Forensic voice comparison using voice source features: An evaluation of the GLOTTEX® software package

Ewald Enzinger, Cuiling Zhang, Geoffrey Stewart Morrison

FORENSIC VOICE COMPARISON LABORATORY
SCHOOL OF ELECTRICAL ENGINEERING & TELECOMMUNICATIONS
UNSW
THE UNIVERSITY OF NEW SOUTH WALES
SYDNEY • AUSTRALIA

IAFPA 2012
Santander

Acknowledgement of Funding

- Data collection was funded by an International Association of Forensic Phonetics and Acoustics (IAFPA) Research Grant.

Why voice source features for FVC?

- Under-investigated source of features
- Potentially complementary information
  - Most acoustic FVC systems based on spectral envelope (MFCC, formants)
- Closely related to speakers' vocal fold biomechanics
  - laryngeal settings (creaky voice etc.)
  - voice pathologies

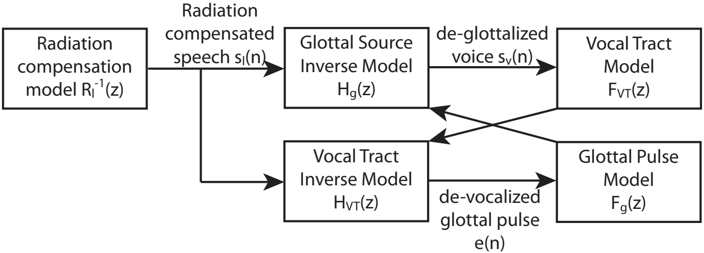
➡ Potentially high between-speaker variability

Voice source features – GLOTTEX®

- Evaluation of the GLOTTEX® software package
  - Originally developed as a non-invasive diagnostic tool for medical applications
  - Provides features related to biomechanics of the vocal folds
- Gómez-Vilda et al. (2008) proposed use in forensic voice comparison
  - Voice source features applied to speaker verification (Mazaira-Fernández et al., 2010)
- Automatic feature extraction from speech segments

Voice source features – GLOTTEX®

Separation of glottal source and vocal tract

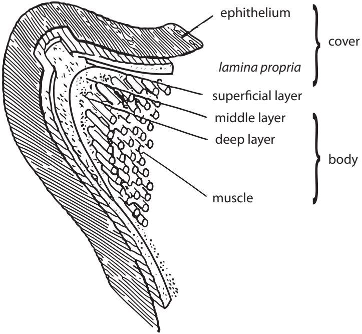


Radiation compensation model $R_l^{-1}(z)$
Radiation compensated speech $s_l(n)$
Glottal Source Inverse Model $H_g(z)$
de-glottalized voice $s_v(n)$
Vocal Tract Model $F_{VT}(z)$
Vocal Tract Inverse Model $H_{VT}(z)$
de-vocalized glottal pulse e(n)
Glottal Pulse Model $F_g(z)$

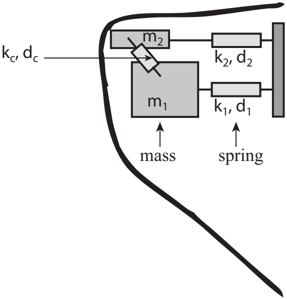
Gómez-Vilda et al. (2009)

Voice source features – GLOTTEX® MWC

- Modeling of vocal fold dynamics using k-mass model
- Decomposition of glottal source signal into
  - average acoustic wave (AAW)
  - mucosal wave correlate (MWC)

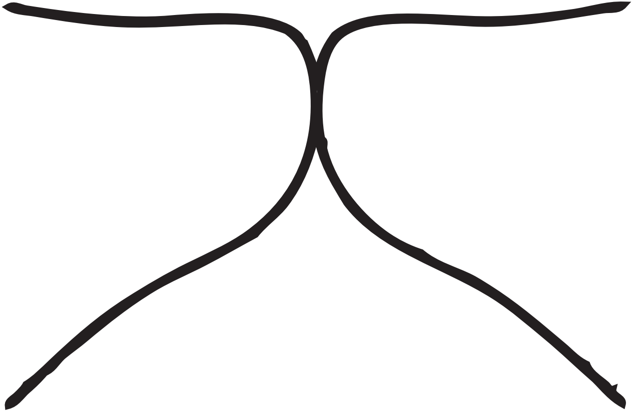


Hirose (2010), p. 140



Two-mass vocal fold model (Story, 2002)

