On smoothing articulatory trajectories obtained from Gaussian mixture model based acoustic-to-articulatory inversion

Prasanta K. Ghosh[a)]
Electrical Engineering, Indian Institute of Science (IISc), Bangalore,
Karnataka 560012, India
prasantg@ee.iisc.ernet.in

Shrikanth S. Narayanan
Department of Electrical Engineering, University of Southern California,
Los Angeles, California 90089
shri@sipi.usc.edu

Abstract: It is well-known that the performance of acoustic-to-articulatory inversion improves by smoothing the articulatory trajectories estimated using Gaussian mixture model (GMM) mapping (denoted by GMM + Smoothing). GMM + Smoothing also provides similar performance with GMM mapping using dynamic features, which integrates smoothing directly in the mapping criterion. Due to the separation between smoothing and mapping, what objective criterion GMM + Smoothing optimizes remains unclear. In this work a new integrated smoothness criterion, the smoothed-GMM (SGMM), is proposed. GMM + Smoothing is shown, both analytically and experimentally, to be identical to the asymptotic solution of SGMM suggesting GMM + Smoothing to be a near optimal solution of SGMM.

© 2013 Acoustical Society of America

PACS numbers: 43.70.Jt, 43.72.Ar, 43.72.Ct [AL]
Date Received: April 19, 2013 **Date Accepted:** June 27, 2013

1. Introduction

The vocal articulators such as jaw, lips, tongue, and velum (VEL) move in a coordinated fashion when a person speaks. The articulators, however, move at a slower rate compared to the vocal tract resonance frequencies. It is sufficient to sample articulatory movements at 200 Hz to capture detailed dynamics of the critical speech articulators[1] (compared to sampling the speech signal which requires a rate of 7.6 kHz even for telephone quality). The slow rate of articulatory movement causes the articulatory trajectories to be smooth and low-pass in nature; this is borne out in measurements obtained using techniques such as Electromagnetic Articulography (EMA)[2] and UltraSound.[3] This fact is exploited in speech modeling, including notably in acoustic-to-articulatory (AtoA) inversion that attempts to recover articulatory details from the speech signal. In particular, inversion performance has been demonstrated to improve by smoothing the estimated articulatory features either in a post-processing step[4] or by incorporating smoothness directly in the estimation criterion.[4,5]

There are several AtoA inversion techniques, and Toutios and Margaritis[6] provide a comprehensive summary of them. Among these techniques, we focus on the AtoA inversion based on Gaussian mixture model (GMM) mapping.[4] In GMM based mapping, articulatory features are estimated separately in each analysis frame from acoustic features using the minimum mean squared error (MMSE) criterion. This

[a)] Author to whom correspondence should be addressed.

causes the estimated articulatory feature trajectory to be rough and jagged in nature. To obtain a realistic articulatory trajectory from the GMM based estimate, the estimated trajectory is low-pass filtered where the cutoff frequency of the low-pass filter is selected to achieve the best inversion performance.[4] Toda *et al.*[4] also proposed a GMM mapping using dynamic features under the maximum-likelihood criterion (thus integrating smoothing directly in the mapping process), which was found to yield a similar performance with GMM mapping followed by low-pass filtering (denoted by GMM + Smoothing). It is important to note that in GMM + Smoothing, the smoothed articulatory features no longer remain optimal in MMSE sense. This suggests that a criterion different from MMSE could correspond to the estimates obtained using GMM + Smoothing which, in fact, yields a better inversion performance than the MMSE criterion and a similar performance with dynamic feature based GMM mapping.

We propose a new smoothness criterion for inversion, called smoothed-GMM (SGMM), which combines smoothness with the information from GMM mapping within the same optimization framework rather than performing them separately. GMM mapping is shown to be a special case of SGMM. It is also analytically shown that GMM + Smoothing matches the solution of SGMM in the limit when the length of the test utterance becomes large. Experimental results on an articulatory database reveal that in practice this asymptotic limit is achieved even for an average utterance length of ~2.75 s with a frame rate of 100 Hz. Thus, both theoretically and experimentally, GMM + Smoothing turns out to be a near optimal solution of SGMM.

