Application of Electrode Arrays for Artifact Removal in an Electromyographic Silent Speech Interface

Michael Wand, Matthias Janke, Till Heistermann, Christopher Schulte, Adam Himmelsbach, and Tanja Schultz

Cognitive Systems Lab, Karlsruhe Institute of Technology
Adenauerring 4, 76137 Karlsruhe, Germany
{michael.wand,matthias.janke,tanja.schultz}@kit.edu
{till.heistermann,christopher.schulte,adam.himmelsbach}@student.kit.edu
http://csl.anthropomatik.kit.edu

Abstract. An electromygraphic (EMG) *Silent Speech Interface* is a system which recognizes speech by capturing the electric potentials of the human articulatory muscles, thus enabling the user to communicate silently. This study deals with the introduction of multi-channel electrode arrays to the EMG recording system, which requires meticulous dealing with the resulting high-dimensional data. As a first application of the technology, Independent Component Analysis (ICA) is applied for automated artifact detection and removal. Without the artifact removal component, the system achieves optimal average Word Error Rates of 40.1% for 40 training sentences and 10.9% for 160 training sentences on EMG signals of audible speech. On a subset of the corpus, we evaluate the ICA artifact removal method, improving the Word Error Rate by 10.7% relative.

Keywords: Electromyography, EMG, Silent Speech Interface, Electrode Array, EMG-based Speech Recognition

1 Introduction

Speech is the most convenient and natural way of human communication and interaction. However all but 150 years ago, the lack of technical devices for amplifying, processing, transmitting, or storing acoustic signals limited spoken communication to face-to-face talk or speeches in front of at most medium-sized audiences.

Nowadays, mobile phone technology has made speech a wide-range, ubiquitous means of communication, and talking to any person, worldwide, has become a reality. Furthermore, many speech-controlled devices and services have been developed, including telephone-based information retrieval systems, voice-operated car navigation systems, and large-vocabulary dictation and translation assistants.

Unfortunately, voice-based communication and interaction suffers from several challenges which arise from the fact that the speech needs to be clearly audible and cannot be masked, including lack of robustness in noisy environments, disturbance for bystanders, privacy issues, and exclusion of speech-disabled people. These challenges may be alleviated by Silent Speech Interfaces, which are systems enabling speech communication to take place without the necessity of emitting an audible acoustic signal, or when an acoustic signal is unavailable [1].

Over the past few years, we have developed a Silent Speech Interface based on surface electromyography (EMG): When a muscle fiber contracts, small electrical currents in form of ion flows are generated. EMG electrodes attached to the subject's face capture the potential differences arising from these ion flows, which are then used to retrace the corresponding speech. This allows speech to be recognized even when it is produced silently, i. e. mouthed without any vocal effort.

Many EMG-based speech recognizers rely on small sets of less than 10 EMG electrodes attached to the speaker's face [2–6]. This setup is based on standard Ag-AgCl gelled electrodes as used in medical applications and imposes some limitations, for example, small shifts in the electrode positioning between recordings are difficult to compensate, and it is impossible to separate superimposed signal sources, thus single active muscles or motor units cannot be discriminated.

In this study we present experiments on using *electrode arrays* for the recording of EMG signals of speech. In a first step, we establish that our existing EMG-based continuous speech recognizer [2] is able to deal with the increased number of signal channels. This requires extending standard Linear Discriminant Analysis with a prior Principal Component Analysis step. Secondly, we present a first application of the EMG array methodology, namely, we show that Independent Component Analysis (ICA) can be used to remove artifacts from the multi-channel EMG signal.

This article is organized as follows: In the following section 2, we describe our new recording system, and section 3 contains a description of the underlying recognizer. Section 4 establishes the multi-channel recording system and presents our experiments on artifact detection and removal. The final section 5 concludes the article.

2 Recording System Setup and Corpus

EMG signals were recorded with the multi-channel biosignal amplifier *EMG-USB2*, which is produced and distributed by *OT Bioelettronica*, Torino, Italy (`http://www.otbioelettronica.it`). The EMG-USB2 device allows to record and process up to 256 EMG channels, supporting a selectable gain of 100 - 10.000 V/V and a recording bandwidth of 3 Hz - 4400 Hz. For line interference reduction, we used the integrated DRL circuit [7]. The electrode arrays were acquired from *OT Bioelettronica* as well. Electrolyte cream was applied to the EMG arrays in order to reduce the electrode/skin impedance.

