AlterEgo: A Personalized Wearable Silent Speech Interface

Type	Conference Paper
Author	Arnav Kapur
Author	Shreyas Kapur
Author	Pattie Maes
URL	http://dl.acm.org/citation.cfm?doid=3172944.3172977
Place	Tokyo, Japan
Publisher	ACM Press
Pages	43-53
ISBN	978-1-4503-4945-1
Date	2018
DOI	10.1145/3172944.3172977
Accessed	2/26/2019, 6:34:37 PM
Library Catalog	Crossref
Conference Name	the 2018 Conference
Language	en
Abstract	We present a wearable interface that allows a user to silently converse with a computing device without any voice or any discernible movements - thereby enabling the user to communicate with devices, AI assistants, applications or other people in a silent, concealed and seamless manner. A user's intention to speak and internal speech is characterized by neuromuscular signals in internal speech articulators that are captured by the AlterEgo system to reconstruct this speech. We use this to facilitate a natural language user interface, where users can silently communicate in natural language and receive aural output (e.g - bone conduction headphones), thereby enabling a discreet, bi-directional interface with a computing device, and providing a seamless form of intelligence augmentation. The paper describes the architecture, design, implementation and operation of the entire system. We demonstrate robustness of the system through user studies and report 92% median word accuracy levels.
Proceedings Title	Proceedings of the 2018 Conference on Human Information Interaction&Retrieval - IUI '18
Short Title	AlterEgo
Date Added	2/26/2019, 6:34:37 PM
Modified	2/26/2019, 6:34:37 PM

Notes:

- Data was collected during two main phases. First, we conducted a pilot study with 3 participants (1 female, average age of 29.33 years) to investigate feasibility of signal detection and to determine electrode positioning. The preliminary dataset recorded with the participants was binary, with the world labels being yes and no.
- This feature representation is passed through a 1**-dimensional convolutional neural network** to classify into word labels with the architecture described as follows. The hidden layer convolves 400 filters of kernel size 3 with stride 1 with the processed input and is then passed through a rectifier nonlinearity. This is subsequently followed by a max pooling layer.
- This unit is repeated twice before globally max pooling over its input. This is followed by a fully connected layer of dimension 200 passed through a rectifier nonlinearity which is followed by another fully connected layer with a sigmoid activation. The network was optimized using a first order gradient descent and parameters were updated using Adam [19] during training. The network was regularized using a 50% dropout in each hidden layer to enable the network to generalize better on unseen data.
- In order to help the users understand silent speech, we showed the user a piece of text and asked the user to read it like (s)he silently read online articles, i.e. read to oneself and not out loud. For each participant, we showed them a total of 750 digits, randomly sequenced on a computer screen, and instructed the users to 'read the number to themselves, without producing a sound and moving their lips'. The digits were randomly chosen from a total of 10 digits (0 to 9), such that each digit exactly appeared 75 times.

Attachments

- Kapur et al. - 2018 - AlterEgo A Personalized Wearable Silent Speech In.pdf

---

Classify ECG Signals Using LSTM Networks

Type Blog Post

URL https://blogs.mathworks.com/deep-learning/2018/08/06/classify-ecg-signals-using-lstm-networks/

Accessed 3/23/2019, 1:41:10 PM

Abstract Today I want to highlight a signal processing application of deep learning. This example, which is from the Signal Processing Toolbox documentation, shows how to classify heartbeat electrocardiogram (ECG) data from the PhysioNet 2017 Challenge using deep learning and signal processing. In particular, the example uses Long Short-Term Memory (LSTM) networks and time-frequency analysis. This example requires Neural Network Toolbox™. Contents

Blog Title Deep Learning

Date Added 3/23/2019, 1:41:10 PM

Modified 3/23/2019, 1:41:10 PM

Attachments

- Snapshot

Deep Learning - The Past, Present and Future of Artificial Intelligen…

Type Presentation

Presenter Lukas Masuch

URL https://www.slideshare.net/LuMa921/deep-learning-the-past-present-and-future-of-artificial-intelligence?ref=https://www.analyticsindiamag.com/popular-presentations-on-artificial-intelligence-and-machine-learning/

