MAXIMUM A-POSTERIORI ESTIMATION OF MISSING SAMPLES WITH CONTINUITY CONSTRAINT IN ELECTROMAGNETIC ARTICULOGRAPHY DATA

Sujith P

Electrical Communication Engineering
Indian Institute of Science (IISc)
Bangalore, India
sujith@ece.iisc.ernet.in

Prasanta Kumar Ghosh

Electrical Engineering
Indian Institute of Science (IISc)
Bangalore, India
prasantg@ee.iisc.ernet.in

ABSTRACT

Electromagnetic Articulography (EMA) technique is used to record the kinematics of different articulators while one speaks. EMA data often contains missing segments due to sensor failure. In this work, we propose a maximum a-posteriori (MAP) estimation with continuity constraint to recover the missing samples in the articulatory trajectories recorded using EMA. In this approach, we combine the benefits of statistical MAP estimation as well as the temporal continuity of the articulatory trajectories. Experiments on articulatory corpus using different missing segment durations show that the proposed continuity constraint results in a 30% reduction in average root mean squared error in estimation over statistical estimation of missing segments without any continuity constraint.

Index Terms— EMA data, Missing sample estimation, Gaussian mixture model, Continuity constraint

1. INTRODUCTION

Electromagnetic Articulography (EMA) [1] provides the kinematics of sensors placed on articulator flesh points in the mid-sagittal plane. This is widely used for research in the articulation of speech, which offers an alternate representation to that in the acoustic domain. Other popular methods for articulatory movement recording are X-ray [2], X-ray microbeam [3], Magnetic Resonance Imaging (MRI) [4] and Ultrasound [5]. EMA data [6] consists of the X and Y coordinates of sensors placed on various articulators including upper lip, lower lip, jaw, tongue tip, tongue body, tongue dorsum and velum at a rate of 500Hz. The recorded data using EMA may contain missing segments because of sensor failure and sensor detachment [6]. However the recording of the articulatory movements using EMA is expensive in the sense that specialist facilities and expertise are typically required. The recording process itself can be tricky, with practical complication that may also increase burden on the subject [7]. Hence missing samples estimation technique [8] could be useful for reconstructing the missing samples rather than discarding the entire data containing missing segments.

In a previous work on missing samples estimation, Qin *et al.* [8] used minimum mean squared error (MMSE) criterion to predict the missing sample in each frame separately. Q. Fang *et al.* [9] used dynamical feature in addition to the articulatory feature to incorporates the temporal information. However they did not consider the fact that the trajectories formed by the movements of the articulators are smooth [10]. Lack of smoothness constraint could cause discontinuities at the boundaries between the missing and known segments as well as within the missing segments. Articulatory trajectories need to be slowly varying within the missing segments and also continuous at the boundaries between the missing segments and the known segments. This means that the reconstructed segments and the known segments should form a continuous trajectory. In contrast to MMSE, in this work we propose maximum a-posteriori estimation with continuity constraint (MAPC), where we constrain the estimates of the articulatory trajectories to be low-pass in nature. According to Toda *et al.* [11] "MMSE-based mapping is not appropriate for multiple probability density distributions because it ignores the covariances of the individual distributions even when they are different from each other". In MAP estimation, the covariance matrices are used as weights in the weighted sum of the means of individual distributions unlike MMSE based estimation. Covariance matrices are regarded as confidence measure for the means from individual mixture components [11]. Thus MAPC makes use of the statistical benefit in MAP estimation as well as the temporal continuity of the articulatory trajectory. We perform missing sample estimation on articulatory database and we find that, on average, MAPC achieves $\sim$30% reduction in RMSE in estimation compared to MMSE.

2. MAXIMUM A-POSTERIORI ESTIMATION WITH CONTINUITY CONSTRAINT

We perform maximum a-posteriori (MAP) estimation of missing samples subject to the constraint that the energy of the high frequency contents of the articulatory trajectory after missing sample estimation is low; this ensures that the trajectory varies slowly in time. Let $\mathbf{x}_n$ be a D dimensional articulatory feature vector, KI_n is the set of indices of known samples and MI_n is the set of indices of missing samples at the frame n. The estimator is a function that maps from known variables $\mathbf{x}_n^p$ to missing variable x_n^m, where $m \in \text{MI}_n$, p is KI_n and $|\text{MI}_n \cup \text{KI}_n|$ =D.