2. The SGMM criterion

Suppose the acoustic and articulatory feature vectors at the *n*th frame are denoted by $\mathbf{x}_n$ and $\mathbf{y}_n$, respectively. $\mathrm{x}_n = [x_n^1\ x_n^2\ \cdots\ x_n^I]^{\mathrm{T}}$ and $\mathrm{y}_n = [y_n^1\ y_n^2\ \cdots y_n^J]^{\mathrm{T}}$, where x_n^i is the *i*th acoustic feature $(1 \le i \le I)$ and y_n^j is the *j*th articulatory feature $(1 \le j \le J)$. T denotes the transpose operator. In GMM based AtoA inversion, a GMM is used to model the joint probability $p(\mathrm{x}_n,\ \mathrm{y}_n|\Theta)$ given by $p(\mathrm{x}_n,\ \mathrm{y}_n|\Theta) = \Sigma_{i=1}^{M} w_i \mathcal{N}(\mathrm{x}_n,\ \mathrm{y}_n;\ \mu_i,\ \Sigma_i)$. Θ are the GMM parameters $\{w_i,\ \mu_i,\ \Sigma_i\}_{i=1}^{M}$, where M is the number of mixture components. w_i denotes the mixture weight for the *i*th mixture. $\mu_i = [\mu_i^{(x)\mathrm{T}}\ \mu_i^{(y)\mathrm{T}}]^{\mathrm{T}}$ is the mean vector of the *i*th mixture and $\mu_i^{(x)}$ and $\mu_i^{(y)}$ denote the mean vectors of the *i*th mixture for x_n and y_n, respectively. Similarly Σ_i denotes the full covariance matrix of *i*th mixture, which is given by

$$\Sigma_i = \begin{bmatrix} \Sigma_i^{(xx)} & \Sigma_i^{(xy)} \\ \Sigma_i^{(yx)} & \Sigma_i^{(yy)} \end{bmatrix},$$

where $\Sigma_i^{(xx)}$ and $\Sigma_i^{(yy)}$ denote the covariance matrices of the *i*th mixture for $\mathbf{x}_n$ and $\mathbf{y}_n$, respectively, and $\Sigma_i^{(xy)}$, $\Sigma_i^{(yx)}$ represent the cross-covariance matrices of the *i*th mixture.

Given an acoustic feature vector sequence of length N frames, x_n, $1 \le n \le N$, the goal of AtoA inversion is to estimate the corresponding articulatory feature vector sequence, $\hat{\mathrm{y}}_n$, $1 \le n \le N$.

In GMM-based inversion,[4] $\hat{\mathrm{y}}_n$ is defined as the MMSE estimate given x_n:[7]

$$\hat{\mathrm{y}}_n \triangleq \mathbb{E}(\mathrm{y}_n|\mathrm{x}_n) = \sum_{i=1}^{M} p(m_i|\mathrm{x}_n,\ \Theta)\, \mathbb{E}(\mathrm{y}_n|\mathrm{x}_n,\ m_i,\ \Theta), \tag{1}$$

where $p(m_i|\mathrm{x}_n, \Theta) = w_i \mathcal{N}\left(\mathrm{x}_n; \mu_i^{(x)},\ \Sigma_i^{(xx)}\right) / \Sigma_{j=1}^{M} w_j \mathcal{N}\left(\mathrm{x}_n;\ \mu_j^{(x)},\ \Sigma_j^{(xx)}\right)$ and $\mathbb{E}(\mathrm{y}_n|\mathrm{x}_n, m_i,\ \Theta) = \mu_i^{(y)} + \Sigma_i^{(yx)} \Sigma_i^{(xx)-1}\left(\mathrm{x}_n - \mu_i^{(x)}\right)$. Note that $\Sigma_i p(m_i|\mathbf{x}_n,\ \Theta) = 1$.