Date 23:58:03 UTC

Accessed 3/19/2019, 11:28:03 PM

Type Technology

Abstract In the last couple of years, deep learning techniques have transformed the

Date Added 3/19/2019, 11:28:03 PM

Modified 3/19/2019, 11:28:03 PM

Deep Learning for the Classification of EEG Time-Frequency Representations

Type Journal Article

Author Audun Eltvik

Pages 122

Library Catalog Zotero

Language en

Abstract This thesis is a report on the implementation and evaluation of a new method classifying EEG signals. The method involves applying either the Short-time Fourier Transform (STFT), Continuous Wavelet Transform (CWT) or Hilbert-Huang Transform (HHT) to produce a two-dimensional time-frequency representation of the signal, known as spectrograms, scalograms and Hilbert spectra, respectively. These two-dimensional representations are then classified using a Convolutional Neural Network (CNN).

Date Added 3/11/2019, 11:49:03 PM

Modified 3/11/2019, 11:49:03 PM

Attachments

- Eltvik - Deep Learning for the Classification of EEG Time-F.pdf

---

Diagnosing Myocardial Infarction using Long-Short Term Memory networks (LSTM's)

Type Web Page

Author Luc Nies

URL https://blog.orikami.nl/diagnosing-myocardial-infarction-using-long-short-term-memory-networks-lstms-cedf5770a257

Date 2018-07-31T11:05:18.365Z

Accessed 3/23/2019, 2:29:03 PM

Abstract Introduction

Website Title Orikami blog

Date Added 3/23/2019, 2:29:03 PM

Modified 3/23/2019, 2:29:03 PM

Attachments

- Snapshot

Direct conversion from facial myoelectric signals to speech using Deep Neural Networks

Type Conference Paper

Author L. Diener

Author M. Janke

Author T. Schultz

Pages	1-7
Date	July 2015
DOI	10.1109/IJCNN.2015.7280404
Library Catalog	IEEE Xplore
Conference Name	2015 International Joint Conference on Neural Networks (IJCNN)
Abstract	This paper presents our first results using Deep Neural Networks for surface electromyographic (EMG) speech synthesis. The proposed approach enables a direct mapping from EMG signals captured from the articulatory muscle movements to the acoustic speech signal. Features are processed from multiple EMG channels and are fed into a feed forward neural network to achieve a mapping to the target acoustic speech output. We show that this approach is feasible to generate speech output from the input EMG signal and compare the results to a prior mapping technique based on Gaussian mixture models. The comparison is conducted via objective Mel-Cepstral distortion scores and subjective listening test evaluations. It shows that the proposed Deep Neural Network approach gives substantial improvements for both evaluation criteria.
Proceedings Title	2015 International Joint Conference on Neural Networks (IJCNN)
Date Added	3/13/2019, 6:06:06 PM
Modified	3/13/2019, 6:06:06 PM

Tags:

feature extraction, electromyography, Electromyography, medical signal processing, acoustic speech signal, articulatory muscle movements, cepstral analysis, deep neural network approach, EMG signal capture, facial myoelectric signal-to-speech conversion, feature processing, feedforward neural network, Gaussian mixture models, Gaussian processes, mixture models, multiple EMG channels, neural nets, neurophysiology, objective Mel-Cepstral distortion scores, prior mapping technique, speech, subjective listening test evaluations, surface electromyographic speech synthesis

Attachments

- IEEE Xplore Abstract Record
- IEEE Xplore Full Text PDF

EMG Pattern Prediction for Upper Limb Movements Based on Wavelet and Hilbert-Huang Transform

Type Journal Article
Author Alvaro Altamirano Altamirano
Pages 134
Library Catalog Zotero
Language en
Date Added 3/12/2019, 6:52:39 PM
Modified 3/12/2019, 6:52:39 PM