Let the q-th articulatory trajectory of length $N+1$ be denoted by $\mathbf{X}^q = \{x_0^q, \cdots, x_{N_1}^q, \cdots, x_N^q\}$, $q \in \text{MI}_n$[1]. Let the samples from the frame N_1 to N_2 are missing and represented by $\mathbf{X}^{q,m}= \{x_{N_1}^{q,m}, \cdots, x_{N_2-1}^{q,m}, x_{N_2}^{q,m}\}$, where $x_n^{q,m} = x_n^q$, $N_1 \leq n \leq N_2$ i.e., the super script 'm' is used to indicate missing sample. Let $\mathbf{X}^p = \{\mathbf{x}_{N_1}^p, \cdots, \mathbf{x}_{N_2-1}^p, \mathbf{x}_{N_2}^p\}$ be the set of vectors comprising samples of the known articulatory trajectories from the frame N_1 to N_2. The objective is to find an estimate of $\mathbf{X}^{q,m}$ that maximizes the joint a-posteriori probability function $f'\left(\mathbf{X}^{q,m}\right)$ $=p\left(x_{N_1}^{q,m}, \cdots, x_{N_2}^{q,m} | \mathbf{x}_{N_1}^p, \cdots, \mathbf{x}_{N_2}^p\right)$ while minimizing the energy of the high frequency components in the estimated trajectory. $p(.)$ denotes probability density function. The energy of the high-frequency components can be expressed as $\sum_{n=0}^{\infty} {y_n^q}^2$, where

[1] q is the index of one of the articulators which contain at least one missing segment in a given utterance.

$y_n^q = \sum_{k=0}^{\infty} h_k x_{n-k}^q$, h_n is the impulse response of a causal IIR high-pass filter[2] with cut-off frequency f_c. It is assumed that at each frame the conditional density is independent of samples of other frames so that the joint density can expressed as a product of that of the individual frames as $f^{'}\left(\mathbf{X}^{q,m}\right)$=$\prod_{n=N_1}^{N_2} p(x_n^{q,m}|\mathbf{x}_n^p)$.

The conditional density function $p(x_n^{q,m}|\mathbf{x}_n^p)$ can be obtained from the joint density function $p(\mathbf{x}_n^{m,p})$, where $\mathbf{x}_n^{m,p} = [x_n^{q,m}\ \mathbf{x}_n^{p\,T}]^T$, where T is the transpose operator. $p(\mathbf{x}_n^{m,p})$ is modeled a-priori using frames with no missing samples. The joint probability density function is expressed using Gaussian Mixture model (GMM) [12] with M number of components as follows:

$$p(\mathbf{x}_n^{m,p}) = \sum_{r=1}^{M} \pi_r \mathcal{N}\left(\mathbf{x}_n^{m,p}; \mu_r, \Sigma_r\right) \tag{1}$$

$$\Sigma_r = \begin{bmatrix} \Sigma_{r,mm} & \Sigma_{r,mp} \\ \Sigma_{r,mp}^T & \Sigma_{r,pp} \end{bmatrix}, \mu_r = \begin{bmatrix} \mu_{r,m} \\ \mu_{r,p} \end{bmatrix},$$

where $\mu_{r,m}$ is the mean of the missing sample of r-th mixture component, $\mu_{r,p}$ is the mean vector of known samples of r-th mixture component, $\Sigma_{r,mm}$ is the variance of the missing sample, $\Sigma_{r,pp}$ is the covariance matrix of the known samples and $\Sigma_{r,mp}$ is the cross-covariance matrix between the missing sample and the known samples of r-th mixture component.

The conditional density function can then be written as a GMM as follows:

$$p(x_n^{q,m}|\mathbf{x}_n^p) = \sum_{r=1}^{M} \pi_n^r \mathcal{N}\left(x_n^{q,m}; \mu_n^r, \sigma_n^r\right) \tag{2}$$

$$\text{where } \pi_n^r = \pi^r \mathcal{N}\left(\mathbf{x}_n^p; \mu_{r,p}, \Sigma_{r,pp}\right)/p(\mathbf{x}_n^p)$$

$$\mu_n^r = \mu_{r,m} + \Sigma_{r,mp}\Sigma_{r,pp}^{-1}\left(\mathbf{x}_n^p - \mu_{r,p}\right)$$

$$\sigma_n^r = \Sigma_{r,mm} - \Sigma_{r,mp}\Sigma_{r,pp}^{-1}\Sigma_{r,mp}{}^T$$