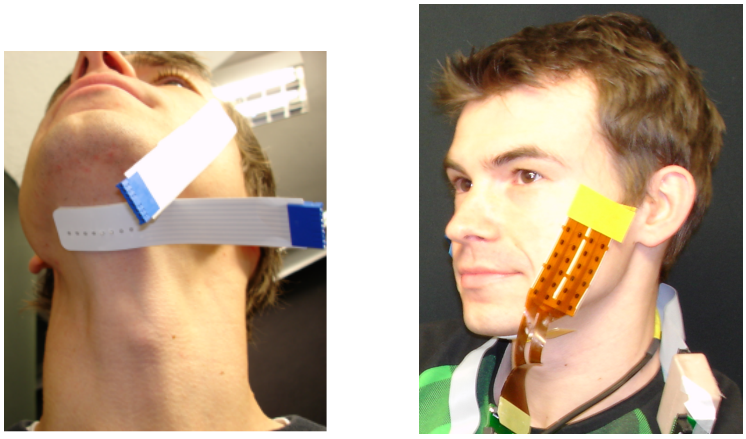

Fig. 1. EMG array positioning for setup A (left) and setup B (right).

We used two different EMG array configurations for our experiments, see figure 1. In *setup A*, we unipolarly recorded 16 EMG channels with two EMG arrays each featuring a single row of 8 electrodes, with 5 mm inter-electrode distance (IED). One of the arrays was attached to the subject's cheek, capturing several major articulatory muscles [3], the other one was attached to the subject's chin, in particular recording signals from the tongue. A reference electrode was placed on the subject's neck.

In *setup B*, we replaced the cheek array with a larger array containing four rows of 8 electrodes, with 10 mm IED. The chin array remained in its place. In this setup, we achieved a cleaner signal by using a *bipolar* configuration, where the potential difference between two adjacent channels in a row is measured. This means that out of $4 \cdot 8$ cheek electrodes and 8 chin electrodes, we obtained $(4+1) \cdot 7 = 35$ signal channels. For both setups, we chose an amplification factor of 1000, a high-pass filter with a cutoff frequency of 3 Hz and a low-pass filter with a cutoff frequency of 900 Hz, and a sampling frequency of 2048 Hz. The audio signal was parallely recorded with a standard close-talking microphone. We used an analog marker system to synchronize the EMG and audio recordings, and according to [8], we delayed the EMG signal by 50ms compared to the audio signal.

The text corpus which we recorded is based on [2]. We used two different text corpora for our recordings: Each session contains a set of ten "BASE" sentences which is used for testing and kept fixed across sessions. Furthermore, each session contains 40 training sentences, which vary across sessions. For reference, we call this basic text corpus "Set 1". A subset of our sessions has been extended to 160 different training sentences and 20 test sentences, where the 20 test sentences consist of the BASE set repeated twice. This enlarged text corpus is called "Set 2".

The recording proceeded as follows: In a quiet room, the speaker read English sentences in normal, audible speech. Note that we also recorded silently mouthed speech, which was not used in this study. The set of sentences was taken from the Broadcast News Domain. The recording was supervised by a member of the research team in order to detect errors (e. g. detached electrodes) and to assure a consistent pronunciation. The training and test sentences were always recorded in randomized order. Thus we finally have four setups to investigate, namely, setups A-1 and A-2 (with 16 EMG channels) and B-1 and B-2 (with 35 EMG channels). The suffixes "1" and "2" refer to the recorded text corpus, e. g. the A-1 sessions consist of 40 + 10 sentences, the A-2 sessions consist of 160 + 20 sentences, etc. At this point we remark that the results on the four setups are not directly comparable, since the number of training sentences, the set of speakers and the number of sessions per speaker differ. Also, our experience indicates that even for one single speaker, the recognition performance may vary drastically between sessions, possibly due to variations in electrode positioning, skin properties, etc. However, it is certainly plausible to compare the effects of different feature extraction methods on the recognition performance of *each* of the setups. Note that the test sets of the four setups consist of identical sentence lists, so their characteristics in terms of perplexity and vocabulary are the same.

The following table summarizes the properties of our corpus.