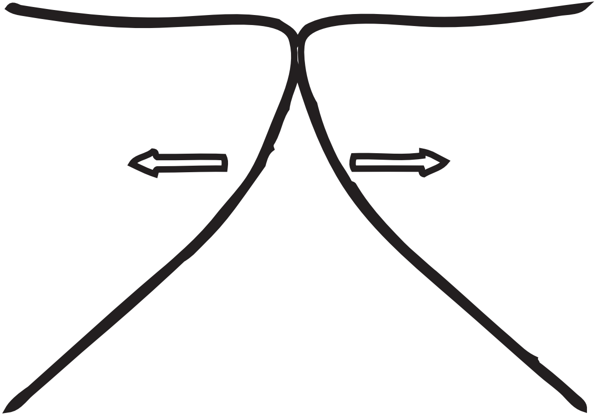
Idealized cycle of vocal fold vibration (Story, 2002, p.197)



(1)

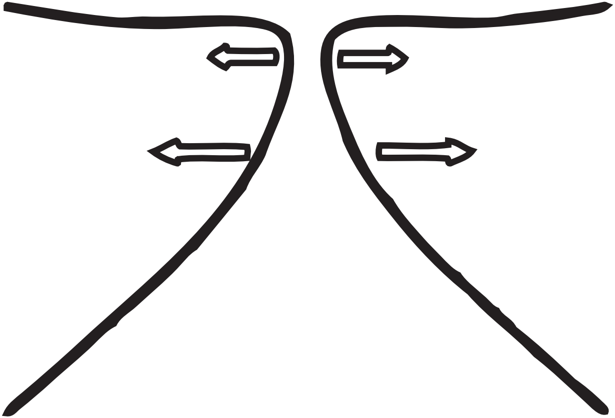
Voice source features – GLOTTEX® MWC

Idealized cycle of vocal fold vibration (Story, 2002, p.197)



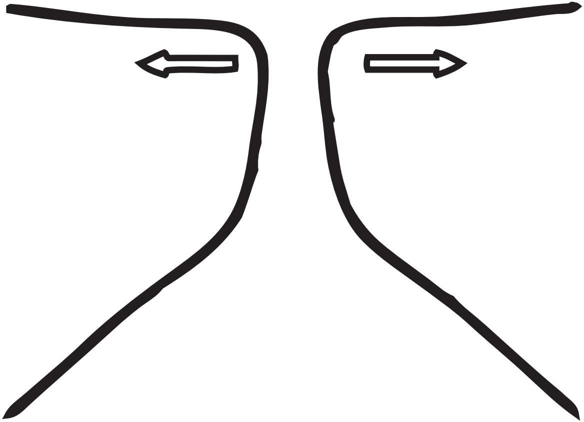
(2)

Idealized cycle of vocal fold vibration (Story, 2002, p.197)



(3)

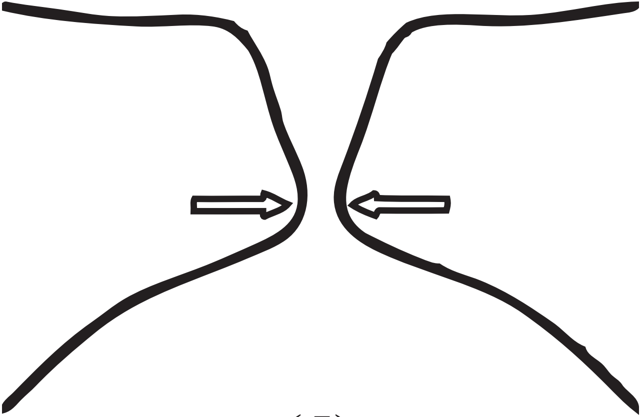
Idealized cycle of vocal fold vibration (Story, 2002, p.197)



(4)