Then

$$f^{'}\left(\mathbf{X}^{q,m}\right) = \prod_{n=N_1}^{N_2} \sum_{r=1}^{M} \pi_n^r \mathcal{N}\left(x_n^{q,m}; \mu_n^r, \sigma_n^r\right) \tag{3}$$

Let random variables $Z_{nr}, n = N_1, \cdots, N_2$ and $r = 1, 2, \cdots, M$ denote the information about which mixture component each sample $x_n^{q,m}$ comes from. For each n, Z_{nr}=1 if $x_n^{q,m}$ came from r-th mixture component. We would have $\sum_r Z_{nr}$=1. Also, we have $\mathbf{P}(Z_{nr}= 1)= \pi_n^r$ and $p(x_n^{q,m}|Z_{nr} = 1)$=$\mathcal{N}\left(x_n^{q,m}; \mu_n^r, \sigma_n^r\right)$. So the complete a-posteriori probability is a function of $\mathbf{Z}$ $\left([\mathbf{Z}]_{nr} = Z_{nr}\right)$ and $\mathbf{X}^{q,m}$, denoted by $f^{'}\left(\mathbf{X}^{q,m}, \mathbf{Z}\right)$.

$$f^{'}\left(\mathbf{X}^{q,m}, \mathbf{Z}\right) = f^{'}\left(\mathbf{X}^{q,m}|\mathbf{Z}\right)p\left(\mathbf{Z}\right)$$

$$= \prod_{n=N_1}^{N_2}\prod_{r=1}^{M}\left(\pi_n^r \mathcal{N}\left(x_n^{q,m}; \mu_n^r, \sigma_n^r\right)\right)^{Z_{nr}} \tag{4}$$

Incorporating the continuity constraint along with logarithm[3] of $f^{'}\left(\mathbf{X}^{q,m}, \mathbf{Z}\right)$ i.e., $\log\left(f^{'}\left(\mathbf{X}^{q,m}, \mathbf{Z}\right)\right)$, the modified objective function to be maximized becomes,

$$f\left(\mathbf{X}^{q,m}, \mathbf{Z}\right) = \log\left(f^{'}\left(\mathbf{X}^{q,m}, \mathbf{Z}\right)\right) - \lambda \sum_{n=0}^{\infty} y_n^{q\,2}$$

$$= \sum_{n=N_1}^{N_2}\sum_{r=1}^{M} Z_{nr} \log\left(\pi_n^r \mathcal{N}\left(x_n^{q,m}; \mu_n^r, \sigma_n^r\right)\right) - \lambda \sum_{n=0}^{\infty} y_n^{q\,2}$$

$$= \sum_{n=N_1}^{N_2}\sum_{r=1}^{M} Z_{nr}\left(\log(\pi_n^r) - \frac{\left(x_n^{q,m} - \mu_n^r\right)^2}{2\sigma_n^r} - \left(\log(\sqrt{2\pi\sigma_n^r}\right)\right)$$

$$- \lambda\left(\sum_{n=0}^{N_1-1} y_n^{q\,2} + \sum_{n=N_1}^{\infty} y_n^{q\,2}\right), \tag{5}$$

where λ is a positive number and $\mathbf{X}^{q,m}$ is the set of optimization variables. We use a negative sign before the second term in eqn (5) because the first term in the objective function $f\left(\mathbf{X}^{q,m}, \mathbf{Z}\right)$ has to be maximized while the second term has to be minimized. The variable Z_{nr} in eqn (5) is also unknown, so we use the expectation maximization (EM) algorithm [13] for maximization. The objective function $f\left(\mathbf{X}^{q,m}, \mathbf{Z}\right)$ in $(t+1)$-th iteration becomes $f\left(\mathbf{X}^{q,m}, \mathbf{Z}, {}_t\hat{\mathbf{X}}^{q,m}\right)$, where ${}_t\hat{\mathbf{X}}^{q,m} \in \mathbb{R}^{N_2-N_1+1}$ is a vector whose elements ${}_t x_n^{q,m}$ are the estimates of the samples of the missing segment $\mathbf{X}^{q,m}$ at the t-th iteration of the EM algorithm. In the first iteration of EM, we initialize ${}_0\hat{\mathbf{X}}^{q,m}$ using the MMSE estimate and in subsequent iterations we use the estimate from previous iteration.