Toda *et al.*[4] reported that smoothing $\hat{\mathbf{y}}_n$ by low-pass filtering makes the estimated articulatory trajectory more realistic and improves the inversion performance.

Here we propose a new smoothness criterion, called SGMM which constrains the estimated trajectory to be smooth to a required degree while estimating the trajectory from the GMM mapping information. Thus, instead of estimating articulatory features in each frame independently (as done in GMM-based inversion), the SGMM criterion estimates the articulatory trajectory for an entire utterance. The jth articulatory feature trajectory is estimated by solving the following optimization problem:

$$\begin{aligned}\{\hat{y}^j_n;\ 1\le n\le N\} &= \underset{\{z^j_n\}}{\operatorname{argmin}}\ J(\{z^j_n;\ 1\le n\le N\})\\ &= \underset{\{z^j_n\}}{\operatorname{argmin}}\ C^j\sum_n\sum_i p(m_i|\mathrm{x}_n,\ \Theta)(z^j_n-\mathbb{E}(y^j_n|\mathrm{x}_n,\ m_i,\ \Theta))^2\\ &\quad+(1-C^j)\sum_n\left(\sum_k z^j_k h^j_{n-k}\right)^2. \end{aligned}\tag{2}$$

J is the objective function comprised of a convex combination of two terms with the convex weight $C^j(0\le C^j\le 1)$. z^j_n is the optimization variable. The second term in J is the total energy of the output of a high-pass filter with impulse response h^j_n (corresponding to the jth articulator) with z^j_n as input. By minimizing the output of a high-pass filter, SGMM constrains the solution to be low-pass or smoothly varying in nature. h^j_n could be designed based on the degree of required smoothness for jth articulator trajectory.

The first term in J is designed so that it utilizes the mapping between acoustic and articulatory spaces using the conditional means with their weights derived from the GMM. The choice of C^j provides a trade-off between the GMM mapping and the smoothness factor. The optimization in Eq. (2) is solved for $j=1,\ \ldots,\ J$ separately to obtain the estimates of all J articulatory feature trajectories.

3. Solution SGMM criterion based optimization

The objective function J in Eq. (2) is a convex (and quadratic) function of the optimization variables $\{z^j_n;\ 1\le n\le N\}$. Thus a global minimum is guaranteed. We define the autocorrelation sequence of the high-pass filter h^j_n as $R^j_{l-k}\triangleq\Sigma_n h^j_{n-k}h^j_{n-l}$. For minimization, the partial derivatives of J with respect to z^j_n are set to zero at $z^j_n=\hat{y}^j_n$ to obtain a set of N equations in the following matrix vector form:

$$\begin{pmatrix}(1-C^j)R^j_0+C^j & (1-C^j)R^j_1 & \cdots & (1-C^j)R^j_{N-1}\\ (1-C^j)R^j_{-1} & (1-C^j)R^j_0+C^j & \cdots & (1-C^j)R^j_{N-2}\\ \vdots & \vdots & \vdots & \vdots\\ (1-C^j)R^j_{-(N-1)} & (1-C^j)R^j_{-(N-2)} & \cdots & (1-C^j)R^j_0+C^j\end{pmatrix}\begin{pmatrix}\hat{y}^j_1\\ \hat{y}^j_2\\ \vdots\\ \hat{y}^j_N\end{pmatrix}=\begin{pmatrix}C^j\Delta^j_1\\ C^j\Delta^j_2\\ \vdots\\ C^j\Delta^j_N\end{pmatrix},\tag{3}$$