Voice source features – GLOTTEX® MWC

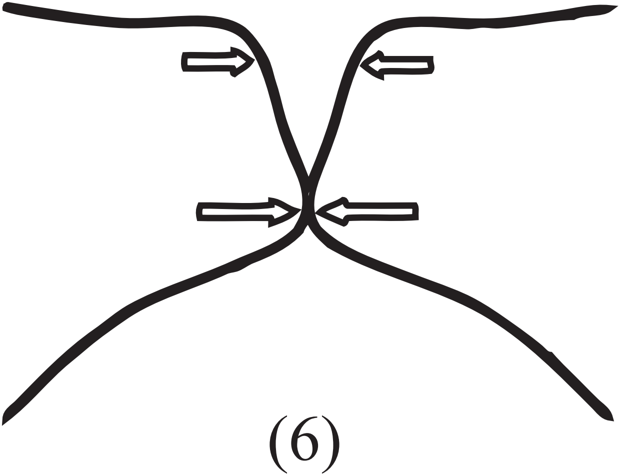
Idealized cycle of vocal fold vibration (Story, 2002, p.197)



(5)

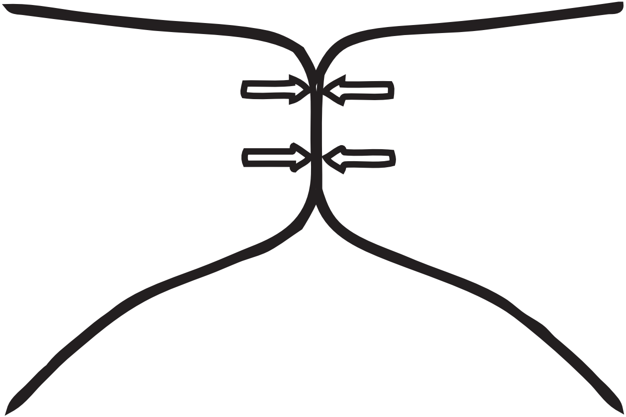
Voice source features – GLOTTEX® MWC

Idealized cycle of vocal fold vibration (Story, 2002, p.197)



Voice source features – GLOTTEX® MWC

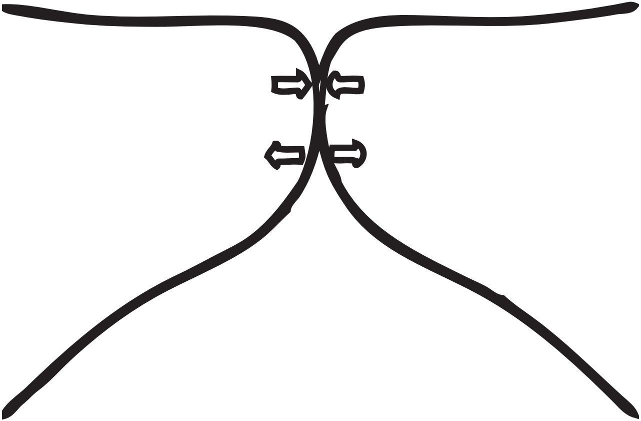
Idealized cycle of vocal fold vibration (Story, 2002, p.197)



(7)

Voice source features – GLOTTEX® MWC

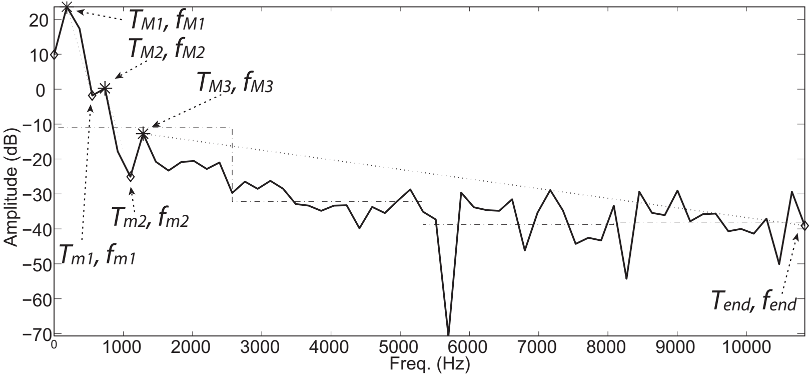
Idealized cycle of vocal fold vibration (Story, 2002, p.197)



(8)

