wave reflections from the glottal opening show origin of strong “glottal” signals.

Sagittal slice through 1.5 cm dia air tube, with 4mm vocal fold membrane & circular hole.

A) open tube, wave reflects from first surface.

B) fold adducting, carried wave into tube.

C) nominal opening in folds, showing reflection locations.

D) closed glottis shows little reflection, wave passes through.
From the simulations we obtain reflected EM wave amplitudes and $\phi$s for an open tube, a solid membrane, and slot openings in the membrane. The graph illustrates these findings, comparing different conditions such as open and solid membranes with specific AR values. Measurement locations are marked at 1.75 cm. The diagram also highlights the location of the trachea and larynx in the context of these measurements.
Both phase and amplitude contribute to the EM sensor signal, $A \cos \phi$, between configurations. A phase and amplitude jump occurs upon initial fold separation.
An air flow function can be generated by associating the EM sensor signal with glottal area (in progress).
What are the Sources of Radar-like Sensor Signals?
Vocal tract wall experiments, e.g., trachea and cheek, show 5-10 micron motions versus 5-10 cm H²O pressure cycles.

- Laser Doppler from Inner Cheek Wall, 15 micron movement due to 7 cm H²O pressure.
- Laser Doppler from Rear Trachea Wall, 14 micron movement due to 7 cm H²O pressure.
- Laser Doppler from sub-glottal region (4 cm below vocal folds), 15 microns movement due to 5 cm H²O pressure.

Gems on sub-glottal region 4 cm below vocal folds, 15 microns movement of anterior sub-glottal wall.

Employed an EM sensor, laser, pressure sensor, EGG, and microphone.
EM sensor signal amplitude and shape can be estimated by simulating changes in position of posterior and anterior surfaces of the tracheal tube.
EM sensor measured trachea wall motion, 4 cm down from larynx, giving 40 mV signals. EM sensed subglottal trachea wall signals show an anterior wall "ballooning" versus pressure.

Onset of Fold closure:

EM sensor measured rear tracheal wall position versus time, from laser doppler velocity signal.

Sub-glottal movement:

Vocal fold signals, 1 V.

Closest of Fold.
What are the Sources of Radar-like Sensor Signals?
Conclusion: Low power, interferometric EM wave sensor signals are understood and very useful for real-time speech processing.

- We can measure Vocal-Tract tissue interface motions as small as $\Delta x = 1\text{ µm}$ using $< 0.5\text{ mW}$ interferometric sensors.
- Two types of Glottal related signals (100-200Hz) are obtained from the neck and head region:
  - Air pressure induced vocal tract wall movements, $Ax = 5-15\text{ µm}$
  - Direct measurement of vocal fold cycle, $Ax = 0.1-1\text{ cm}$

Applications in human speech characteristic for many low bandwidth vocal coding, $> 300\text{ Hz}$, including:

- Low Bandwidth Vocoding, $< 300\text{ Hz}$
- Denoising
- Speaker Verification
- Speech Recognition
- Speech Recognition
The human vocal tract can be well characterized if a sufficiently good excitation function can be obtained.
glottis and jaw

2 EM sensor measurements for "butter"
Glottal-synchronous transfer functions for "butter" using EM sensor excitation & ARMA
Additional information is available at the following:

• See web sites for additional information:
  - http://speech.llnl.gov (many reports and papers)
• Recent working draft, available as preprint, describes details of EM wave interaction with glottal and tracheal structures (in review for publication). Titled: "EM Wave Measurements of Glottal Structure Dynamics" Holzrichter, Ng, Burke, Champaign, Kallmann, Sharpe -- LLNL Kobler, Rosowski, and Hillman -- Mass. Eye and Ear Infirmary available as Livermore Report number: UCRL-JC-14775 and above llnl.gov web site.
Major collaborators have been:

- Prof. Robert Hillman - Harvard and Mass. Eye & Ear Infirmary
- Profs. Neville Luhman and Richard Freeman – UC Davis
- Drs. Greg Burnett and Todd Gable – Former students: Lawrence Livermore and UC Davis/LLNL campus
- Prof. John Ohala – UCB
- Dr. Rebecca Leonard – UCD Hospital
- Prof. Ingo Titze – U of Iowa
- Dr. Wayne Lea – Speech Sciences Institute
- Prot. Luhman and Richard Freeman – UC Davis