where $\Delta^j_l=\Sigma_i p(m_i|\mathrm{x}_l,\ \Theta)\ \mathbb{E}(\mathrm{y}^j_l|\mathrm{x}_l,\ m_i,\ \Theta)$. We can further write the set of equations as $((1-C^j)\mathbf{R}^j+C^j\mathbf{I})\hat{\mathbf{y}}^j=C^j\mathbf{d}^j$, where $\mathbf{R}^j=\{R^j_{kl}\}=\{R^j_{k-l}\}=\{R^j_{|k-l|}\}$ (since the autocorrelation matrix is symmetric), $\mathbf{I}$ is $N\times N$ identity matrix, $\hat{\mathrm{y}}^j=[\hat{y}^j_1,\cdots\hat{y}^j_N]^{\mathrm{T}}$ and $\mathbf{d}^j=[\Delta^j_1,\ \cdots,\Delta^j_N]^{\mathrm{T}}$. $\mathbf{R}^j$ is an autocorrelation matrix and hence symmetric toeplitz. Thus, $((1-C^j)\,\mathbf{R}^j+C^j\mathbf{I})$ is invertible for any choice of $C^j(0<C^j<1)$. The estimate of the jth articulatory feature trajectory thus can be obtained as follows:

$$\hat{\mathbf{y}}^j=C^j((1-C^j)\mathbf{R}^j+C^j\mathbf{I})^{-1}\mathrm{d}^j=\left(\mathbf{I}+\frac{1-C^j}{C^j}\mathbf{R}^j\right)^{-1}\mathbf{d}^j.\tag{4}$$

When $C^j=0$ in Eq. (4) (i.e., only the second term in J is considered), the estimated trajectory $\hat{\mathrm{y}}^j$ is trivially zero. In other words, when no GMM mapping information is included in the objective function J, the maximally smooth solution is an all zero trajectory. On the other hand when $C^j = 1$ (i.e., no smoothness constraint is imposed on the estimated articulatory trajectory), $\hat{\mathbf{y}}^j = \mathbf{d}^j$ or $\hat{y}^j_n = \Delta^j_n = \Sigma_i p(m_i|\mathbf{x}_n, \Theta)$ $\mathbb{E}(y^j_n|\mathbf{x}_n, m_i, \Theta)$, which is identical to the GMM mapping based estimate [Eq. (1)]. Thus, GMM based inversion is a special case of the optimization using the proposed SGMM criterion. For $0 < C^j < 1$, the estimated trajectory lies between the extremes of the all-zero trajectory and the jagged trajectory obtained using GMM based inversion.

$\mathbf{R}^j$ is, in general, an $N \times N$ positive semi-definite symmetric toeplitz matrix with its entries coming from the autocorrelation sequence of h^j_n (i.e., R^j_n) with the corresponding spectrum $|H^j(\omega)|^2$. $\mathbf{R}^j$ is also a convolution matrix with the corresponding impulse response R^j_n. Let $\rho^j = (1 - C^j)/C^j$. Hence, $\mathbf{I} + \rho^j\mathbf{R}^j$ is an $N \times N$ positive definite symmetric toeplitz matrix with the related spectrum $1 + \rho^j|H^j(\omega)|^2$. Note that $1 + \rho^j|H^j(\omega)|^2 > 0, \forall\omega$; the addition of "1" acts as a regularization ensuring the invertibility of the spectrum $1 + \rho^j|H^j(\omega)|^2$ (similar to $\mathbf{I}$ for the invertibility of $\mathbf{I} + \rho^j\mathbf{R}^j$). Using a result from the inverse of the toeplitz matrix [Eq. (5.5) in Ref. 8], it is easy to show that $(\mathbf{I} + \rho^j\mathbf{R}^j)^{-1}$ is asymptotically (as $N \to \infty$) toeplitz with the corresponding spectrum $|G^j(\omega)|^2 = 1/(1 + \rho^j|H^j(\omega)|^2)$. Since $|H^j(\omega)|^2$ is a high-pass spectrum and $\rho^j > 0$, it is easy to see that $|G^j(\omega)|^2$ is a low-pass spectrum, where ρ^j controls the stop band attenuation of the low-pass filter. Hence, in the limit $N \to \infty$, $(\mathbf{I} + \rho^j\mathbf{R}^j)^{-1}$ acts as a convolution matrix with a corresponding impulse response Q^j_n of a low-pass filter with spectrum $|G^j(\omega)|^2$, where Q^j_n is the inverse Fourier transform of $|G^j(\omega)|^2$. Thus, asymptotically $\hat{\mathbf{y}}^j$ [Eq. (4)] is a low-passed or smoothed version of $\mathbf{d}^j$, the GMM based estimate. Thus we prove that the solution of SGMM asymptotically matches GMM + Smoothing.