Attachments

- Altamirano - EMG Pattern Prediction for Upper Limb Movements Ba.pdf

---

EMG-to-Speech: Direct Generation of Speech From Facial Electromyographic Signals

Type Journal Article
Author Matthias Janke
Author Lorenz Diener
URL http://ieeexplore.ieee.org/document/8114359/
Volume 25
Issue 12
Pages 2375-2385
Publication IEEE/ACM Transactions on Audio, Speech, and Language Processing
ISSN 2329-9290, 2329-9304
Date 12/2017
DOI 10.1109/TASLP.2017.2738568
Accessed 2/26/2019, 6:34:53 PM
Library Catalog Crossref
Language en
Abstract Silent speech interfaces are systems that enable speech communication even when an acoustic signal is unavailable. Over the last years, public interest in such interfaces has intensified. They provide solutions for some of the challenges faced by today’s speech-driven technologies, such as robustness to noise and usability for people with speech impediments. In this paper, we provide an overview over our silent speech interface. It is based on facial surface electromyography (EMG), which we use to record the electrical signals that control muscle contraction during speech production. These signals are then converted directly to an audible speech waveform, retaining important paralinguistic speech cues for information such as speaker identity and mood. This paper gives an overview over our state-of-the-art direct EMG-to-speech

transformation system. This paper describes the characteristics of the speech EMG signal, introduces techniques for extracting relevant features, presents different EMG-to-speech mapping methods, and finally, presents an evaluation of the different methods for real-time capability and conversion quality.

Short Title	EMG-to-Speech
Date Added	2/26/2019, 6:34:53 PM
Modified	2/26/2019, 6:34:53 PM

Notes:

- used a 3-hidden-layer feed forward neural network
- To avoid bias towards numerically larger EMG- or audio features, the signal is normalized to zero mean and unit variance. Dropout [75] is used to reduce overfitting.
- Used LSTM - The LSTMs used in this work are bidirectional LSTMs. Graves et al. [80].
- According to extensive experiments on electrode positioning [84], EMG-based speech processing requires, at the very least, signals from the cheek area and the throat (to capture tongue activity).
- All measurements were obtained on an Intel Core i7-2700 CPU running at 3.5 GHz.
- For the evaluation of LSTM networks, we used the CURRENNT implementation

Attachments

- Janke and Diener - 2017 - EMG-to-Speech Direct Generation of Speech From Fa.pdf

---

Interspeech 2007 Abstract: Wand et al.

Type	Web Page
URL	https://www.isca-speech.org/archive/interspeech_2007/i07_0686.html
Accessed	3/7/2019, 11:11:48 AM
Date Added	3/7/2019, 11:11:52 AM
Modified	3/7/2019, 11:11:52 AM

Notes:

- Non-audible speech is favored for the sake of privacy,
  for example when making a confidential phone call in public
  spaces. Last but not least, alternative input methods for speech
  recognition may be useful for patients with medical speech impairments.
-

Attachments

- wand_is07.pdf

Pattern learning with deep neural networks in EMG-based speech recognition

Type	Conference Paper
Author	Michael Wand
Author	Tanja Schultz
URL	http://ieeexplore.ieee.org/document/6944550/
Place	Chicago, IL
Publisher	IEEE
Pages	4200-4203
ISBN	978-1-4244-7929-0
Date	8/2014
DOI	10.1109/EMBC.2014.6944550
Accessed	2/26/2019, 6:35:18 PM
Library Catalog	Crossref
Conference Name	2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
Language	en
Abstract	We report on classification of phones and phonetic features from facial electromyographic (EMG) data, within the context of our EMG-based Silent Speech interface. In this paper we show that a Deep Neural Network can be used to perform this classification task, yielding a significant improvement over conventional Gaussian Mixture models. Our central contribution is the visualization of patterns which are learned by the neural network. With increasing network depth, these patterns represent more and more intricate electromyographic activity.
Proceedings Title	2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society
Date Added	2/26/2019, 6:35:18 PM

Modified 2/26/2019, 6:35:18 PM

Notes:

- 25 sessions from 20 speakers, each
comprising 200 read English-language utterances spoken in
normal, audible speech.
- This data was recorded in bipolar fashion, where the difference of two
adjacent channels is taken to reduce common mode artifacts,
thus we finally got 35 (5 7) EMG channels. Sampling was
performed at 2048Hz.
- We assume
the input data to be Gaussian distributed and modify the
RBM algorithm to get a Gaussian-Bernoulli RBM [14]. Our
code is based on the original scripts by Hinton

- Uses GMM and Restricted Boltzmann Machine Algorithm, not LSTM or RNN or Convolutional neural networks.