Below we describe the steps involved in the estimation process using the EM algorithm at the $(t + 1)$-th iteration after discarding the terms that are independent of the optimization variables in eqn (5).

Expectation step:

$$f\left(\mathbf{X}^{q,m}\right) = E_{\mathbf{Z}}\left[f\left(\mathbf{X}^{q,m}, \mathbf{Z}, {}_t\hat{\mathbf{X}}^{q,m}\right)\right] = f_1\left(\mathbf{X}^{q,m}\right) + f_2\left(\mathbf{X}^{q,m}\right)$$

$$= \sum_{n=N_1}^{N_2}\sum_{r=1}^{M} - {}_t\hat{\pi}_n^r\left(\frac{\left(x_n^{q,m} - \mu_n^r\right)^2}{2\sigma_n^r}\right) - \lambda \sum_{n=N_1}^{\infty} y_n^{q\,2}, \tag{6}$$

where, ${}_t\hat{\pi}_n^r = E\left[Z_{nr}|{}_t\hat{x}_n^{q,m}\right] = p\left(r|\mathbf{x}_n^p, {}_t\hat{x}_n^{q,m}\right)$, conditional probability of the r-th mixture component. $E(.)$ denotes the expectation operator.

Maximization step:

For maximization we need to satisfy the first order necessary condition – differentiate $f\left(\mathbf{X}^{q,m}\right)$ and equate it to zero.

$$\frac{\partial f_1}{\partial x_i^{q,m}} = x_i^{q,m}\left(-\sum_{r=1}^{M}\frac{{}_t\hat{\pi}_i^r}{\sigma_i^r}\right) + \sum_{r=1}^{M}\frac{{}_t\hat{\pi}_i^r}{\sigma_i^r}\mu_i^r \tag{7}$$

$$\frac{\partial f_2}{\partial x_i^{q,m}} = -\lambda\sum_{n=N_1}^{\infty}\left(\sum_{k=0}^{\infty} h_k x_{n-k}^q\right)h_{n-i}$$

$$= -\lambda\left(\sum_{l=N_1}^{N_2} a_{il}x_l^{q,m} + b_i\right), \tag{8}$$

$$\text{where } a_{ij} = R_{(i-j)}^h \triangleq \sum_{n=0}^{\infty} h_n h_{n-(i-j)}, \tag{9}$$

$$b_i = \sum_{n=i}^{\infty}\left(\sum_{k=0}^{\infty}\tilde{x}_{n-k}^q h_k\right)h_{n-i} = \tilde{x}_n^q \star R_n^h \text{ at } i,$$

$$\text{where, } \tilde{x}_k^q = \begin{cases} 0, & \text{if } N_1 \le k \le N_2 \\ x_k^q, & \text{otherwise} \end{cases}$$

where $N_1 \le i, j \le N_2$ and $\star$ denotes the convolution operator. Note that we need to know only the autocorrelation sequence of the high-pass filter h_n for optimization. Using eqn (7), (8) and considering $-\frac{\partial f}{\partial x_i^{q,m}} = -\left(\frac{\partial f_1}{\partial x_i^{q,m}} + \frac{\partial f_2}{\partial x_i^{q,m}}\right) = 0$, $N_1 \le i \le N_2$, we can form

[2]Although we have used causal IIR filter in the present formulation, one can choose any FIR or IIR filter

[3]Note that, as logarithm is a monotonic function, this does not change the solution of the optimization

a matrix-vector equation as follows:

$$(\mathbf{A}_1 + \lambda\mathbf{A}_2)_{t+1}\hat{\mathbf{X}}^{q,m} + (\mathbf{B}_1 + \lambda\mathbf{B}_2) = \mathbf{0}, \tag{10}$$

where matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ come from eqn (7), $\mathbf{A}_2$ and $\mathbf{B}_2$ come from eqn (8).