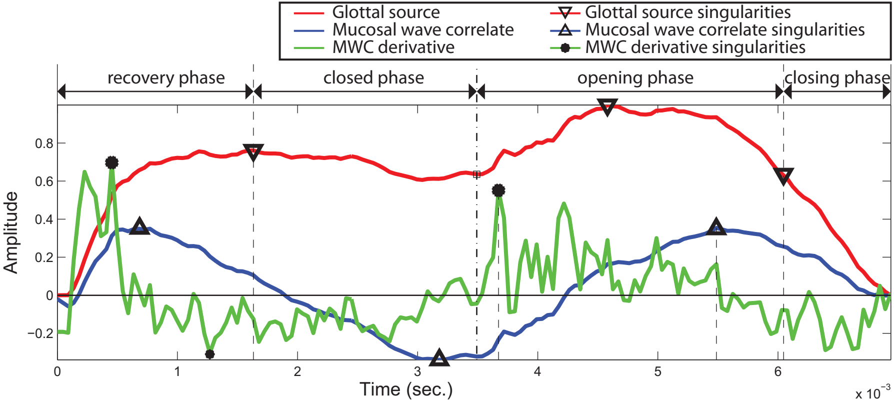
Voice source features – GLOTTEX®

- Mucosal wave correlate (MWC) features:
  - 1-14th cepstral coefficient of MWC power spectrum
  - MWC power spectral singularities (frequencies and amplitudes of minima and maxima):



Voice source features – GLOTTEX® Time-domain

- Relative times of singularities in glottal source, MWC, and MWC derivative time signals



Voice source features – GLOTTEX® f0+Distortion

- Distortion features and fundamental frequency
  - Fundamental frequency (f0)
  - f0 jitter
  - Amplitude shimmer
  - Slenderness shimmer (glottal pulse spike $\frac{height}{width}$)
  - Area shimmer (area under source signal per cycle)
  - Glottal-to-noise excitation ratio

Data

- 60 female Standard Chinese speakers
- Split into 3 groups of 20 speakers
  - background database
  - development set
  - test set
- Information-exchange task over the telephone
- High quality and mobile-to-landline data
- Two recording sessions separated by 2–3 weeks

`http://databases.forensic-voice-comparison.net/`

Likelihood ratio calculation

- Score obtained using GMM–UBM system
- Logistic-regression calibration and fusion
- 2 approaches:
  - segmental: pause fillers (sustained /n/ tokens)
  - speech-active portion of recording (200 ms blocks)

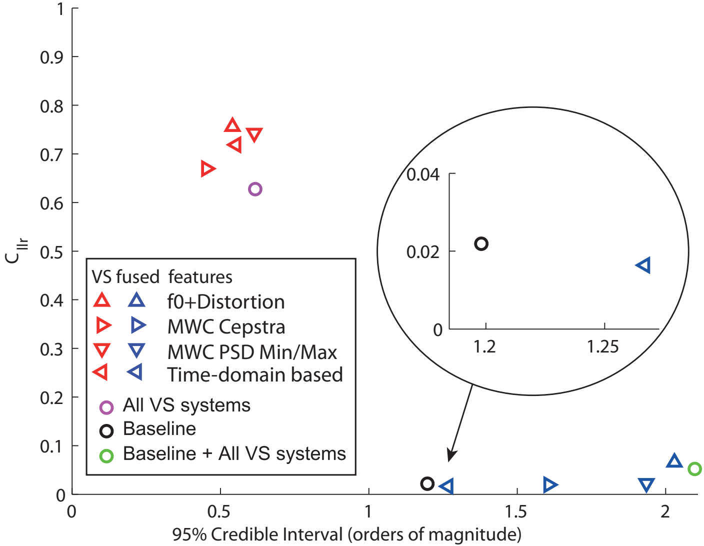
- Baseline MFCC GMM-UBM system (Reynolds et al., 2000)
  - Entire speech-active portion of recording
  - 16 MFCC+Δ, Feature warping
  - 1024 Gaussian mixture components (UBM)

Evaluation measures

- Validity / Accuracy
  - Log-likelihood ratio cost (C_{llr}) metric
- Reliability / Precision
  - 95% credible interval (Morrison, 2011)
  - Parametric estimation method