Attachments

- Wand and Schultz - 2014 - Pattern learning with deep neural networks in EMG-.pdf

---

Session independent non-audible speech recognition using surface electromyography

Type	Conference Paper
Author	L. Maier-Hein
Author	F. Metze
Author	T. Schultz
Author	A. Waibel
URL	http://ieeexplore.ieee.org/document/1566521/
Place	San Juan, Puerto Rico
Publisher	IEEE
Pages	331-336
ISBN	978-0-7803-9479-7
Date	2005
DOI	10.1109/ASRU.2005.1566521
Accessed	3/13/2019, 5:52:21 PM
Library Catalog	Crossref

Conference Name IEEE Workshop on Automatic Speech Recognition and Understanding, 2005.

Language en

Abstract In this paper we introduce a speech recognition system based on myoelectric signals. The system handles audible and non-audible speech. Major challenges in surface electromyography based speech recognition ensue from repositioning electrodes between recording sessions, environmental temperature changes, and skin tissue properties of the speaker. In order to reduce the impact of these factors, we investigate a variety of signal normalization and model adaptation methods. An average word accuracy of 97.3% is achieved using seven EMG channels and the same electrode positions. The performance drops to 76.2% after repositioning the electrodes if no normalization or adaptation is performed. By applying our adaptation methods we manage to restore the recognition rates to 87.1%. Furthermore, we compare audibly to non-audibly spoken speech. The results suggest that large differences exist between the corresponding muscle movements. Still, our recognition system recognizes both speech manners accurately when trained on pooled data.

Proceedings Title IEEE Workshop on Automatic Speech Recognition and Understanding, 2005.

Date Added 3/13/2019, 5:52:21 PM

Modified 3/13/2019, 5:52:21 PM

Notes:

- A high-pass filteris applied to avoid aliasing artefacts whereasalow-pass filteris used to reduce movement artefacts in the sig-nals
- in this study, isolated word recognition was performed on a vocab-ulary consisting of the ten English digits “zero” to “nine”.
- Figure 1 the electrodes were positioned such that theyobtain the EMG signal of six articular muscles: thelevator angulioris(EMG2,3), thezygomaticus major(EMG2,3), theplatysma(EMG4,5) thedepressor anguli oris(EMG5), theanterior bellyof thedigastric(EMG1) and thetongue(EMG1,6,7) [10, 6]. Forthree of the seven EMG channels (EMG2,6,7) a classical bipolarelectrode configuration with a 2cm center-to-center inter-electrodespacing was used. For the remaining four channels one of thedetection electrodes was placed directly on the articulatory mus-cles and was referenced to either the nose (EMG1) or to both ears(EMG3,4,5) (Figure 1). The positioning of the electrodes was op-timized in previous experiments, not reported here.

Attachments

- Maier-Hein et al. - 2005 - Session independent non-audible speech recognition.pdf

Speech Recognition with Deep Recurrent Neural Networks

Type Journal Article
Author Alex Graves
Author Abdel-rahman Mohamed
Author Geoffrey Hinton
URL http://arxiv.org/abs/1303.5778
Publication arXiv:1303.5778 [cs]
Date 2013-03-22
Extra arXiv: 1303.5778
Accessed 2/26/2019, 6:35:27 PM
Library Catalog arXiv.org
Language en
Abstract Recurrent neural networks (RNNs) are a powerful model for sequential data. End-to-end training methods such as Connectionist Temporal Classification make it possible to train RNNs for sequence labelling problems where the input-output alignment is unknown. The combination of these methods with the Long Short-term Memory RNN architecture has proved particularly fruitful, delivering state-of-the-art results in cursive handwriting recognition. However RNN performance in speech recognition has so far been disappointing, with better results returned by deep feedforward networks. This paper investigates deep recurrent neural networks, which combine the multiple levels of representation that have proved so effective in deep networks with the flexible use of long range context that empowers RNNs. When trained end-to-end with suitable regularisation, we find that deep Long Short-term Memory RNNs achieve a test set error of 17.7% on the TIMIT phoneme recognition benchmark, which to our knowledge is the best recorded score.
Date Added 2/26/2019, 6:35:27 PM
Modified 2/26/2019, 6:35:27 PM