Setup	# of Speakers / Sessions	Average data length in sec.		
		Training	Test	Total
A-1	3 / 6	144	37	181
A-2	2 / 2	528	74	602
B-1	6 / 7	149	42	191
B-2	4 / 4	570	83	653

3 Feature Extraction, Training and Decoding

The feature extraction is based on *time-domain features* [8]. We first split the incoming EMG signal channels into a high-frequency and a low-frequency part, after this, we perform framing and compute the features. The whole feature extraction proceeds as follows.

For any given feature $\mathbf{f}$, $\bar{\mathbf{f}}$ is its frame-based time-domain mean, $\mathbf{P_f}$ is its frame-based power, and $\mathbf{z_f}$ is its frame-based zero-crossing rate. $S(\mathbf{f}, n)$ is the stacking of adjacent frames of feature $\mathbf{f}$ in the size of $2n + 1$ ($-n$ to n) frames.

For an EMG signal with normalized mean $x[n]$, the nine-point double-averaged signal $w[n]$ is defined as

$$w[n] = \frac{1}{9} \sum_{k=-4}^{4} v[n+k], \quad \text{where} \quad v[n] = \frac{1}{9} \sum_{k=-4}^{4} x[n+k].$$

The high-frequency signal is $p[n] = x[n] - w[n]$, and the rectified high-frequency signal is $r[n] = |p[n]|$. Mathematically, the computation of $w[n]$ from $x[n]$ is an application of a weighted moving average filter.

The final feature $\mathbf{TD}n$ is defined as follows:

$$\mathbf{TD}n = S(\mathbf{TD}0, n), \text{where } \mathbf{TD}0 = [\bar{\mathbf{w}}, \mathbf{P_w}, \mathbf{P_r}, \mathbf{z_p}, \bar{\mathbf{r}}]. \qquad (※)$$

Frame size and frame shift are set to 27 ms respective 10 ms.

The $\mathbf{TD}0$ feature is the most basic feature used in this study, consisting of a channel-wise combination of five standard time-domain features, which are computed frame by frame. In the final step defined by equation (※), a stacking of adjacent feature vectors with context width $2 \cdot n + 1$ is performed, with varying n. This process is performed for each channel, and the combination of all channel-wise feature vectors yields the final $\mathbf{TD}n$ feature vector. For the stacking context widths, different n have been used in prior work, e. g. [8] uses a context width of 5, however on a different corpus, [9] shows that increasing the context width to 15 frames, i. e. 150 ms per side, yields improved results.

In all cases, we apply Linear Discriminant Analysis (LDA) on the $\mathbf{TD}n$ feature. The LDA matrix is computed by dividing the training data into 136 classes corresponding to the begin, middle, and end parts of 45 English phonemes, plus one silence phoneme. From the 135 discriminant dimensions which are yielded by the LDA algorithm, we always retain 32 dimensions. As shown in section 4.2, it may be necessary to perform Principal Component Analysis (PCA) before computing the LDA matrix, see section 4.2 for further details. In the experiments described in section 4.3, Independent Component Analysis (ICA) and possible artifact suppression is applied *before* the feature extraction step, on the raw EMG data.

The recognizer is based on three-state left-to-right fully continuous Hidden-Markov-Models. All experiments use bundled phonetic features (BDPFs) [2] for training and decoding. In order to obtain phonetic time-alignments as a reference for training, the parallely recorded acoustic signal is forced-aligned with an English Broadcast News (BN) speech recognizer. Based on these time-alignments, the HMM states are initialized by a merge-and-split training step [10], followed by four iterations of Viterbi training.

For decoding, we use the trained acoustic model together with a trigram Broadcast News language model. The test set perplexity is 24.24. The decoding vocabulary is restricted to the words appearing in the test set, which results in a test vocabulary of 108 words. Note that we do *not* use lattice rescoring for our experiments.

Further information can be found in [2], the recognizer presented therein serves as the baseline for this study.

4 Experiments and Results

In this section we first present results on applying our baseline system towards the new, high-dimensional signals. Subsequently, we present our ICA-based automated artifact removal method, and evaluate its performance.