$$[\mathbf{A}_1]_{i_1 i_1} = \sum_{r=1}^{M} \frac{{}_t\hat{\pi}^r_{i_1+N_1-1}}{\sigma^r_{i_1+N_1-1}}, \quad [\mathbf{A}_2]_{i_1 i_2} = a_{i_1+N_1-1, i_2+N_1-1},$$

$$[\mathbf{B}_1]_{i_1} = \sum_{r=1}^{M} \frac{{}_t\hat{\pi}^r_{i_1+N_1-1}}{\sigma^r_{i_1+N_1-1}} \mu^r_{i_1+N_1-1}, \quad [\mathbf{B}_2]_{i_1} = b_{i_1+N_1-1},$$

$$1 \leq i_1, i_2 \leq N_2 - N_1 + 1$$

$\mathbf{A}_1$ and $\mathbf{A}_2$ are square matrices and $(\mathbf{B}_1 + \lambda\mathbf{B}_2)$ is a column matrix. $\mathbf{A}_2$ is the auto-correlation matrix of filter impulse response. Hence, $\mathbf{A}_2$ is a toeplitz and positive semidefinite matrix. $\mathbf{A}_1$ is positive definite because it is a diagonal matrix with positive entries. Thus, $(\mathbf{A}_1 + \lambda\mathbf{A}_2)$ is a positive definite matrix and the inverse of $(\mathbf{A}_1 + \lambda\mathbf{A}_2)$ exists. From eqn (7) and (8) the Hessian is $(-\mathbf{A}_1 - \lambda\mathbf{A}_2)$ which is a negative definite matrix that gives the second order sufficient condition. Thus, from eqn (10), we obtain the estimate at the $(t+1)$-th iteration as follows:

$${}_{t+1}\hat{\mathbf{X}}^{q,m} = -(\mathbf{A}_1 + \lambda\mathbf{A}_2)^{-1}(\mathbf{B}_1 + \lambda\mathbf{B}_2) \tag{11}$$

We use the following stopping criterion for the EM iterations:

$$\frac{1}{N_2 - N_1 + 1} \|{}_{t+1}\hat{\mathbf{X}}^{q,m} - {}_t\hat{\mathbf{X}}^{q,m}\|_2^2 < 10^{-4}$$

3. EXPERIMENTAL EVALUATION

3.1. DATASET

For the experiments, we use the Multichannel Articulatory (MOCHA) [6] database which contains 460 utterances each spoken by one male and one female subjects. Preprocessing is done by following the steps outlined in [14] and thus the acoustic and articulatory data at a rate of 100 frames per second are obtained. We use 14 dimensional EMA feature vector in each frame denoted by 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 an analysis window of 20 ms length with 10 ms shift.

3.2. EXPERIMENTAL SETUP

We follow the same experimental procedure used by Qin *et al.* [8] by reconstructing artificially blacked-out portion of one articulator's trajectory, which is treated as the missing segment. The initial point of the missing segment is selected at random and we consider four different missing segment durations (MSD) 100ms, 200ms, 400ms and 800ms. We also black out both X and Y trajectories of one sensor simulating the failure of that sensor. Thus missing segments of two trajectories are estimated using remaining 12 sensor's data. The joint distribution $p(\mathbf{x}_n^{m,p})$ is modeled using the training data with GMM with 8 components. We have not got significant performance gain by using higher number of mixture components. We consider a five-fold cross validation set-up using 368 sentences for training and remaining 92 as the test set in each fold. Here we compare the performances of the MMSE and the MAPC for missing sample estimation task. In addition to the known articulatory features we also used acoustic features in the respective frames to investigate if there is any performance benefit by using acoustic features. We refer to this scheme by MAPCA. We also estimate the missing samples using a linear function (LIN). In LIN the missing samples are estimated as the points on the line joining the two end points of the missing segment x_{N_1-1} and x_{N_2+1}. The equation of the line is $\hat{x}_n^{q,m} = \alpha n + x^q_{N_2+1} - \alpha(N_2+1)$, where n is the frame index and $\alpha = \left(\frac{x^q_{N_2+1} - x^q_{N_1-1}}{N_2 - N_1 + 2}\right)$.