Tags:

Computer Science - Computation and Language, Computer Science - Neural and Evolutionary Computing

Notes:

Comment: To appear in ICASSP 2013

Attachments

- Graves et al. - 2013 - Speech Recognition with Deep Recurrent Neural Netw.pdf

---

Surface Electromyography for Speech and Swallowing Systems: Measurement, Analysis, and Interpretation

Type	Journal Article
Author	Cara E. Stepp
URL	http://pubs.asha.org/doi/10.1044/1092-4388%282011/11-0214%29
Volume	55
Issue	4
Pages	1232-1246
Publication	Journal of Speech, Language, and Hearing Research
ISSN	1092-4388, 1558-9102
Date	08/2012
DOI	10.1044/1092-4388(2011/11-0214)
Accessed	3/17/2019, 6:25:57 PM
Library Catalog	Crossref
Language	en
Abstract	Purpose: Applying surface electromyography (sEMG) to the study of voice, speech, and swallowing is becoming increasingly popular. An improved understanding of sEMG and building a consensus as to appropriate methodology will improve future research and clinical applications. Method: An updated review of the theory behind recording sEMG for the speech and swallowing systems is provided. Several factors that are known to affect the content of the sEMG signal are discussed, and practical guidelines for sEMG recording and analysis are presented, focusing on special considerations within the context of the speech and swallowing anatomy. Results: Unique challenges are seen in application of sEMG to the speech and swallowing musculature owing to the small size of the muscles in relation to the sEMG detection volume and the present lack of knowledge about innervation zone locations. Conclusions: Despite the challenges discussed, application of sEMG to speech and swallowing has potential as a clinical and research tool when used correctly and is specifically suited to noninvasive clinical studies using between-condition or between-group comparisons for which detection of specific isolated muscle activities is not necessary.
Short Title	Surface Electromyography for Speech and Swallowing Systems
Date Added	3/17/2019, 6:25:57 PM
Modified	3/17/2019, 6:25:57 PM

Notes:

Diagram of neck muscles as seen from the front illustratingexamples of bipolar electrode configurations. A: Incorrect bilateralconfiguration. B: Suggested configuration with electrodes placedparallel to the longitudinal axis of the muscle body, in line with thefibers of the muscle. TH = thyrohyoid; OH = omohyoid, SCM =sternocleidomastoid; SH = sternohyoid; ST = sternothyroid

Attachments

- Stepp - 2012 - Surface Electromyography for Speech and Swallowing.pdf

---

Syllable-based speech recognition using EMG

Type	Conference Paper
Author	E. Lopez-Larraz
Author	O. M. Mozos
Author	J. M. Antelis
Author	J. Minguez
Pages	4699-4702
Date	August 2010
DOI	10.1109/IEMBS.2010.5626426
Library Catalog	IEEE Xplore
Conference Name	2010 Annual International Conference of the IEEE Engineering in Medicine and Biology
Abstract	This paper presents a silent-speech interface based on electromyographic (EMG) signals recorded in the facial muscles. The distinctive feature of this system is that it is based on the recognition of syllables instead of phonemes or words, which is a compromise between both approaches with advantages as (a) clear delimitation and identification inside a word, and (b) reduced set of classification groups. This system transforms the EMG signals into robust-in-time feature vectors and uses them to train a boosting classifier. Experimental results demonstrated the effectiveness of our approach in three subjects, providing a mean classification rate of almost 70% (among 30 syllables).
Proceedings Title	2010 Annual International Conference of the IEEE Engineering in Medicine and Biology
Date Added	2/26/2019, 6:53:01 PM
Modified	2/26/2019, 6:53:01 PM

Tags:

EMG, feature extraction, boosting classifier, Decision trees, Electrodes, electromyographic signals, electromyography, Electromyography, facial muscles, Facial muscles, Facial Muscles, medical signal processing, Muscles, Natural Language Processing, Pattern Recognition, Automated, robust-in-time feature vectors, Semantics, signal classification, silent-speech interface, Speech, Speech Production Measurement, speech recognition, Speech recognition, Speech Recognition Software, syllable-based speech recognition

Attachments

- IEEE Xplore Abstract Record
- IEEE Xplore Full Text PDF