Blind Separation of TMJ Sounds by Incorporating Fractional Delay Estimation

By Clive Cheong Took, Kianoush Nazarpour, Saeid Sanei, and Jonathon Chambers

Centre of Digital Signal Processing, Cardiff University, Cardiff CF24 3AA, Wales, UK,
Phone: +44-29-20879063, Email: \{cheongc, nazarpourk, saneis, chambersj\}@cf.ac.uk

Abstract

Temporomandibular joint (TMJ) sounds are generated from the two joints holding the lower jaw and the temporal bone (the skull) in patients suffering from temporomandibular disorder (TMD). In this study, we shall show how the inherent fractional delay between a pair of TMJ sources can be estimated by time localised sparse component analysis. In effect, the active periods of one of the sources will assist us in estimating the delay by employing modified versions of either the maximum likelihood (MLE) delay estimator or the mixed modulated Lagrange explicit time delay estimator (MMLETDE). Thereafter, we consider a strategy based on delaying the mixtures to solve the blind source separation of the TMJ sources. Simulation studies confirm the efficacy of the proposed scheme.

1. Introduction

TMD refers to all medical problems related to the region of mandible (lower jaw) and the temporal bone. TMD is traditionally diagnosed by stethoscope auscultation. Hence, TMJ sounds have long been associated with TMD [Leader et al. (2001)]. TMJ sources can be categorised into four types: soft click, hard click, soft crepitus, and hard crepitus [Took et al. (2006), Cheong Took et al. (2006)]. The task of detecting the right TMJ source is particularly difficult if the TMJ sounds (recorded from a pair of stethoscopes) are mixtures of the two TMJ sources. In our study the mixing model is represented as:

\[ x_i(t) = \sum_{j=1}^{2} h_{ij} s_j(t - \delta_{ij}) + v_i(t) \quad (1.1) \]

where \( x_i(t) \) is the \( i \)th TMJ mixture signal and \( v_i(t) \) represents additive zero mean white Gaussian noise at discrete time \( t \) for \( i = 1,2 \). \( h_{ij} \) are the attenuation coefficients and \( \delta_{ij} \) are the time delays associated with the path from the \( j \)th source to the \( i \)th sensor (stethoscope). The problem of recovering \( s_i(t) \) from the mixtures can be formulated as blind source separation (BSS). Likewise, the mixing matrix \( H \) can be defined as follows:

\[
\begin{pmatrix}
h_{11}z^{-d} & h_{12}z^{-(d+D)} \\
h_{21}z^{-(d+D)} & h_{22}z^{-d}
\end{pmatrix}
\quad (1.2)
\]

where \( d \) is the time for a TMJ source to reach to its ipsi (same side of the head) sensor and \( D \) is the time interval between the two contra (opposite side of the head) sources. Both delays are in terms of samples. Besides, we have considered \( d \) and \( D \) to be fractional since we do not know the exact head size, sampling rate, and velocity of sound in the tissue. The following assumptions are made: 1) all the source signals are super-Gaussian and mutually
Blind Separation of TMJ Sounds by Incorporating Fractional Delay Estimation

2 statistically independent (as empirically investigated); 2) the mixing matrix is full rank; 3) there is at least one active period of one source within the mixtures. The aim of this paper is to investigate how time localised sparse component analysis (SCA) can help to determine $D$ in the undertaken anechoic mixtures. The organisation of the paper is as follows; firstly we briefly review the MMLETDE and MLE delay estimators and propose their modified versions. Also, we introduce the concept of sparsity and highlight the words "time localised" in the framework of SCA. Thereafter, we consider the algorithm presented in [Cheong Took et al. (2006)] to reconstruct the sources. Section 3 compares the performance of our algorithm against algorithms such as that of Parra [Parra, L. & Spence, C. (2000)], exploiting the nonstationarity of the sources (convolutive modelling), and DUET of Özgür and Rickard [Özgür, Y. & Rickard, S. (2004)], which performs time-frequency masking (anaechoic modelling). Finally, we draw conclusions in Section 4.

2. Background

2.1. Fractional Delay Estimators

For discrete time signals, whenever we delay a signal $s(t)$ by a non-integer delay $D$, we require the computation of the subsample between $s(t-\lfloor D \rfloor)$ and $s(t-\lfloor D \rfloor-1)$ where $\lfloor \cdot \rfloor$ denotes the floor operation. To approximate this non-integer delay, we can formulate $s(t-D)$ as [Cheng, Z. & Tjhung, T. T. (2003)]:

$$s(t-D) = \sum_{k=-\infty}^{k=\infty} U(k)s(t-k)$$  \hspace{1cm} (2.1)

where $U(k) = \prod_{i=-M_a}^{M_b} \frac{D-i}{k-i}$, $i \neq k$ is the well-known Lagrange interpolation finite impulse response (FIR) filter. Note that either $M_a = M_b = N/2$ when $N$ is even, or $M_a = (N-1)/2$, $M_b = (N+1)/2$ when $N$ is odd, and $N = \text{round}(2D)$ is the order of the filter. The mixed modulated Lagrange explicit time delay estimator (MMLETDE) [Cheng, Z. & Tjhung, T. T. (2003)] can be summarised as follows:

$$\text{error}(q) = s(t-D) - \sum_{k=-M_a}^{k=M_b} \left( \hat{U}(k)e^{j\omega\nu} \right)s(t-k)$$  \hspace{1cm} (2.2)

where $\nu = k - \hat{D}(q)$

$$\hat{D}(q+1) = \hat{D}(q) - 2\mu \text{Re} \left\{ \text{error}(q) \sum_{k=-M_a}^{k=M_b} g(\nu)s(t-k) \right\}$$  \hspace{1cm} (2.3)

where

$$g(\nu) = e^{j\omega\nu} \left( \frac{\cos(\pi\nu) - \text{sinc}(\nu)}{\nu} - j\omega\text{sinc}(\nu) \right) \text{ and } j \triangleq \sqrt{-1}$$