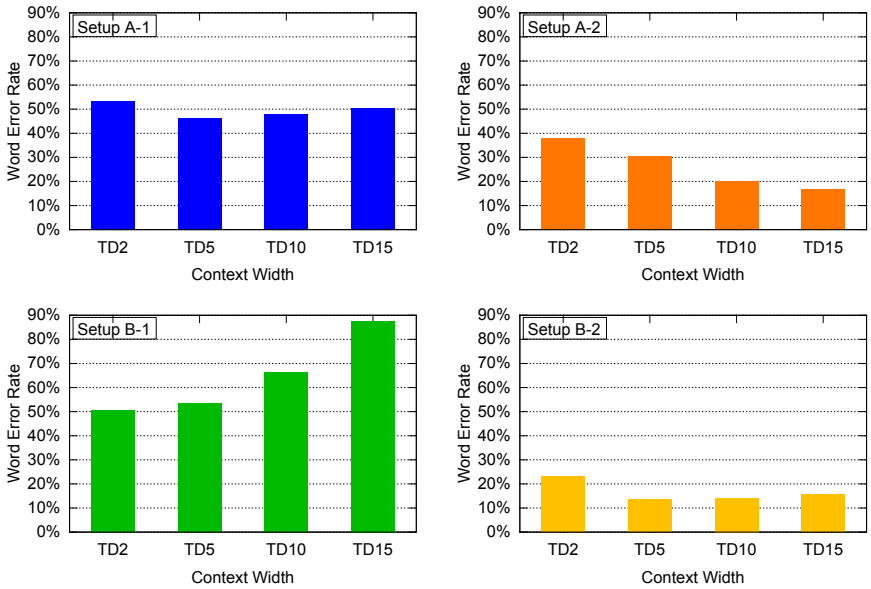

Fig. 2. Word Error Rates for the baseline system with different stacking context widths. PCA or ICA were not applied.

4.1 Baseline Recognition System

In the first experiment, we use our baseline recognizer, as described in section 3, and feed it with the EMG features from the array recording system. Figure 2 shows the Word Error Rates for different stacking widths, averaged over all sessions of each setup.

We consider the optimal context widths for the four setups. Our observations for the four distinct setups are very different: For setup A-1, with 16 channels and 40 training sentences, the Word Error Rate (WER) varies between 46.3% and 53.2%, with the optimum reached at a context width of 5 (i. e. **TD**5). For the B-1 setup, with 35 channels but the same amount of training data, the optimal context width appears to be **TD**2 with a WER of 50.5%, widening the context causes deteriorating results, the worst WER is 87.6% for the **TD**15 stacking.

For the setups with 160 training sentences, the recognition performance is consistently better due to the increased training data amount. With respect to context widths, we observe a behavior which vastly differs from the results above: For 16 EMG channels (setup A-2), the optimal context width is 15, with a WER of only 13.2%. For setup B-2, **TD**5 stacking is optimal, with a WER of 10.6%.

The behavior described in this section is quite consistent across recording sessions. The variation in optimal stacking width indicates a deep inconsistency between our setups, which leads us to the series of experiments described in the following section.