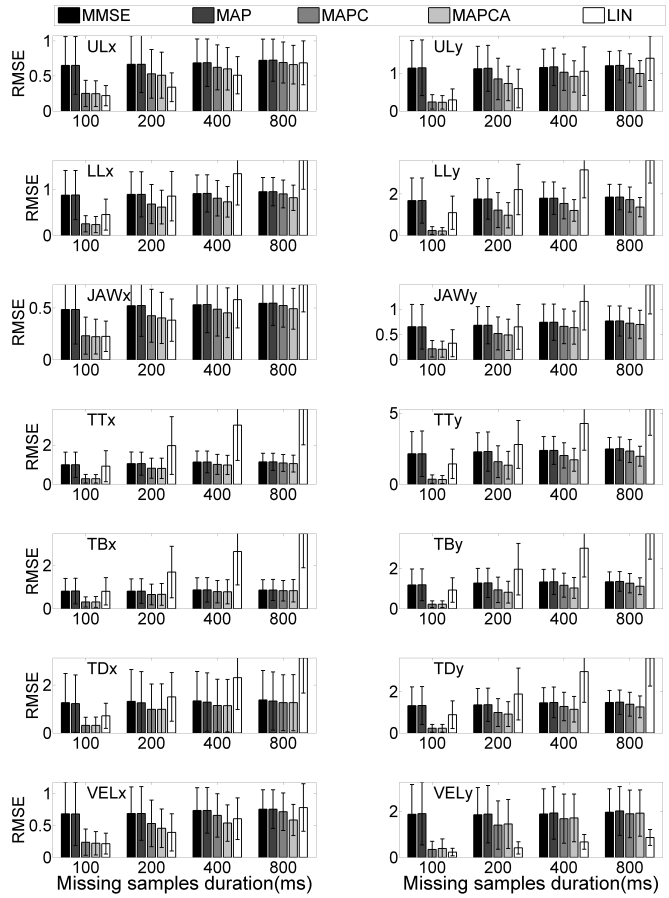

Fig. 1. RMSE (average± one standard deviation) of estimation for different articulators for different MSD for male speaker.

Experiments are conducted using a Chebyshev FIR[4] high-pass filter of order L. We consider different values of the hyper-parameters – f_c, L, and λ. Experiments are repeated with the following parameter sets: $f_c = \{5\text{Hz}, 10\text{Hz}, 20\text{Hz}, 40\text{Hz}\}$, $L = \{20, 30, 60\}$ and $\lambda = \{1, 10^{-1}, 10^{-4}, 10^{-7}, 10^{-10}, 10^{-12}\}$. We report the result corresponding to the parameter set which yields the best performance on the test set, i.e., L=30, f_c = 20Hz and $\lambda = 10^{-7}$.

3.3. RESULTS AND DISCUSSION

Fig. 1 and Fig. 2 show the reconstruction performances in terms of the root mean square error (RMSE) between the original and estimated segments for different MSD for male and female subjects respectively. The x-axis represents various MSD and the y-axis represents the RMSE in mm scale. From the figures we can see that in terms of the average RMSE, MAPC[5] performs better than MMSE and the performance gain of MAPC decreases with increasing MSD. For example, the reductions in the average RMSE by using MAPC over MMSE for the male subject are 76.3%, 25.6%, 11.9% and

[4]Although an IIR filter could be chosen as the high-pass filter, we have experimented with FIR filter in this work.

[5]We have also performed ML estimation by maximizing $p(\mathbf{x}_n^p|x_n^{q,m})$ but the results are inferior to that of the MMSE and MAP estimation.

5.5% for the MSD of 100ms, 200ms, 400ms and 800ms respectively. In the case of female subject, these reductions are 78.9%, 26.9%, 12.9% and 6.2%. The performance of MMSE is almost constant for different MSD. This is because in MMSE estimation each sample of a trajectory is estimated independently from the other samples in that trajectory.