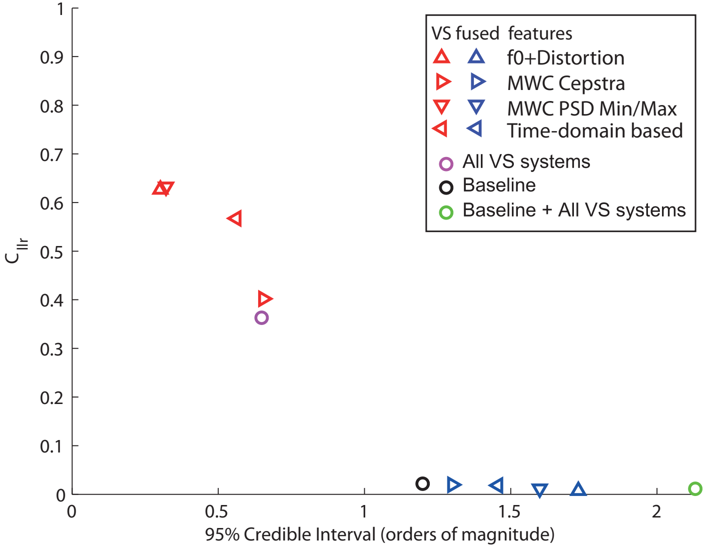
Results - segmental /n/ approach

high-quality v high-quality



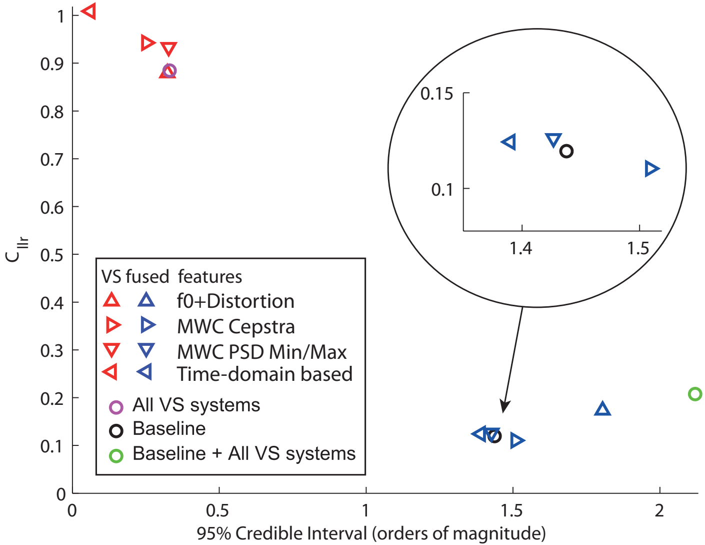
Results - speech-active portion

high-quality v high-quality



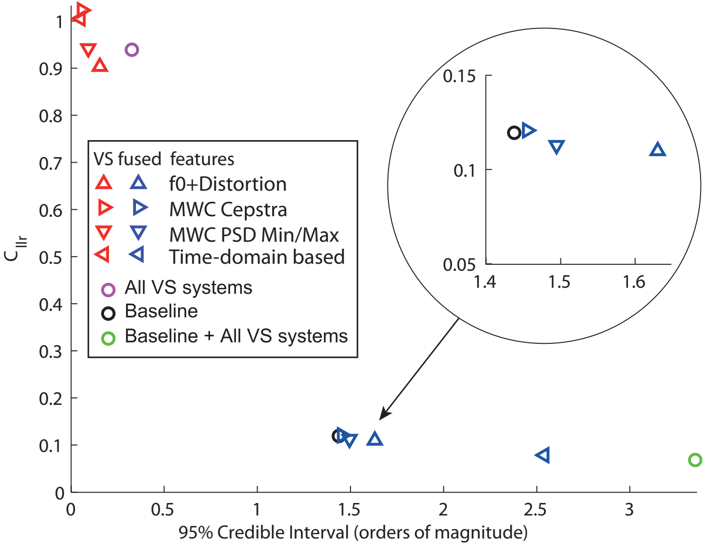
Results - segmental /n/ approach

mobile-to-landline v high-quality



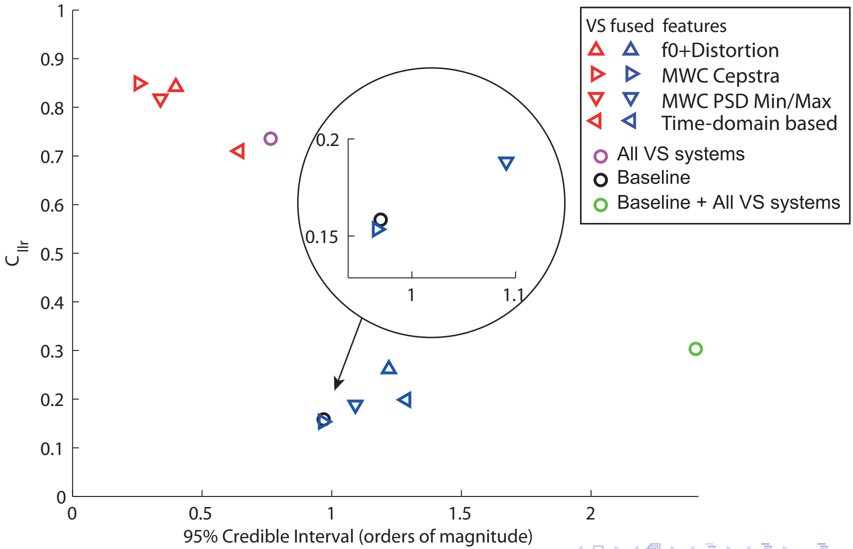
Results - speech-active portion

mobile-to-landline v high-quality



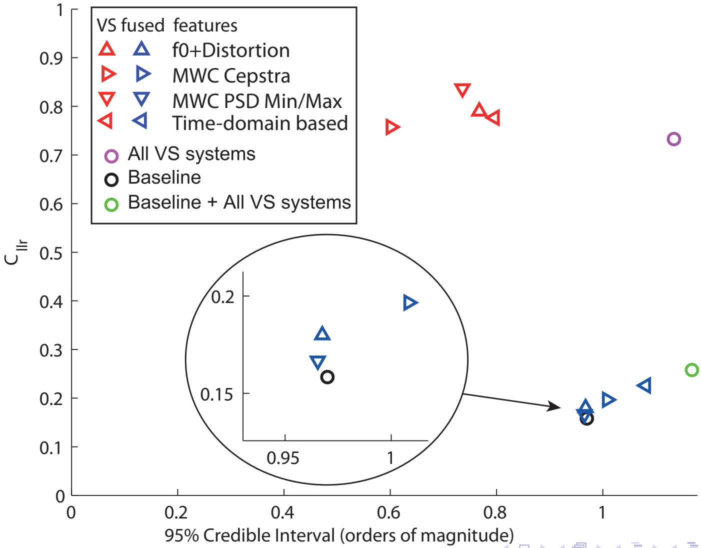
Results - segmental /n/ approach

mobile-to-landline v mobile-to-landline



Results - speech-active portion

mobile-to-landline v mobile-to-landline



Conclusion

- Caveats
  - only one database in one language tested
  - segment tested was a nasal
    (developer said ok)
    (all-speech-active also tested)

- Adding glottal-source features extracted by GLOTTEX® did **not** lead to substantial improvement in performance relative to a baseline forensic-voice-comparison system.

Thanks!

References

Gómez-Vilda, P., Álvarez, A., Mazaira, L., Fernández-Baillo, R., Nieto, V., Martínez, R., Muñoz, C., & Rodellar, V. (2008). Decoupling vocal tract from glottal source estimates in speaker's identification. *Language Design (Special Issue)*, (pp. 111–118).

