

3-10-2006

Blind separation of speech with a switched sparsity and temporal criteria

Daniel Smith
University of Wollongong

Ian Burnett
University of Wollongong, ianb@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/infopapers>



Part of the [Physical Sciences and Mathematics Commons](#)

Recommended Citation

Smith, Daniel and Burnett, Ian: Blind separation of speech with a switched sparsity and temporal criteria 2006.
<https://ro.uow.edu.au/infopapers/540>

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Blind separation of speech with a switched sparsity and temporal criteria

Abstract

A blind signal separation algorithm (SCAtemp) that exploits both the sparse time-frequency representation and temporal structure of speech is proposed. SCAtemp compares each speech signal's adherence to the sparsity and temporal criteria, before switching to the most appropriate criteria to estimate each signal. This algorithm is shown to improve the real time separation performance of conventional BSS algorithms exclusively exploiting either the temporal structure, sparsity or statistical independence of signals. The improvement of SCAtemp over conventional BSS algorithms can be attributed to the use of additional a priori knowledge of speech in the temporal short term.

Disciplines

Physical Sciences and Mathematics

Publication Details

This article was originally published as: Smith, D & Burnett, I, Blind separation of speech with a switched sparsity and temporal criteria, IEEE 8th Workshop on Multimedia Signal Processing 2006 (MMSP-06), Victoria, BC, Canada, 3-6 October 2006, 136-140. Copyright IEEE 2006.

Blind Separation of Speech with a Switched Sparsity and Temporal Criteria

Daniel Smith and Ian Burnett

School of Electrical, Computer and Telecommunication Engineering
University of Wollongong
Email: dvs02@uow.edu.au

Abstract—A Blind Signal Separation algorithm (SCAtemp) that exploits both the sparse time-frequency representation and temporal structure of speech is proposed. SCAtemp compares each speech signal's adherence to the sparsity and temporal criteria, before switching to the most appropriate criteria to estimate each signal. This algorithm is shown to improve the real time separation performance of conventional BSS algorithms exclusively exploiting either the temporal structure, sparsity or statistical independence of signals. The improvement of SCAtemp over conventional BSS algorithms can be attributed to the use of additional a priori knowledge of speech in the temporal short term.

I. INTRODUCTION

The application of Blind Signal Separation (BSS) to speech is of considerable interest in the research community. This interest is motivated by its use in developing adaptive, intelligent solutions to the 'cocktail party problem', a problem in which any speaker in an acoustic environment can be independently retrieved (or made the focus of listening attention) amidst other concurrent speakers and noise [1]. Consequently, BSS has the potential to improve the performance of voice technologies operating in noisy environments, such as teleconferencing and hands free mobile telephony [2].

These voice technologies require real time processing, as processing delays of 300ms or greater are considered intolerable in interactive two way communication [3]. In order for BSS algorithms to operate in conjunction with these interactive technologies, the algorithms must separate frames sized for real time application. Therefore, the purpose of this work is to improve real time separation of speech, by merging the temporal [1], [4], [5] and Sparse Component Analysis (SCA) [6]–[8] criteria of BSS into a single framework.

In order to separate speech mixtures, the algorithms in [1], [4], [5] exploit the temporal structure of speech and the SCA methods utilise the knowledge that the time-frequency (t-f) representation of speech are approximately disjoint orthogonal (non-overlapping) in the mixture [6]. Although the SCA and temporal approaches generally adhere to the properties of speech, there are occasions when short-time frames of speech fail to comply to the assumptions of these BSS algorithms. For instance, the t-f disjoint orthogonality assumption of SCA will be violated for short stationary sections of the mixture where formants of different speech signals overlap. Whilst, the temporal algorithms will be violated across short time frames

of mixed speech, which contain unvoiced or transient regions that do not possess temporal structure [3].

In order to improve the modeling of speech within BSS, a novel approach (SCAtemp) to combine the SCA and temporal criteria is proposed. The SCAtemp framework specifically exploits the sequential SCA algorithm proposed in [8] and a temporal algorithm that jointly models speech production mechanisms [5]. Estimates of each separation criteria are compared and a switching mechanism is employed to retrieve the signal estimate of the highest accuracy. Hence, it is hypothesised that the separation performance of the unified framework will become more robust, particularly across short frames of speech with weaker conformance to either the SCA or temporal criteria.