Note that $\hat{U}(k)$ is the estimate of (2.1) when $D$ is substituted with $\hat{D}$, while $q$ denotes the $q$th iteration of the algorithm. From (2.2), we can see that the Lagrange interpolation FIR filter coefficients $U(k)$ are modulated by $e^{j\omega\nu}$ where $\omega$ is an arbitrary angular frequency. We suggest it is reasonable to remove $e^{j\omega\nu}$ from (2.2). Its verification is omitted due to the lack of space. On the other hand, the maximum likelihood delay estimator (MLE) given in [Kay (1993)] is based on the cross-correlation between the delayed signal and
the original one. However, we cannot use this estimator in its present form since the unknown delay can be fractional. Therefore, we propose to employ \( \sum_{k=-M_a}^{M_b} \hat{U}(k)s(t-k) \) to model the fractional delay and maximise mutual information (MI) between \( s(t-D) \) and \( s(t-D) \) instead of the cross-correlation, i.e. :

\[
\hat{D} = \arg \max_{\hat{D}} \left\{ \text{MI} \left( s(t-D), \sum_{k=-M_a}^{M_b} \hat{U}(k)s(t-k) \right) \right\} \tag{2.4}
\]

Due to limited space, we will not show the simulation result demonstrating the superiority of the modified MLE estimator over the original one.

2.2. Sparsity of TMJ sources

Sparsity in this work refers to the situation where a relatively small number of source signals are active over any particular time interval. For the case of a single active source, sparsity [Took et al. (2006)] can be mathematically described as

\[
\{ s_i(t); \ i = 1, \ldots, n \}
\]

where \( \forall t \exists k \in 1, \ldots, n \) where \( |s_k(t)| >> |s_j(t)| \) (2.5)

and for \( j \neq k \) \( s_j(t) \approx 0 \)

where \( s_i(t) \) is a given source signal and \( s_j(t) \) is another arbitrary source signal. We have assumed that this sparsity condition is fulfilled for all values of \( t \) in equation (2.5). In this work, we have considered that this condition is satisfied only for a short time interval, i.e. \( T_1 \leq t \leq T_2 \), hence it is referred to as time localised SCA.

2.3. Blind detection of the active periods of a single source

In the simple model [Cheong Took et al. (2006)] due to the proximity of the stethoscope to the ipsi TMJ source, it was assumed that \( d=0 \), while the differential delay \( D=1 \) in \( H \) (1.2). However, the accuracy of this model relies on the exact adjustment of the sampling frequency to the size of the head. We relax these assumptions and consider:

\[
0 \leq d \leq 0.5 \quad 0.8 \leq D \leq 1.5 \tag{2.6}
\]

If only one source prevails in both mixtures during a time interval, MI between the two mixture signals for that segment is greater than those of the other segments. Since \( D > 0 \), we can deduce that the MLE function cannot exhibit a maximum at \( D=0 \). Otherwise, it is obvious that during this period (\( T_1 \leq t \leq T_2 \)) both sources exist. Hence, we can employ these criteria to blindly detect those regions where the sparsity prevails.

3. Simulations

We have considered the scenario where crepitus is present (due to osteoarthrosis). These TMJ sources (when measured separately) were mixed by a randomly generated \( H \):

\[
\begin{pmatrix}
0.8823z^{-0.17} & -0.2491z^{-1.51} \\
0.4708z^{-1.51} & 0.9685z^{-0.17}
\end{pmatrix}
\]

We employ the algorithm presented in [Cheong Took et al. (2006)], which requires the estimation of fractional delay \( D \) for source reconstruction. Thereafter, we compare Parra’s algorithm [Parra, L. & Spence, C. (2000)], and the time-frequency approach of Özgür and Rickard [Özgür, Y. & Rickard, S. (2004)] with our approach in terms of SIR (signal-to-interference ratio) as SNR varies in Figure 1.
4. Concluding Remarks

Figure 1 supports the superiority of our method in terms of higher SIR as a consequence of the inclusion of the delay estimates from the modified MLE and MMLETDE schemes. The much lower performance of Rickard’s algorithm is due to its reliance on k-means clustering (which in turn depends on good initialisation of $H$). Parra’s algorithm uses a long filter length to approximate linear convolution by circular convolution (owing to its frequency approach). This suggests that the filter length of its separating matrix is longer than what is required, which leads to an increase in interference. In summary, we have demonstrated how time localised SCA can assist in estimating the fractional differential delay $D$ from the TMJ mixtures, by employing modified versions of MLE and the MMLETDE estimators.

REFERENCES


