ESTIMATING MISSING DATA SEQUENCES IN X-RAY MICROBEAM RECORDINGS

Chao Qin and **Miguel Á. Carreira-Perpiñán**

EECS, School of Engineering, University of California, Merced

1 Abstract

Techniques for recording the vocal tract shape during speech such as X-ray microbeam or EMA track the spatial location of pellets attached to several articulators. Limitations of the recording technology result in most utterances having sequences of frames where one or more pellets are missing. Rather than discarding such sequences, we seek to reconstruct them. We use an algorithm for recovering missing data based on learning a density model of the vocal tract shapes, and predicting missing articulator values using conditional distributions derived from this density. Our results with the Wisconsin X-ray microbeam database show we can recover long, heavily oscillatory trajectories with errors of 1 to 1.5 mm for all articulators.

2 Motivation and idea

- Mistracked or missing pellets often occur during recordings with X-ray microbeam or EMA. Reasons for mistracks:
  - Pellets fall off or sensor malfunction.
  - Microbeam can't find the pellets it is looking for, or follows the wrong pellet.
- We reconstruct the missing tongue pellets from the locations of present ones.
- Fitting a single mapping from the present to the missing components is unsatisfactory because the missing components vary from frame to frame. We want to obtain a flexible, efficient way to construct mappings "on demand" between an arbitrary set of input and output variables.

Idea of the method

❶ Offline, estimate a joint density model $p(\mathbf{x})$ using a complete data set $\{\mathbf{x}_n\}$.

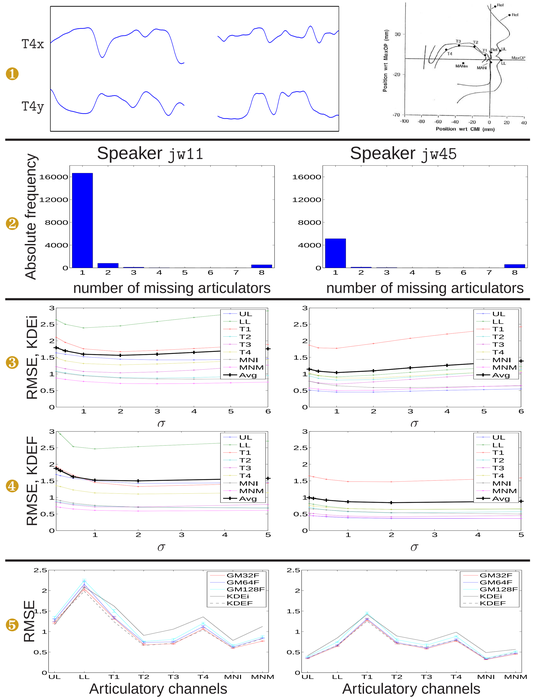
❷ At run time, for each frame $\mathbf{x}_t$, determine the missing components $\mathcal{M}_t$ and the present ones $\mathcal{P}_t$, and reconstruct $\mathcal{M}_t$ as $\mathbf{f}_{\mathcal{M}_t}(\mathbf{x}_{\mathcal{P}_t}) = \mathrm{E}_p\{\mathbf{x}_{\mathcal{M}_t}|\mathbf{x}_{\mathcal{P}_t}\}$.