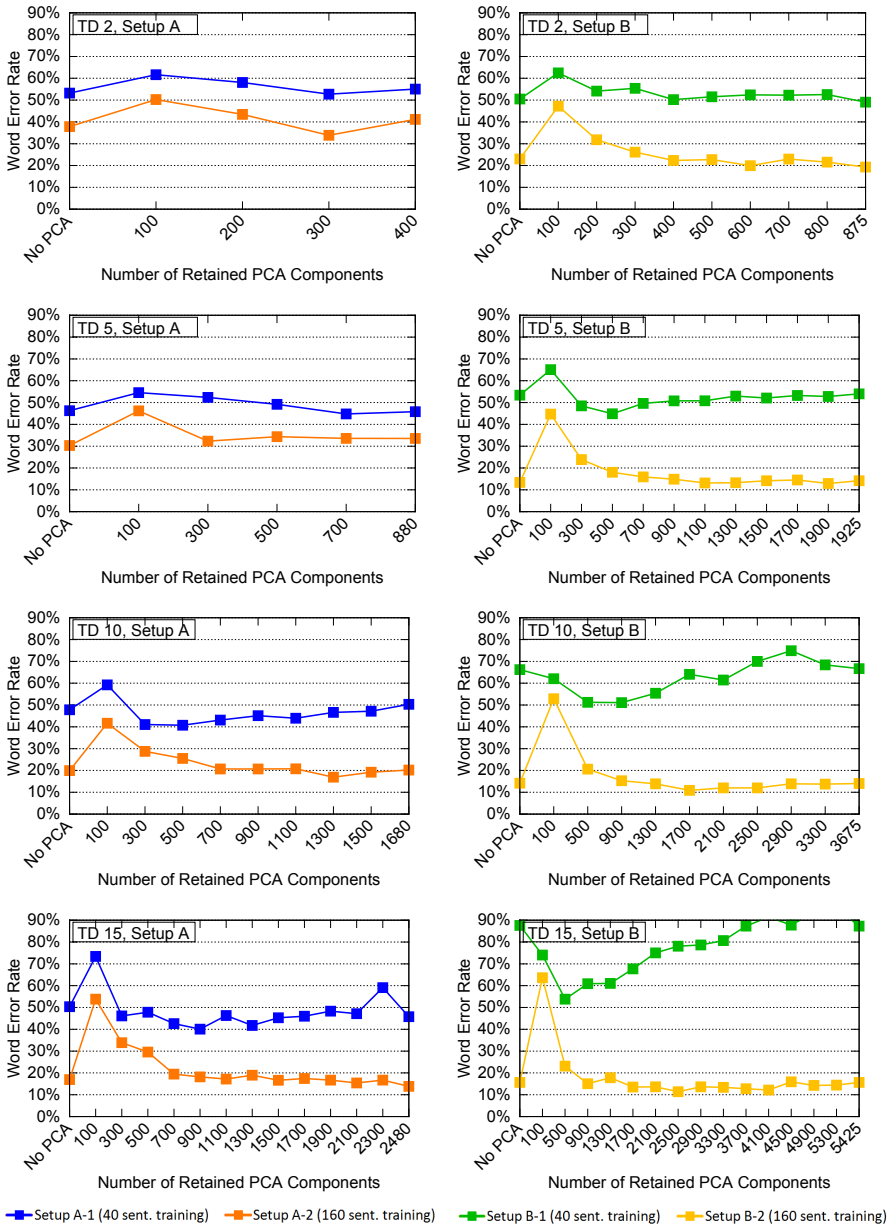

Fig. 3. Word Error Rates for different PCA dimension reductions. Observe that the feature space dimension *before* the PCA step increases from left to right and from top to bottom.

4.2 PCA Preprocessing to Avoid LDA Sparsity

Machine learning tasks frequently exhibit a challenge known as the "Curse of Dimensionality", which means that high-dimensional input data, relative to the

amount of training data, causes undertraining, diminishes the effectiveness of machine learning algorithms, and reduces in particular the generalization capability of the generated models. The maximal feature space dimension which allows robust training depends on the amount of available training data.

The dimensionality of the feature space in our experiments depends on the number of EMG channels and the stacking width in the feature extraction. From the results of section 4.1, we observe

- that for both setups A and B, increasing the amount of training data increases the optimal context width
- and that for both the 40-sentence training corpus (set 1) and the 160-sentence training corpus (set 2), the optimal context width with setup B is lower than the optimum for setup A.

This strongly suggests that the "Curse of Dimensionality" is the cause of the discrepancy we observed. However, since the LDA algorithm *always* reduces the feature space dimensionality to 32 channels, the GMM training itself is not affected by varying feature dimensionalities.

We propose that the deterioration of recognition accuracy for small amounts of training data and high feature space dimensionalities is caused by the LDA computation step. It has been shown that when the amount of training data is small relative to the sample dimensionality, the LDA within-scatter matrix becomes sparse, which reduces the effectivity of the LDA algorithm [11]. This may be the case in our setup, since with only a few minutes of training data, we may have a sample dimensionality before LDA of up to $35 \cdot 5 \cdot 31 = 5425$ for the 35-channel system with a **TD**15 stacking.

The following set of experiments deals with coping with the LDA sparsity problem. We expect an improved recognition accuracy and, in particular, a more consistent result regarding the optimal feature stacking width. Our method is to apply an additional PCA dimension reduction step before the LDA computation, as advocated for visual face recognition [12]. This step should allow an improved LDA estimation, however, if the PCA cutoff dimension is chosen too low, one will lose information which is important for discrimination.

The computation works as follows: On the training data set, we first compute a PCA transformation matrix. We apply PCA and keep a certain number of components from the resulting transformed signal, where the components are, as usual, sorted by decreasing variance. Then we compute an LDA matrix of the PCA-transformed training data set, finally keeping 32 dimensions. The resulting PCA + LDA preprocessing is now applied to the entire corpus, normal HMM training and testing is performed, and we use the Word Error Rate as a measure for the quality of our preprocessing.