To demonstrate the improvement in data efficiency achieved by merging the sparse and temporal structure of speech into a single framework, the separation performance of SCAtemp is compared to its constituent SCA [8] and temporal [5] approaches, in addition to benchmark BSS algorithms exploiting Independent Component Analysis (ICA) [9], [10]. The analysis is conducted across mixtures of three moving speakers.

II. FORMULATION OF THE BSS PROBLEM

The BSS problem can be formulated as follows: The vector of sensor signals ($X(t)$) contain observations of the vector of signals ($S(t)$) linearly mixed according to the system A :

$$X(t) = A \cdot S(t) \quad (1)$$

where $X(t) = [x_1(t) \dots x_N(t)]^T$ is a $N \times 1$ vector of mixed observations, $S(t) = [s_1(t) \dots s_M(t)]^T$ is an unknown $M \times 1$ vector of signals and A is an unknown $N \times M$ non-singular matrix. In this approach it is assumed that A contains scalar elements (instantaneous mixing) and the system is square i.e. the number of signals is equal to the number of sensors ($M = N$).

Given only mixed observations $X(t)$, an $M \times N$ separation matrix W (estimating A^{-1}) must be computed and then multiplied by $X(t)$ in order to obtain a scaled, permutation of the original signals $c \cdot S(t)$. In contrast to the majority of BSS algorithms that simultaneously estimate the entire separation matrix, the method presented in this paper is a sequential approach. In a sequential approach, each column of

the separation matrix (W_{ie}) is individually estimated and then employed to separate a single signal. This process is repeated until all signals are separated.

III. ELEMENTS OF THE PROPOSED METHOD

The new SCAtemp architecture is comprised of two BSS approaches that are detailed in this Section. The SCAtemp algorithm compares these two BSS criteria, switching estimation to the criteria that best conforms to the underlying speech signals. The first criteria that is exploited by SCAtemp is a sequential approach to SCA estimation proposed in [8]. The second criteria used is a temporal approach [5] that employs a joint model of speech production mechanisms. The deflation process used to remove estimated signals from the mixture is then outlined.

A. Sequential SCA Algorithm

This sequential SCA algorithm proposed in [8] (SeqTIF) estimates the mixing column A_i of each signal using the Time Frequency Ratio Of Mixtures (TIFROM) approach that is detailed in [7]. The Short Time Fourier Transform (STFT) is computed across the mixture $X(t)$ to form the t-f representation $X(m, k)$, centered on the short time window m and frequency k . Ratios $\xi(\hat{m}, k) = [\xi_1(m, k) \dots \xi_{N-1}(m, k)]$ are then computed between the corresponding t-f windows of the 1st and j th mixed observation for $j = 2$ to N :

$$\begin{aligned} \xi_{j-1}(m, k) &= \frac{x_j(m, k)}{x_1(m, k)} \\ &= \frac{a_{j1}s_1(m, k) + \dots + a_{jM}s_M(m, k)}{a_{11}s_1(m, k) + \dots + a_{1M}s_M(m, k)} \end{aligned} \quad (2)$$

where a_{ji} are the elements of A . When $s_i(t)$ is the only signal present in a t-f window (m, k) , the ratios $A_{ie} = [1 \ \xi(\hat{m}, k)]^T$ correspond to the mixing column of the signal $s_i(t)$.

In order to estimate the mixing column of each signal, the following assumptions are necessary:

- 1) For each signal s_i (where $i \in \{1 \dots M\}$), there exists at least a series (Υ_u, k) of time-adjacent t-f windows of mixed observations where either s_i occurs alone or where the energy of s_i is far greater than all other signals. This series is considered the most ‘visible’ in the mixture.
- 2) Every t-f series (Υ_u, k) must always possess at least one signal.
- 3) Signals should be non stationary across the t-f windows (m, k) of the series (Υ_u, k) .

Under the condition of the second and third assumptions, when one signal is present across adjacent t-f windows (m, k) , the ratio $\xi(\hat{m}, k)$ is constant across these windows. However, under the third assumption of non stationary signals, if there is more than one signal present in the window set (m, k) , $\xi(\hat{m}, k)$ will vary across this set of windows.

As a consequence of this property, SeqTIF uses the variance across a series of TF windows to estimate mixing columns.