3 Deriving mappings with varying sets of inputs and outputs from a density model

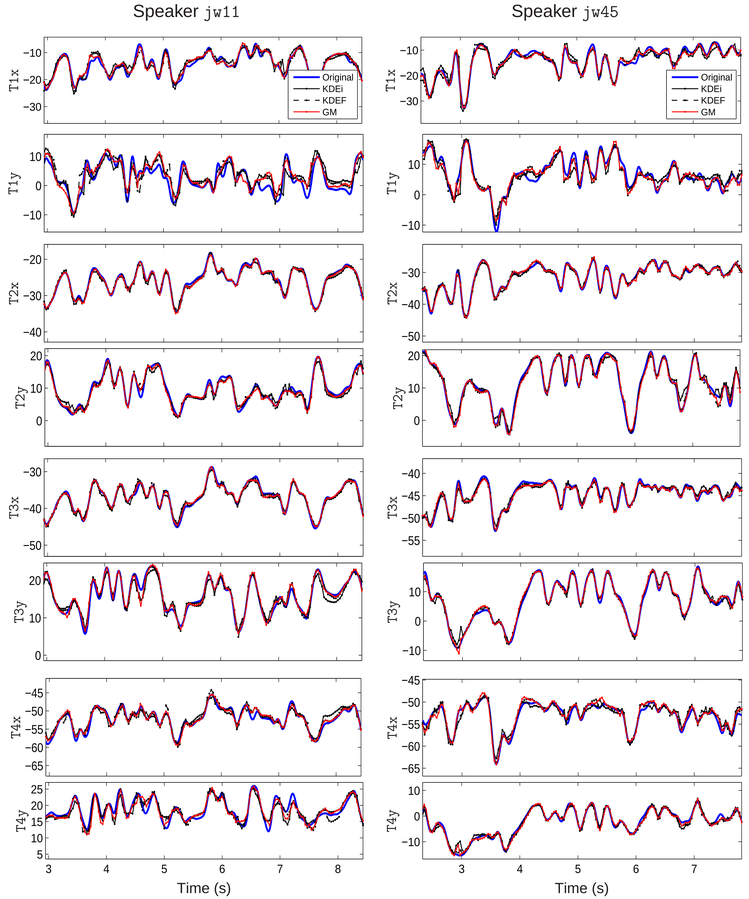
- To estimate the joint density model $p(\mathbf{x})$, we try two models:
  1. *Nonparametric Gaussian kernel density estimate (KDE)*. We try isotropic (KDEi) and full-covariance matrices (KDEF).
  2. *Parametric density estimate by Gaussian mixtures (GM)*. We try various numbers M of components, each with a full-covariance matrix $\boldsymbol{\Sigma}_m$.
- To compute the conditional distribution $p(\mathbf{x}_{\mathcal{M}}|\mathbf{x}_{\mathcal{P}}) = p(\mathbf{x})/p(\mathbf{x}_{\mathcal{P}})$ in terms of the joint and marginal distributions:

$$
\begin{aligned}
p(\mathbf{x}_{\mathcal{P}}) &= \textstyle\sum_{m=1}^{M} \pi_m \mathcal{N}(\mathbf{x}_{\mathcal{P}}; \boldsymbol{\mu}_{m,\mathcal{P}}, \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{P}}) \\
p(\mathbf{x}_{\mathcal{M}}|\mathbf{x}_{\mathcal{P}}) &= \textstyle\sum_{m=1}^{M} \pi_{m,\mathcal{M}|\mathcal{P}} \mathcal{N}(\mathbf{x}_{\mathcal{M}}; \boldsymbol{\mu}_{m,\mathcal{M}|\mathcal{P}}, \boldsymbol{\Sigma}_{m,\mathcal{M}|\mathcal{P}}) \\
\pi_{m,\mathcal{M}|\mathcal{P}} &= \pi_m \mathcal{N}(\mathbf{x}_{\mathcal{P}}; \boldsymbol{\mu}_{m,\mathcal{P}}, \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{P}}) / p(\mathbf{x}_{\mathcal{P}}) \\
\boldsymbol{\mu}_{m,\mathcal{M}|\mathcal{P}} &= \boldsymbol{\mu}_{m,\mathcal{M}} + \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{M}}^T \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{P}}^{-1} (\mathbf{x}_{\mathcal{P}} - \boldsymbol{\mu}_{m,\mathcal{P}}) \\
\boldsymbol{\Sigma}_{m,\mathcal{M}|\mathcal{P}} &= \boldsymbol{\Sigma}_{m,\mathcal{M}\mathcal{M}} - \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{M}}^T \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{P}}^{-1} \boldsymbol{\Sigma}_{m,\mathcal{P}\mathcal{M}} \\
\mathbf{f}(\mathbf{x}_{\mathcal{P}}) &= \mathrm{E}\{\mathbf{x}_{\mathcal{M}}|\mathbf{x}_{\mathcal{P}}\} = \textstyle\sum_{m=1}^{M} \pi_{m,\mathcal{M}|\mathcal{P}}(\mathbf{x}_{\mathcal{P}})\, \boldsymbol{\mu}_{m,\mathcal{M}|\mathcal{P}}(\mathbf{x}_{\mathcal{P}}).
\end{aligned}
$$