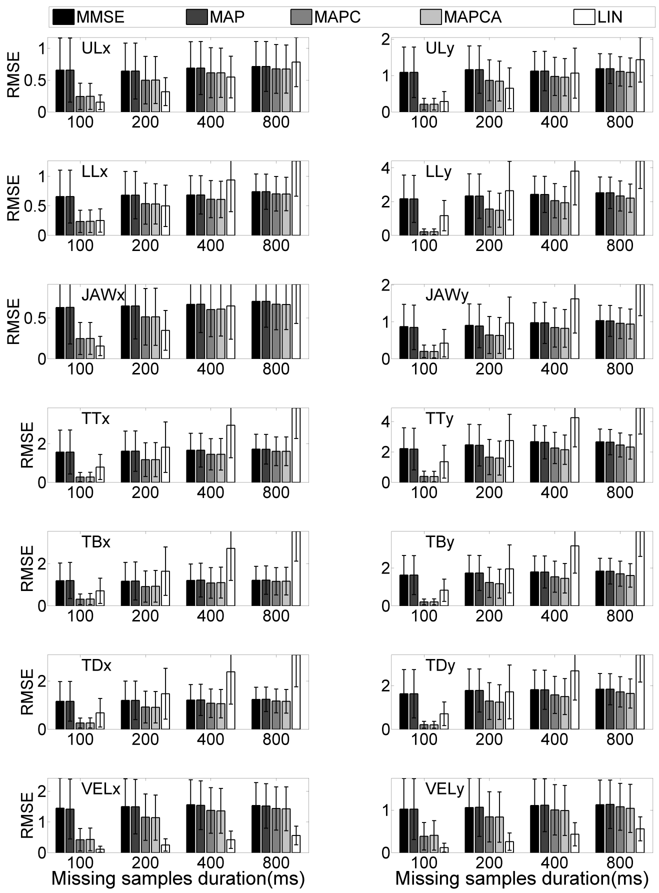

Fig. 2. RMSE (average± one standard deviation) of estimation for different articulators for different MSD for female speaker.

The performance of LIN [6] decreases with increasing MSD. Considering all articulators of both subjects, the average values of RMSE in the case of LIN are 0.56, 1.22, 1.96 and 2.48 mm for MSD of 100ms, 200ms, 400ms and 800ms respectively. We can see that for small MSD, LIN is better than MMSE but MAPC is even better than LIN in most of the cases except for VELx and VELy. In the case of MAPC, average RMSE reduces by 52%, 20%, 39% and 46% (for four MSD) over LIN. Thus, overall MAPC is found to yield lower average RMSE compared to both the baseline (LIN) and the MMSE approach. MAPCA is better than MAPC but the average reductions in RMSE are 0.3%, 2.5%, 2.8% and 2.9% for the MSD 100ms, 200ms, 400ms, and 800ms respectively. Thus there is no significant benefit by using acoustic features in addition to the known articulatory features for missing samples estimation task. We observe that MAP (i.e., MAPC without continuity constraint) performs similar to MMSE. This is because the conditional density $p(x_n^{q,m}|\mathbf{x}_n^p)$ is found to be dominated by a single Gaussian in most of the frames resulting in similar estimates using MAP and MMSE. Since the performance of MAP[7] and MMSE are almost similar, the benefit is coming from the continuity constraint.

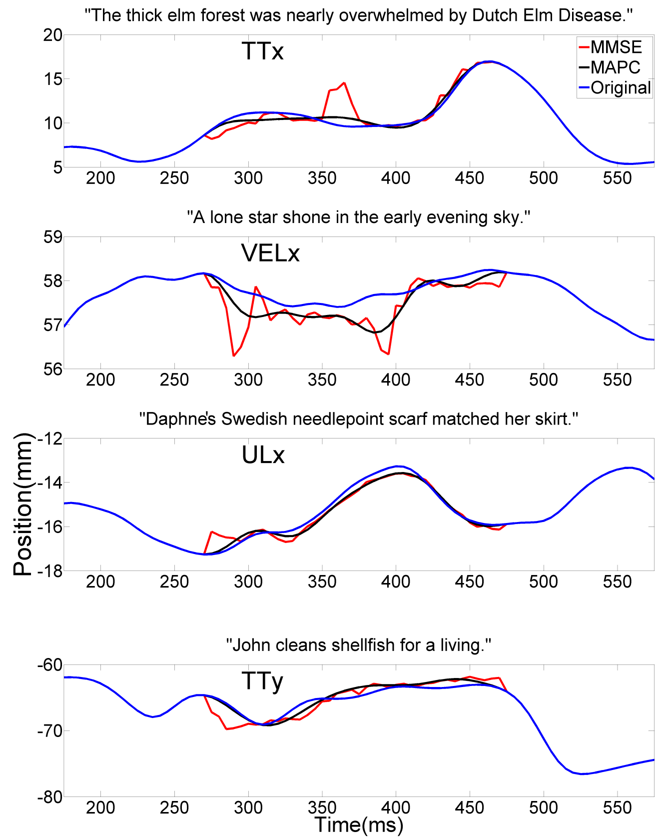

Fig. 3. Reconstructed and original articulatory trajectories for different utterances from the female speaker of MOCHA database .

Fig. 3 illustrates a few reconstructed and the original articulatory trajectories for different utterances from the female subject in the MOCHA database for MSD of 200ms. We can see from Fig. 3 that the reconstructed trajectories in the case of MAPC [8] are smoother compared to MMSE. Also there is no discontinuity at the boundaries between the known segments and the reconstructed segments in the case of MAPC unlike MMSE.