The mean $me(\Upsilon_u, k)$ and variance $var(\Upsilon_u, k)$ are computed across series (Υ_u, k) of time adjacent windows of the ratio $\xi(\hat{m}, k)$. The series with minimum $var(\Upsilon_u, k)_{min}$ is considered the most visible series (non-overlapping) in the mixture, and hence, selected as the mixing column estimate $A_{ie} = me(\Upsilon_u, k)$.

The estimate A_{ie} is then used to eliminate its corresponding signal $s_{ie}(t)$ from the mixture $X(t)$ as follows:

$$\begin{aligned} x_{mod,j-1}(t) &= x_1(t) - A_{ie} \cdot x_j(t) \\ &= (a_{11} \cdot s_1 + \dots + a_{1M} \cdot s_M) \\ &\quad - \frac{a_{1i}}{a_{ji}} \cdot (a_{j1} \cdot s_1 + \dots + a_{jM} \cdot s_M) \\ \text{where } j &= 2 \text{ to } N \text{ and } i \in \{1, \dots, M\} \end{aligned} \quad (3)$$

A mixing column is then estimated from the ‘cancelled’ mixture $X_{mod}(t)$, before the signal associated with this mixture estimate is again eliminated using (3). This process is repeated until only one (least visible) signal remains in the mixture, becoming the estimated signal $s_{re}(t)$.

B. Joint Model of Speech Production Mechanisms

BSS approaches, such as those presented in [1], have demonstrated that a speech signal can be extracted from a mixture by exploiting the following assumption:

- (a) *A single speaker has more temporal correlation than any linear combination of mixed speakers*

It is the temporal correlation generated by the production mechanisms of speech that make assumption (a) hold true [3]. The BSS approach developed in [5] (AR-F0 algorithm), exploits assumption (a), employing a more complete model of speech production mechanisms than existing temporal approaches. The AR-F0 algorithm jointly models the AR structure (short term correlation) and periodicity $\frac{1}{F_0}$ (long term correlation) of speech in the cost function $C(W_j, b_j, B_j)$:

$$\begin{aligned} C(W_j, b_j, B_j) &= 1/2 * E[\xi^2] \\ \xi(t) &= W_j^T \cdot X_a(t) - B_j \cdot W_j^T \cdot X_l(t) \end{aligned} \quad (4)$$

where b_j is the prediction filter, B_j is the prediction gain and $\hat{X}(t) = [X(t-1) \dots X(t-P)]$. The term $X_a(t) = X(t) - \hat{X}(t) \cdot b_j^T$ is the short term temporal prediction error of the mixtures and $X_l(t) = X(t - \frac{1}{F_0}) - \hat{X}(t - \frac{1}{F_0}) \cdot b_j^T$ is the short term, period delayed prediction error of the mixtures. Therefore, the error function $\xi(t)$ jointly describes the short and long term temporal prediction error of the estimated speech. The first term in $\xi(t)$ (containing $X_a(t)$) represents the short term prediction model and the second term (containing $X_l(t)$) represents the long term prediction model.

This is a sequential approach to separation, using gradient adaptation of the parameters W_j, b_j and B_j in order to estimate a signal $s_{re}(t)$ from the mixture. The learning rules of the parameters are derived and then shown in [5].

TABLE I

THIS TABLE SHOWS THE RESULT OF AN EMPIRICAL STUDY CONDUCTED TO DETERMINE EFFECT THAT THE THRESHOLD VALUE c_{comp} HAS ON SEPARATION PERFORMANCE. THE SCAtemp ALGORITHM IS APPLIED TO A SET OF 20 STATIONARY MIXTURES AS c_{comp} IS VARIED BETWEEN 0.004 AND 0.4. THE SNR PERFORMANCE (IN DECIBELS (dB)) IS SHOWN FOR A SUBSET OF c_{comp} VALUES FOR ANALYSIS BLOCKS SPANNING IN SIZE FROM 70MS TO 0.56S.

c_{comp}	Size of Analysis Block (s)								
	0.07	0.08	0.09	0.11	0.16	0.26	0.36	0.46	0.56
0.004	12.56	13.33	14.72	16.70	20.94	25.70	27.93	29.29	30.04
0.02	17.86	18.42	19.73	21.04	23.70	26.67	28.52	29.63	30.17
0.04	18.19	19.07	20.04	21.39	24.06	26.76	28.52	29.63	30.17
0.06	17.37	17.98	19.40	20.72	23.60	26.67	28.52	29.63	30.17
0.4	16.56	17.31	18.74	20.09	22.99	26.12	28.10	29.44	30.08