4 Reconstruction of artificially blacked-out data

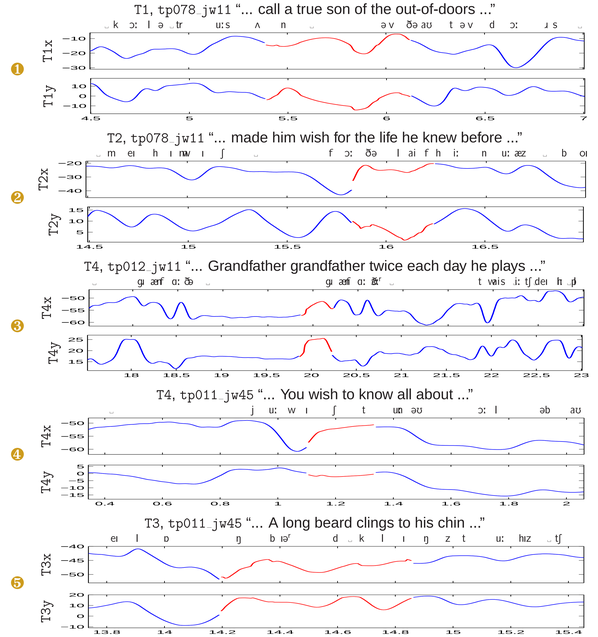


❶: typical mistracks for one pellet in the Wisconsin X-ray microbeam database (XRMB, pellet schematic at right); mistrack duration ≈ 0.5 sec. ❷–❺ show results for speakers `jw11` (left) and `jw45` (right). ❷: histogram of the number of missing articulators in XRMB for `jw11` and `jw45`, with mistrack percentage 11.32% and 3.55%, resp. ❸ (KDEi) and ❹ (KDEF): effect of the bandwidth σ on reconstructing missing articulators. Training and testing sets contain 50,000 and 10,000+ frames, resp. Each covariance is $\sigma^2\mathbf{I}$ in KDEi and $\sigma_F^2\boldsymbol{\Sigma}_m$ in KDEF, where $\boldsymbol{\Sigma}_m$ is estimated from the 100 nearest neighbors for each mixture component in the training set. The "Avg" curve is a weighted sum of the reconstruction error of each articulator, with weights inversely proportional to the respective error. ❺: reconstruction error for each missing articulator, for KDEi, KDEF and GM with $M = 32$, 64 and 128 components. Errorbars over 10 random initialisations of EM for training GMs.



Reconstruction of articulators `T1,T1,T3,T4` (top to bottom) artificially blacked-out over 5+ sec. for the utterance `tp011` by KDEi, KDEF, and GM with $M = 32$. The missing tracks are accurately reconstructed (average error 0.5 to 1.5 mm) even though they are highly variable. `T1` is more difficult to reconstruct than other tongue pellets. Note our method uses no temporal information: frames are reconstructed independently from each other.

5 Reconstruction of truly missing data



Reconstructions of truly missing articulators for speaker `jw11` (❶–❸) and `jw45` (❹–❺), all using the GM with $M = 32$ components.

6 Conclusion

We have extended an algorithm for missing data reconstruction and applied it to recovering missing pellet tracks in X-ray microbeam recordings, where the pellets are missing over extended periods, and the subset of missing pellets changes over time. A surprisingly parsimonious density model was sufficient to produce very accurate reconstructions for most pellets, even when the trajectory oscillates drastically over the period where it is missing. One limitation of the approach is that it relies on estimating a density model of the data ahead of time using a complete dataset (with no missing values). While this is not a problem with existing, large articulatory databases, future work should address reconstruction in more challenging situations, such as (near) real-time, or where little or no complete data are available for training.

Work funded by NSF award IIS-0754089.

INTERSPEECH 2010 Spoken Language Processing for All