For illustration, we consider a fifth order rational transfer function $(H(\omega))$ of a type II Chebyshev high-pass filter with a 40 dB attenuation at 10 Hz with sampling frequency 100 Hz as shown in Fig. 1. For finite N, we pick the N/2th (N/2+1th for even N) row of $(\mathbf{I} + \rho\mathbf{R})^{-1}$ as the representative impulse response P_n for $(\mathbf{I} + \rho\mathbf{R})^{-1}$. We compute the mean squared error (MSE) E_{P-Q} between P_n and Q_n over the same support as shown in Fig. 1(d) for different values of ρ. It is clear that E_{P-Q} becomes zero for $N = 150$ (corresponds to 1.5 s with 100 Hz frame rate) for $\rho = 999$. For $\rho = 99$ and 9, E_{P-Q} becomes zeros even for lower values of N indicating the asymptotic equivalence between $(\mathbf{I} + \rho\mathbf{R})^{-1}$ and $|G(\omega)|^2$.

4. Experimental evaluation

While we argue in Sec. 3 that the solution of SGMM asymptotically matches GMM + Smoothing, it is important to note that the low-pass filter in the limit has a frequency response of the form $(1 + \rho|H(\omega)|^2)^{-1}$ [i.e., infinite impulse response (IIR) filter] in the case of SGMM, but in the case of GMM + Smoothing the low-pass filter can be either finite impulse response or IIR and its frequency response need not have a specific form. We conduct AtoA inversion experiments on an articulatory dataset comprising utterances of different lengths to examine the role that the particular form of a

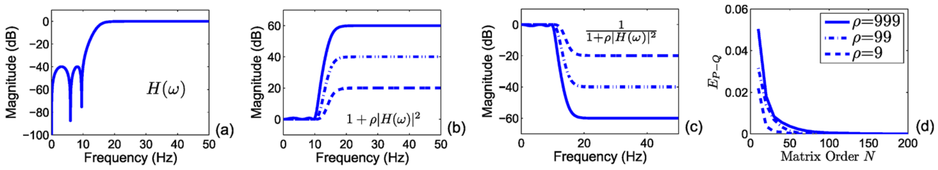


Fig. 1. (Color online) Illustration of the asymptotic equivalence between $(\mathbf{I} + \rho\mathbf{R})^{-1}$ and a low-pass convolution matrix with spectrum $|G(\omega)|^2 = (1 + \rho|H(\omega)|^2)^{-1}$ (index "j" is omitted for simplicity): (a) A high-pass spectrum $H(\omega)$, (b) $1 + \rho|H(\omega)|^2$, (c) $|G(\omega)|^2$, and (d) MSE between impulse response P_n representing $(\mathbf{I} + \rho\mathbf{R})^{-1}$ and Q_n, the inverse Fourier transform of $|G(\omega)|^2$.

low-pass filter in SGMM may play on the inversion performance, specifically for different choices of N. The dataset and experimental details are described below.

4.1 Dataset and pre-processing

For the AtoA experiment, we use the Multichannel Articulatory (MOCHA) database[9] that contains speech and the corresponding EMA data from one male and one female talker of British English. The EMA data consists of dynamic positions of the EMA sensors in the mid-sagittal plane of the talker. A total of seven sensors are placed on the upper lip (UL), lower lip (LL), lower incisor (JAW), tongue tip (TT), tongue body (TB), tongue dorsum (TD), and VEL. Following the preprocessing steps outlined by Ghosh and Narayanan,[5] we obtain parallel acoustic and articulatory data at a frame rate of 100 observations/s. We use 14 dimensional raw EMA features for representing the articulatory space (i.e., X and Y co-ordinates of 7 EMA sensors), namely ULx, LLx, JAWx, TTx, TBx, TDx, VELx, ULy, LLy, JAWy, TTy, TBy, TDy, and VELy. Acoustic features are represented by 39 dimensional Mel-frequency cepstral coefficients (MFCCs) and are computed using 20 msec analysis frame length with 10 msec shift.