Figure 3 plots the Word Error Rates of the recognizer for setups A and B and different stacking widths versus the number of retained dimensions after the PCA step. In all cases, we jointly plot the WERs for training data sets 1 and 2.

The figures show that the PCA step indeed helps to overcome LDA sparsity. For example, in the A-1 setup, the optimal context width without PCA application is 5, yielding a WER of 46.3%. With PCA application, the optimal number

Setup	A-1	A-2	B-1	B-2
Best Result without PCA ("Baseline")	46.3%	17.0%	50.5%	13.4%
Optimal Stacking Width without PCA	5	15	2	5
Optimal Number of Dimensions without PCA	880	2480	875	1925
Best Result with PCA	40.1%	13.9%	44.9%	10.9%
Optimal Stacking Width with PCA	15	15	5	10
Optimal Number of Dimensions with PCA	900	2480	500	1500
Relative Improvement by PCA Application	13.4%	18.2%	11.1%	18.7%

Table 1. Optimal Results and Parameters with and without PCA

of retained PCA dimensions for the **TD**5 context width is 700, yielding a WER of 44.8%. However, we can still do better: With a vastly increased context width of 15, we get the best WER of 40.1%, at a dimensionality of 900 after PCA application.

This is also true for the other four setups, see table 1 for an overview. In all cases, we obtain WER reductions of more than 10% relative, and also, in all cases the optimal context width increases.

So far, we have found the optimal context width for the EMG speech classification task to lie around 10 to 15 frames on each side, which makes a context of around 200-300 ms. It may be possible to try even wider contexts, however, close examination of the results in figure 3 show that between the context widths of 10 and 15, the respective WERs with optimal PCA dimensionality are rather close for each of the four setups, so it may be assumed that increasing the context width beyond around 15 frames will not yield further improvements.

4.3 ICA Application

Having established a baseline recognizer for array recordings of EMG, we now consider applications of this technology. One well-established means of identifying signal sources in multi-channel signals is *Independent Component Analysis (ICA)* [13]. ICA is a linear transformation which is used to obtain independent components within a multi-channel signal; the underlying idea is that the statistical independence between the estimated components is maximized. We use the Infomax ICA algorithm according to [14], as implemented in the Matlab EEGLAB toolbox [15].

In this study, we apply ICA for *artifact removal*. For this purpose, we interpret ICA as a method of (blind) source separation: We run ICA on the raw EMG signal, obtaining a set of 16 resp. 35 ICA components. The ICA decomposition matrix is computed separately for the two arrays. We then develop a heuristical measure to determine whether any detected component should be considered an artifact or a "target" EMG signal component. This method has been presented in detail in [16]. For our experiments, we choose the B-1 setup with 40 training sentences.

As an example, figure 4 shows a short extract of EMG signals recorded with the chin array, together with their ICA decomposition. The example shows that

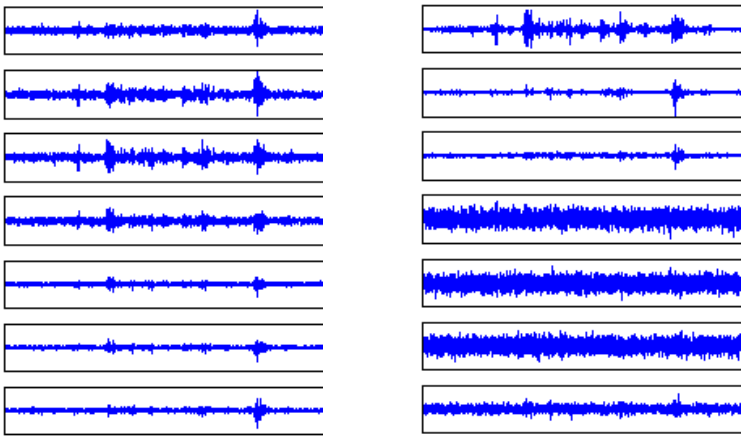

Fig. 4. EMG signals of the chin array before ICA processing (left) and after ICA processing (right). The ICA decomposition shows visibly distinct EMG signal components and artifact noise.

artifact components may be vastly different from target components: the 7 original EMG channels of the chin array of one utterance (left) are decomposed into three "target" components which look like EMG signals, and four "noise" components. Therefore we expect that the removal of the noise channels improves the recognition results. In [16] we present two strategies:

- **Direct method:** We take the ICA components, identify and remove artifact components, and then compute the EMG features on the *remaining components*.
- **Backprojection:** We take the ICA components, identify and remove artifact components as before, and then back-project these components to the original signal. Mathematically, this can be described as applying the ICA decomposition, setting the artifact ICA components to zero, and then multiplying the altered set of ICA components with the *inverse* of the ICA matrix.