C. Deflation Technique

In a sequential approach to estimation, once a signal has been estimated, it is necessary to permanently remove it from the original mixture $X(t)$. The SeqTIF and AR-F0 algorithms can only estimate $s_{re}(t)$ up to an undetermined scaling factor of it's original signal. Thus, a scaling factor g is required to remove $s_{re}(t)$ from $X(t)_v$ as follows:

$$X(t)_{v+1} = X(t)_v - g \cdot s_{re}(t) \quad (5)$$

In order to deflate $s_{re}(t)$ from the mixture, the optimal solution of g is obtained as the Minimum Mean Squared Error (MMSE) of $X(t)_{v+1}$.

IV. THE PROPOSED UNIFIED METHOD

The new SCAtemp algorithm initially employs the SeqTIF algorithm (outlined in Section III-A) and only switches to the temporal AR-F0 algorithm (outlined in Section III-B), if the performance of SeqTIF is considered inadequate. This initial application of SeqTIF can be justified as it provides a more accurate measure of the estimation quality (A_e) than the AR-F0 algorithm. Under the assumptions of Section III-A, the $var(\Upsilon_u, k)$ of SeqTIF can be used to directly measure the level of signal overlap in each t-f series, and hence, the accuracy of mixing column estimates within each t-f series. The $var(\Upsilon_u, k)$ is a true measure of the estimation quality with $var(\Upsilon_u, k) = 0$ corresponding to perfect estimation of a mixing column.

The AR-F0 algorithm uses the $C(W_j, b_j, B_j)$ from (4) to measure the estimation quality. In contrast to $var(\Upsilon_u, k)$, $C(W_j, b_j, B_j)$ is not a true measure of separation performance. $C(W_j, b_j, B_j)$ is the error associated with modeling a mixture of speech signals using (4) and does not necessarily indicate the level of separation performance. For instance, it was revealed in [11], that the modeling error of voiced speech mixtures can still be small without achieving any form of separation. However, the modeling error does usually provide some indication of separation quality. In particular, a large $C(W_j, b_j, B_j)$ corresponds to poor conformance to the temporal model in (4), and hence, poor quality speech estimates.

A. Comparing Separation Criteria

SCAtemp operates by comparing $var(\Upsilon_u, k)_{min}$ of the SeqTIF estimate to a pre defined threshold c_{min} . The threshold, c_{min} , is an upper bound of variance, below which the quality of SeqTIF estimation is considered high enough to avoid AR-F0 estimation. However, when $var(\Upsilon_u, k)_{min} > c_{min}$, it is considered that the separation performance of SCAtemp may be substantially improved by incorporating AR-F0 estimation. In this case, it is unknown whether the SeqTIF or AR-F0 estimate provides superior separation. Consequently, the $var(\Upsilon_u, k)_{min}$ of the SeqTIF estimate and $C(W_j, b_j, B_j)_{min}$ of the AR-F0 estimate are used to compare separation performance. As $C(W_j, b_j, B_j)_{min}$ is generally larger than $var(\Upsilon_u, k)_{min}$, another heuristic c_{comp} is required to provide a fair comparison between the $var(\Upsilon_u, k)_{min}$ and $c_{comp} \cdot C(W_j, b_j, B_j)_{min}$ terms. This heuristic is chosen a priori, and must ensure that SCAtemp consistently selects the best quality estimate.

In order to determine a suitable c_{comp} threshold in this experiment, an empirical study was conducted across a training set of 20 stationary mixtures, each consisting of three different speech signals. The size of the analysis blocks spanned from 70ms to 0.56s. The value of c_{comp} was varied between 0.004 and 0.4 to determine which value most consistently selected the best estimate between the SeqTIF and AR-F0 algorithms. A subset of these results demonstrating the performance trend are shown in Table I. Table I shows that a c_{comp} value of 0.04 had the highest estimation quality, as it had exhibited a maximum SNR (defined in (6)). The separation performance degraded as the value of c_{comp} varied in either direction from 0.04. As c_{comp} decreased in value from 0.04, SCAtemp was more likely to employ SeqTIF estimates at the expense of AR-F0 estimates that were often of a higher quality. When the value of c_{comp} became larger than 0.04, the reverse occurred.

The SCAtemp algorithm operates by switching between the SCA and temporal criteria, as shown in the flow diagram in Fig. 1.