4. CONCLUSIONS

We have proposed a maximum a-posteriori estimation with continuity constraint (MAPC) for the task of missing sample reconstruction in articulatory trajectory. We have found that adding continuity constraint improves the performance of the missing sample reconstruction. However, the benefit of using continuity constraint decreases with increasing missing segment length. When a single sensor data is missing, we have found that there is no additional reconstruction benefit by using acoustic features in addition to the articulatory data from remaining known sensors. The estimation performance varies across different articulators; this could be due to different correlation between missing articulator and the known articulators. MAPC approach could also be useful for missing value imputation in other applications such as gene expression data [15, 16] and feature reconstruction in automatic speech recognition [17, 18].

[6]Polynomial interpolation is also carried out but the results are similar to that of LIN method.

[7]MAP+low-pass filter is not considered because it does not guarantee the continuity at the boundary of known and missing segments.

[8]A Matlab implementation of MAPC is available for download at http://www.ee.iisc.ernet.in/new/people/faculty/prasantg/Softwares/MAPC_missingsample_ICASSP2014.tar.gz

5. REFERENCES

[1] P. W. Schönle, K. Gräbe, P. Wenig, J. Höhne, J. Schrader, and B. Conrad, "Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract," *Brain and Language*, vol. 31, no. 1, pp. 26–35, 1987.

[2] Y. Laprie and M.-O. Berger, "Extraction of tongue contours in x-ray images with minimal user interaction," in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, vol. 1, pp. 268–271, 1996.

[3] J. Westbury, P. Milenkovic, G. Weismer, and R. Kent, "X-ray microbeam speech production database," *The Journal of the Acoustical Society of America*, vol. 88, p. S56, 1990.

[4] E. Bresch, Y.-C. Kim, K. Nayak, D. Byrd, and S. Narayanan, "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging," *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 123–132, 2008.

[5] M. Stone, "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data," *The Journal of the Acoustical Society of America*, vol. 87, pp. 2207–2217, 1990.

[6] A. A. Wrench, "A multi-channel/multi-speaker articulatory database for continuous speech recognition research.," *Phonus.*, vol. 5, pp. 1–13, 2000.

[7] I. Steiner, K. Richmond, I. Marshall, and C. D. Gray, "The magnetic resonance imaging subset of the mngu0 articulatory corpus," *The Journal of the Acoustical Society of America*, vol. 131, no. 2, pp. EL106–EL111, 2012.

[8] C. Qin and M. A. Carreira-Perpinán, "Estimating missing data sequences in x-ray microbeam recordings.," in *Proc.INTERSPEECH*, pp. 1592–1595, 2010.

[9] Q. Fang, J. Wei, F. Hu, A. Li, and H. Wang, "Estimating the position of mistracked coil of ema data using gmm-based methods," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2013 Asia-Pacific*, pp. 1–4, IEEE, 2013.

[10] P. K. Ghosh and S. S. Narayanan, "On smoothing articulatory trajectories obtained from gaussian mixture model based acoustic-to-articulatory inversion," *The Journal of the Acoustical Society of America*, vol. 134, no. 2, pp. EL258–EL264, 2013.

[11] T. Toda, A. W. Black, and K. Tokuda, "Statistical mapping between articulatory movements and acoustic spectrum using a gaussian mixture model," *Speech Communication*, vol. 50, no. 3, pp. 215–227, 2008.

[12] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, vol. 1. springer New York, 2006.

[13] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.

[14] P. K. Ghosh and S. Narayanan, "A generalized smoothness criterion for acoustic-to-articulatory inversion," *The Journal of the Acoustical Society of America*, vol. 128, pp. 2162–2172, 2010.

[15] H. Kim, G. H. Golub, and H. Park, "Missing value estimation for dna microarray gene expression data: local least squares imputation," *Bioinformatics*, vol. 21, no. 2, pp. 187–198, 2005.

[16] X. Wang, A. Li, Z. Jiang, and H. Feng, "Missing value estimation for dna microarray gene expression data by support vector regression imputation and orthogonal coding scheme," *BMC bioinformatics*, vol. 7, no. 1, p. 32, 2006.

[17] B. Raj and R. M. Stern, "Missing-feature approaches in speech recognition," *Signal Processing Magazine, IEEE*, vol. 22, no. 5, pp. 101–116, 2005.

[18] M. Cooke, P. Green, L. Josifovski, and A. Vizinho, "Robust automatic speech recognition with missing and unreliable acoustic data," *Speech communication*, vol. 34, no. 3, pp. 267–285, 2001.