In [16], it is shown that the direct method yields better results than backprojection. Therefore in this article we present results on artifact removal with the direct method, i. e. we compute EMG features on the remaining ICA components.

Artifact components are identified by the following three measures, which are computed on the ICA decomposition:

- Autocorrelation measure: This method typically identifies very regular (periodic) artifacts, like power line noise. We compute the autocorrelation sequence of the input component and then take the value of the *first maximum* after the first zero of the sequence. If this value is greater than 0.5, this component is deemed an artifact.

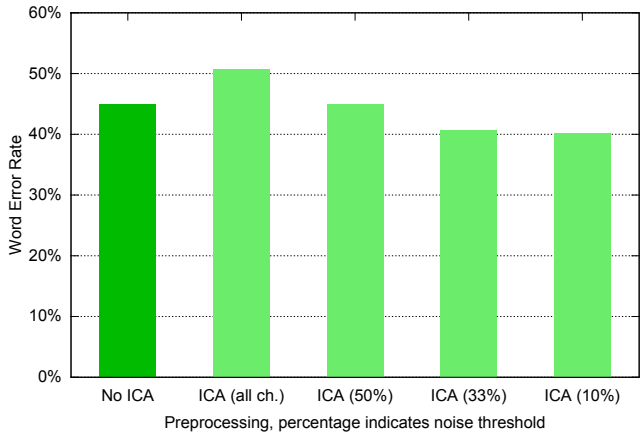

Fig. 5. Results for ICA with noise removal with different noise threshold percentages on the B-1 setup. See text for details.

- High-frequency noise detection: The surface EMG signal has a frequency range of 0Hz - 500Hz [17]. Therefore a component with distinct high-frequency parts is considered an artifact. We compute the discrete-time Fourier transform of the input component and divide the frequency axis into two intervals: The "signal" interval from 0Hz to 500Hz, and the "noise" interval from 500Hz to 1024Hz (the Nyquist frequency). We then compute the areas of the amplitude of the Fourier transform over the two intervals and divide the "signal" area by the "noise" area. If the quotient is smaller than 1.3, this component is deemed an artifact.
- EMG signal range: The main energy of the EMG signal is found between 50Hz and 150Hz [17]. As before, we divide the frequency axis into two parts: A "signal" interval from 50Hz to 150Hz, and "noise" part from 0Hz to 50Hz and from 150Hz to 1024Hz. Then we divide the "signal" area by the "noise" area. If the quotient is below 0.25, we deem this component an artifact. For this measure, we found that the power spectral density yielded slightly more robust estimates than a standard Fourier transformation.

Our measures are first computed on each ICA component of *each utterance* of the training data set. In a second step, we combine the results: For a component to be considered an artifact, we require that *at least one* of the three methods considers this component an artifact on *a minimum percentage* of (training) utterances. This "threshold percentage" is varied between 10% and 50%, where a lower value causes more components to be removed. We observed that the threshold makes a difference when components vary across utterances, e. g. when the contact between electrode and skin deteriorates over time.

For our experiments, we used the *optimal setup* of the B-1 corpus, namely TD5 stacking with PCA dimensionality reduction to 500. Figure 5 shows Word Error Rates for this experiment with different strategies: The baseline is attained

without ICA application, here we obtain 44.9% WER. ICA increases the WER to 50.6%—this result is in line with [18], where it is shown that direct ICA application may cause both improvement or deterioration of the WER. With a slightly different preprocessing, in [16] we obtained an insignificant improvement with direct ICA application.

Using our noise removal strategies clearly improves the WER: With a threshold percentage of 50%, we obtain 44.9% WER, about as much as without ICA. But we can still improve this result: The best WER is achieved with a 10% noise threshold, where the WER is reduced to 40.1%. Compared with the baseline, this is a relative improvement of 10.7%.