V. EXPERIMENTAL RESULTS AND DISCUSSION

An experiment was conducted to compare the separation performance of the SCAtemp algorithm to the SeqTIF and

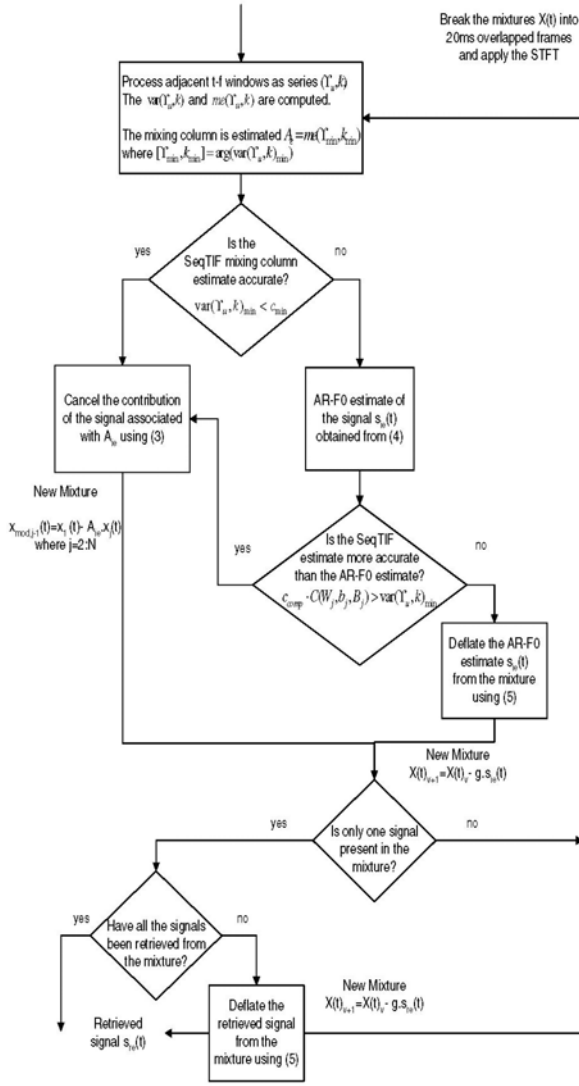


Fig. 1. The SCAtemp algorithm is structured to switch between SeqTIF and AR-F0 estimation. The switching is based upon a comparison of each criteria's estimation quality, the variance of SeqTIF estimates and MMSE of AR-F0 estimates, such that the highest quality estimate is chosen.

AR-F0 algorithms of which it consists, in addition to two benchmark BSS algorithms, Extended Infomax [10] and FastICA [9]. The experiment was conducted on 10 time varying mixtures, each consisting of three different speech signals that were 5s in length and sampled at 8000Hz. The mixing system was generated under the assumption that the sound pressure of each audio signal varied inversely with the distance to each sensor. This mixing system was updated every 125ms and produced an average SNR of 1.27dB between the mixtures $X(t)$ and original signals $s_i(t)$. The physical path of the speakers are shown in Fig. 2.

The performance of the algorithms were compared across analysis blocks of the mixture ranging in size from 70ms to 0.56s. The SeqTIF approach further processed these analysis blocks by breaking them into short time windows of 20ms with

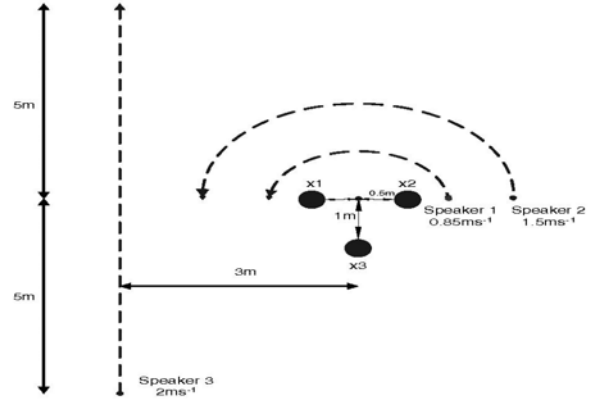


Fig. 2. The physical path of the acoustic environment in which the mixing system was generated. The first two speakers moved in a circular path at constant velocities of $0.85m/s$ and $1.5m/s$. The third speaker moved in a straight line at a constant velocity of $2m/s$. The microphones were located at x_1, x_2 and x_3 .