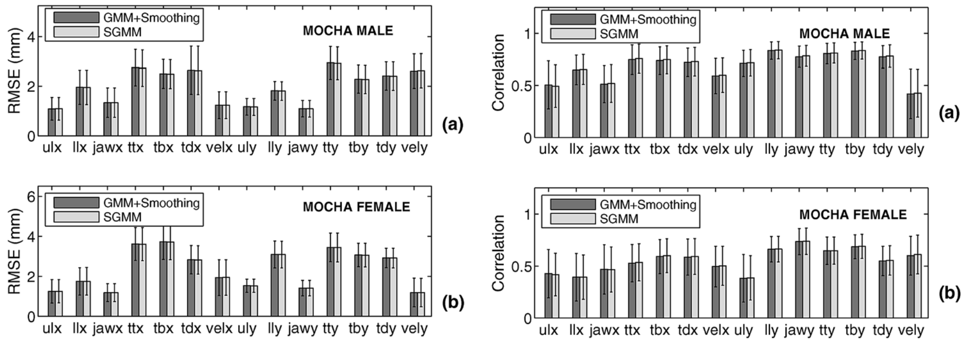
4.2 Experimental setup

AtoA inversion is performed separately on the male and female subjects of the MOCHA corpus using a fivefold cross-validation setup. Inversion performance is measured over all sentences of all folds through average root mean squared error (RMSE) and Pearson correlation coefficient (PCC)[10] between the original and estimated articulatory trajectories.

Following the finding by Toda *et al.*,[4] 64 mixture component GMMs are used to model the acoustic-articulatory map in the training data separately for each fold. In the case of SGMM, we use a fifth order type II Chebyshev high-pass filter with 40 dB stop band attenuation as h_n^j with cut-off frequency f_c^j for the jth articulator. Different values of f_c^j and C^j were experimented with -$f_c^j \in \{3 + 0.5\,(k-1)\,\text{Hz},\ k = 1,\cdots,45\}$ and $C^j \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 0.9, 0.99, 0.999\}$. We report AtoA inversion performance corresponding to the f_c^j and C^j combination which gives the least average RMSE. In the case of GMM + Smoothing, a fifth order type II Chebyshev low-pass filter with 40 dB stop band attenuation is used for smoothing whose cut-off frequency is also varied over the same range as that for h_n^j and the best performance among these is reported for each articulator.

4.3 Results and discussions

Figure 2 shows the AtoA inversion performance in terms of RMSE and PCC for each articulator of both subjects in the MOCHA corpus. It is evident that the inversion



I. inversion performance using RMSE II. inversion performance using PCC

Fig. 2. Comparison of SGMM and GMM + Smoothing - error bars indicate average inversion performance with ± one standard deviation.

performances using SGMM and GMM + Smoothing are not significantly different. Thus, inversion experiments support our theoretical finding that GMM + Smoothing is a near optimal solution of the SGMM criterion. An advantage of using SGMM over GMM + Smoothing is that the solution of SGMM can be computed in a recursive manner.[5] Optimal choices of C^j in the case of the male and female subjects turn out to be in the range of 0.005 to 0.1 suggesting low-pass filters with high stop band attenuation to be preferred in SGMM. If the length of a sentence is too short to satisfy the asymptotic limit, the length of the sentence could be increased by appending it with silence and then considering articulatory features only in the segment of interest. It should also be noted that the functional forms of generalized smoothness criterion (GSC)[5] and SGMM appear to be similar except that in GSC the training data is used in a non-parametric fashion while SGMM uses parameters of a GMM learned from the training data.