5 Conclusion

In this study we introduced a new EMG-based speech recognition technology, based on electrode arrays instead of single electrodes. We presented two basic recording setups and evaluated their potential on data sets with different amount of training data. An unexpected inconsistency with respect to the optimal feature stacking width led us to the introduction of a PCA preprocessing step before the LDA matrix is computed, which gives us consistent relative Word Error Rate improvements between 11% to 19%, even for small training data sets of only 40 sentences. As a first application of the new array technology, we showed that Independent Component Analysis (ICA) can be applied for artifact detection and removal, improving the Word Error Rate by 10.7% relative. However, we observed that application of ICA without the noise detection heuristics did not improve our result.

References

1. Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, and James Gilbert. Silent Speech Interfaces. *Speech Communication*, 52(4):270 – 287, 2010.
2. Tanja Schultz and Michael Wand. Modeling Coarticulation in Large Vocabulary EMG-based Speech Recognition. *Speech Communication*, 52(4):341 – 353, 2010.
3. Lena Maier-Hein, Florian Metze, Tanja Schultz, and Alex Waibel. Session Independent Non-Audible Speech Recognition Using Surface Electromyography. In *Proc. IEEE ASRU*, San Juan, Puerto Rico, 2005.
4. Joao Freitas, Antonio Teixeira, and Miguel Sales Dias. Towards a Silent Speech Interface for Portuguese. In *Proc. Biosignals*, 2012.
5. Charles Jorgensen and Sorin Dusan. Speech Interfaces based upon Surface Electromyography. *Speech Communication*, 52(4):354 – 366, 2010.
6. Eduardo Lopez-Larraz, Oscar M. Mozos, Javier M. Antelis, and Javier Minguez. Syllable-Based Speech Recognition Using EMG. In *Proc. IEEE EMBS*, 2010.
7. Bruce B. Winter and John G. Webster. Driven-right-leg Circuit Design. *IEEE Trans. Biomed. Eng.*, BME-30:62 – 66, 1983.
8. Szu-Chen Jou, Tanja Schultz, Matthias Walliczek, Florian Kraft, and Alex Waibel. Towards Continuous Speech Recognition using Surface Electromyography. In *Proc. Interspeech*, Pittsburgh, PA, 2006.

9. Michael Wand and Tanja Schultz. Speaker-Adaptive Speech Recognition Based on Surface Electromyography. In Ana Fred, Joaquim Filipe, and Hugo Gamboa, editors, *Biomedical Engineering Systems and Technologies*, volume 52 of *Communications in Computer and Information Science*, pages 271–285. Springer Berlin Heidelberg, 2010.
10. Naonori Ueda, Ryohei Nakano, Zoubin Ghahramani, and Geoffrey E. Hinton. Split and Merge EM Algorithm for Improving Gaussian Mixture Density Estimates. *Journal of VLSI Signal Processing*, 26:133 – 140, 2000.
11. Zhihua Qiao, Lan Zhou, and Jianhua Z. Huang. Sparse Linear Discriminant Analysis with Applications to High Dimensional Low Sample Size Data. *International Journal of Applied Mathematics*, 39:48 – 60, 2009.
12. Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs Fisherface: Recognition using Class-specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:711 – 720, 1997.
13. Aapo Hyvrinen and Erkki Oja. Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13:411 – 430, 2000.
14. Anthony J. Bell and Terrence I. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*, 7:1129 – 1159, 1995.
15. Scott Makeig et al. EEGLAB: ICA Toolbox for Psychophysiological Research. WWW Site, Swartz Center for Computational Neuroscience, Institute of Neural Computation, University of San Diego California, www.sccn.ucsd.edu/eeglab/, 2000. [World Wide Web Publication].
16. Michael Wand, Adam Himmelsbach, Till Heistermann, Matthias Janke, and Tanja Schultz. Artifact Removal Algorithm for an EMG-based Silent Speech Interface. In *Proc. IEEE EMBC*, 2013.
17. Hanqing Zhao and Guoqiang Xu. The Research on Surface Electromyography Signal Effective Feature Extraction. In *Proc. 6th International Forum on Strategic Technology*, 2011.
18. Michael Wand, Christopher Schulte, Matthias Janke, and Tanja Schultz. Array-based Electromyographic Silent Speech Interface. In *Proc. Biosignals*, 2013.