50% overlap and then forming series (Y_u, k) that consist of 6 time adjacent windows.

The criteria used to measure the separation performance was the Signal to Noise Ratio (SNR). It is calculated between the estimated signal $s_e(t)$ and the corresponding original signal $s_i(t)$:

$$SNR = 10 \cdot \log_{10} \frac{\sigma^2}{E\{(s_i(t) - s_e(t))^2\}} \text{ dB} \quad (6)$$

where σ^2 is the variance of $s_i(t)$. When calculating the SNR , both $s_e(t)$ and $s_i(n)$ are normalized to a variance of 1, in order to avoid the scaling ambiguity of $s_e(n)$. Thus, the SNR is simplified to $-10 \cdot \log_{10} E\{(s_i(t) - s_e(t))^2\}$, such that a higher SNR corresponds to a smaller estimation error of $s_e(t)$, and hence, better separation performance.

A. A Comparison of the Unified and Single Criteria Algorithms

Figure 3 compares the average SNR of the unified SCAtemp algorithm, its constituent SeqTIF and AR-F0 algorithms, in addition to benchmark FastICA and Extended Infomax algorithms in the analysis. All algorithms exhibit a maximum SNR at a block size of 100ms, as this is the longest analysis block that can estimate a constant mixing system. As block sizes become longer than 125ms, estimation performance will decrease as the blocks comprise of a change in the mixing system.

It is evident from Fig. 3 that the separation performance of the SCAtemp algorithm is superior to the algorithms which exploit either the sparse t-f representation (SeqTIF algorithm) or temporal structure of speech (AR-F0 algorithm). In particular, the SCAtemp algorithm offers a substantial separation advantage over the AR-F0 algorithm across all data sizes, with the average SNR advantage ranging from 5.3dB to 7.8dB. The advantage of the SCAtemp algorithm increases across longer block sizes, as the level of conformance to the AR-F0 model in (4) declines across non stationary speech.

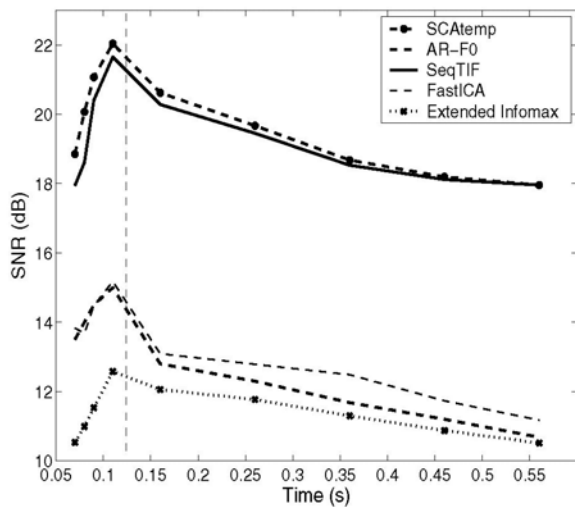


Fig. 3. The separation performance (SNR) of the SCAtemp algorithm is compared to its constituent SeqTIF and AR-F0 algorithms, along with benchmark ICA (FastICA and Extended Infomax) algorithms for analysis blocks that are sized from 70ms to 0.56s. The experimental set consisted of 10 mixtures each consisting of three different speech signals. The mixtures changed every 125ms, as shown by the dotted vertical line.

The SCAtemp algorithm exhibits a significantly smaller separation advantage over SeqTIF than the AR-F0 algorithm, with the average SNR advantage ranging from $0dB$ to $0.91dB$ (maximum 5% advantage). This smaller advantage can be attributed to the superior quality of SeqTIF estimates over AR-F0 estimates, particularly across longer block sizes. As blocks of speech become longer, more visible t-f series are obtained, improving SeqTIF estimation. Simultaneously, the quality of AR-F0 estimation is reduced by the lack of conformance to the temporal model, mentioned in the previous paragraph. Therefore, as analysis blocks lengthen, the SCAtemp algorithm exploits SeqTIF estimates more frequently. Most notably, the SNR of the algorithms converge at a block size of 0.56s, indicating SCAtemp only uses SeqTIF estimates at this block size. As block sizes are reduced towards real time operation, speech exhibits a higher level of compliance to the AR-F0 model. In this case, the SCAtemp algorithm exploits AR-F0 estimates with greater regularity (particularly in regions without visible series) to improve upon SeqTIF estimation.