Given a high-pass filter $|H(\omega)|^2$ in SGMM, one can always find a low-pass filter $|G(\omega)|^2 = (1+\rho|H(\omega)|^2)^{-1}$ and perform GMM + Smoothing with $|G(\omega)|^2$ to achieve an inversion performance similar to SGMM with $|H(\omega)|^2$. However, the opposite is not true in general. This is because any arbitrary low-pass filter $A(\omega)$ cannot be put in the form $(1+\rho|H(\omega)|^2)^{-1}$. For example, the type II Chebyshev low-pass filter used in AtoA experiments is not in this particular form. In spite of that the AtoA inversion performances using GMM + Smoothing and SGMM turn out to be similar. This suggests that although it could be difficult to find a high-pass filter $H(\omega)$ in SGMM corresponding to an arbitrary low-pass filter $A(\omega)$ in GMM + Smoothing, SGMM with a high-pass filter different from $H(\omega)$ could lead to a similar inversion performance as that of GMM + Smoothing with $A(\omega)$. For a given $A(\omega)$, one could also find a low-pass filter of the form $(1+\rho|H(\omega)|^2)^{-1}$ that best approximates $A(\omega)$ and then SGMM with the corresponding $H(\omega)$ as the high-pass filter will lead to an inversion performance similar to that of GMM + Smoothing with $A(\omega)$.

5. Conclusions

We present a new unified criterion (SGMM) for estimation and smoothing for AtoA inversion; its solution is shown, both theoretically and experimentally, to be identical to the individually optimized GMM + Smoothing based solution in the limiting case. In practice, these results seem to hold for utterances just a few seconds long. Since in GMM + Smoothing based inversion the GMM mapping and smoothing are performed separately, this finding offers an additional insight as to what underlying criterion is being optimized in GMM + Smoothing.

References and links

[1] S. Ouni and Y. Laprie, "Studying pharyngealization using an articulograph," *International Workshop on Pharyngeals and Pharyngealisation* (2009).

[2] S. J. Perkell, M. Cohen, M. Svirsky, M. Matthies, I. Garabieta, and M. Jackson, "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," J. Acoust. Soc. Am. **92**, 3078–3096 (1992).

[3] T. Shawker, M. Stone, and B. Sonies, "Tongue pellet tracking by ultrasound: Development of a reverberation pellet," J. Phonetics **13**, 134–146 (1985).

[4] T. Toda, A. Black, and K. Tokuda, "Acoustic-to-articulatory inversion mapping with Gaussian mixture model," in *Proceedings of the ICSLP*, Jeju Island, Korea (2004), pp. 1129–1132.

[5] P. K. Ghosh and S. S. Narayanan, "A generalized smoothness criterion for acoustic-to-articulatory inversion," J. Acoust. Soc. Am. **128**(4), 2162–2172 (2010).

[6] A. Toutios and K. Margaritis, "Acoustic-to-articulatory inversion of speech: A review," in *Proceedings of the International 12th TAINN* (2003).

[7] F. Faubel, J. McDonough, and D. Klakow, "Bounded conditional mean imputation with Gaussian mixture models: A reconstruction approach to partly occluded features," IEEE Trans. Acoust., Speech, Signal Process. **1**, 3869–3872 (2009).

[8] R. M. Gray, “Toeplitz and circulant matrices: A review,” Found. Trends Commun. Inf. Theory **2**(3), 155–329 (2005) (available at http://ee.stanford.edu/ gray/toeplitz.pdf).

[9] A. A. Wrench and H. J. William, “A multichannel articulatory database and its application for automatic speech recognition,” in *5th Seminar on Speech Production: Models and Data*, Bavaria (2000), pp. 305–308.

[10] D. R. Cox and D. V. Hinkley, *Theoretical Statistics* (Chapman and Hall, London, 1974), Appendix 3.