Gómez-Vilda, P., Fernández-Baillo, R., Rodellar, V., Nieto, V., Álvarez, A., Mazaira-Fernándeza, L., Martínez, R., & Godino-Llorenteb, J. (2009). Glottal source biometrical signature for voice pathology detection. *Speech Communication*, *51*(9), 759–781.

Hirose, H. (2010). Investigating the physiology of laryngeal structures. In W. Hardcastle, & J. Laver (Eds.) *The Handbook of Phonetic Sciences*, (pp. 130–152). Oxford: Blackwell.

Mazaira-Fernández, L., Álvarez-Marquina, A., Gómez-Vilda, P., Martínez-Olalla, R., & Muñoz, C. (2010). Glottal Source Cepstrum Coefficients Applied to NIST SRE 2010. In *V Jornadas de Reconocimiento Biométrico de Personas (JRBP10)*. Huesca, Spain: Facultad de Informática (UPM).

Morrison, G. S. (2011). Measuring the validity and reliability of forensic likelihood-ratio systems. *Science & Justice*, *51*, 91–98.

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted gaussian mixture models. *Digital Signal Process.*, *10*, 19–41.

Story, B. (2002). An overview of the physiology, physics and modeling of the sound source for vowels. *Acoustical Science and Technology*, *23*(4), 195–206.