B. A Comparison of the Unified and ICA Algorithms

Fig. 3 indicates that the separation performance of the unified SCAtemp approach is far superior to the FastICA and Extended Infomax algorithms across all block sizes. The SCAtemp algorithm achieves an average SNR advantage of between $7.85dB$ and $9.54dB$ over the Extended Infomax algorithm and an average SNR that is between $5.17dB$ and $6.97dB$ larger than FastICA.

The SeqTIF demonstrates superior separation performance to both ICA algorithms in this analysis. The AR-F0 algorithm has a separation advantage over the Extended Infomax algorithm, however, only a similar level of performance to FastICA

across block sizes less than 160ms. The FastICA algorithm shows a clear separation advantage (up to $0.89dB$) over the AR-F0 algorithm for block sizes greater than 160ms, due to the degradation of AR-F0 estimation across non stationary speech.

VI. CONCLUSIONS AND FUTURE WORK

A unified approach to BSS is proposed (SCAtemp), combining the sparse t-f representation and temporal structure of speech into a switched framework. This algorithm offers a superior data efficiency in comparison with constituent algorithms, which exploit either the temporal structure or SCA criteria, exclusively. Furthermore, it exhibits a separation advantage over two benchmark ICA algorithms. These results suggest that modeling additional a priori knowledge of speech, improves separation performance, particularly in the realm of real time operation. In addition, the improved data efficiency of SCAtemp enhances the tracking of time varying mixtures.

A limitation of this algorithm is the heuristic approach used to select between the temporal and SCA estimates. Although training of c_{comp} greatly reduces the likelihood of estimates being incorrectly selected, there is still the possibility of the weaker estimate being employed in SCAtemp. An avenue of future work will involve improving the robustness of the selection criteria. Furthermore, the SCA and temporal approaches will be incorporated into a unified framework to accommodate reverberation within the environment (convolutive mixtures).

REFERENCES

- [1] A. Cichoki and S. Amari, *Adaptive Blind Signal and Image Processing : Learning Algorithms and Applications*. John Wiley & Sons, 2002.
- [2] K. Torkola, "Blind Separation for Audio Signals - Are we there yet?" in *Proc. Workshop on Independent Component Analysis (ICA99)*, vol. 2, 1999, pp. 239-244.
- [3] W. Kleijn and K. Paliwal, *Speech Coding and Synthesis*, W. Kleijn and K. Paliwal, Eds. Elsevier Science, 1995.
- [4] N. Murata, S. Ikeda, and A. Ziehe, "An Approach to Blind Source Separation Based on the Temporal Structure of Speech Signals," *Neurocomputing*, vol. 41, pp. 1-24, October 2001.
- [5] D. Smith, J. Lukasiak, and I. Burnett, "Blind Speech Separation Using a Joint Model of Speech Production," *IEEE Signal Processing Letters*, vol. 12, no. 11, pp. 784-787, November 2005.
- [6] O. Yilmaz and S. Rickard, "Blind Separation of Speech Mixtures via Time-Frequency Masking," *Proc. IEEE Transaction on Signal Processing*, vol. 52, no. 7, pp. 1830-1847, 2004.
- [7] F. Abrard, Y. Deville, and P. White, "From Blind Source Separation to Blind Source Cancellation in the Undetermined Case: A New Approach based on Time-Frequency Analysis," in *Proc. International Conference on Independent Component Analysis (ICA2001)*, San Diego, December 2001, pp. 734-739.
- [8] D. Smith, J. Lukasiak, and I. Burnett, "A Sequential Approach to Sparse Component Analysis," in *The 7th IEEE International Workshop on Multimedia Signal Processing (MMSP05)*, Shanghai, October 2005, pp. 129-132.
- [9] A. Hyvarinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis," *Neural Computation*, vol. 9, pp. 1482-1492, 1997.
- [10] T. Lee and T. Sejnowski, "Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources," *Neural Computation*, vol. 11, no. 2, pp. 417-441, 1999.
- [11] D. Smith, J. Lukasiak, and I. Burnett, "An Investigation of Temporal Modeling in Blind Signal Separation," in *The 8th IEEE International Symposium on Signal Processing and its Applications (ISSPA05)*, vol. 2, Sydney, August 2005, pp. 503-506.