Effect of Noise Reduction Gain Errors on Simulated Cochlear Implant Speech Intelligibility

It has been suggested that the most important factor for obtaining high speech intelligibility in noise with cochlear implant (CI) recipients is to preserve the low-frequency amplitude modulations of speech across time and frequency by, for example, minimizing the amount of noise in the gaps between speech segments. In contrast, it has also been argued that the transient parts of the speech signal, such as speech onsets, provide the most important information for speech intelligibility. The present study investigated the relative impact of these two factors on the potential benefit of noise reduction for CI recipients by systematically introducing noise estimation errors within speech segments, speech gaps, and the transitions between them. The introduction of these noise estimation errors directly induces errors in the noise reduction gains within each of these regions. Speech intelligibility in both stationary and modulated noise was then measured using a CI simulation tested on normal-hearing listeners. The results suggest that minimizing noise in the speech gaps can improve intelligibility, at least in modulated noise. However, significantly larger improvements were obtained when both the noise in the gaps was minimized and the speech transients were preserved. These results imply that the ability to identify the boundaries between speech segments and speech gaps may be one of the most important factors for a noise reduction algorithm because knowing the boundaries makes it possible to minimize the noise in the gaps as well as enhance the low-frequency amplitude modulations of the speech.
The impact of noise power estimation on speech intelligibility in cochlear-implant speech coding strategies

The advanced combination encoder (ACE™) is an established speech-coding strategy in cochlear-implant processing that selects a number of frequency channels based on amplitudes. However, speech intelligibility outcomes with this strategy are limited in noisy conditions. To improve speech intelligibility, either noise-dominant channels can be attenuated prior to ACE™ with noise reduction or, alternatively, channels can be selected based on estimated signal-to-noise ratios. A noise power estimation stage is, therefore, required. This study investigated the impact of noise power estimation in noise-reduction and channel-selection strategies. Results imply that estimation with improved noise-tracking capabilities does not necessarily translate into increased speech intelligibility.

General information
State: Published
Organisations: Hearing Systems, Department of Electrical Engineering, Department of Health Technology, Cochlear, Electromagnetic Systems Group
Contributors: Bentsen, T., Mauger, S. J., Kressner, A. A., May, T., Dau, T.
Pages: 818-821
Publication date: 1 Feb 2019
Peer-reviewed: Yes

Publication information
Journal: Journal of the Acoustical Society of America
Volume: 145
Issue number: 2
ISSN (Print): 0001-4966
Ratings:
BFI (2019): BFI-level 2
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 1.77 SJR 0.695 SNIP 1.224
Web of Science (2017): Impact factor 1.605
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 1.83 SJR 0.819 SNIP 1.271
Web of Science (2016): Impact factor 1.547
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 1.77 SJR 0.854 SNIP 1.416
Web of Science (2015): Impact factor 1.572
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 1.8 SJR 0.887 SNIP 1.402
Web of Science (2014): Impact factor 1.503
The benefit of combining a deep neural network architecture with ideal ratio mask estimation in computational speech segregation to improve speech intelligibility

Computational speech segregation attempts to automatically separate speech from noise. This is challenging in conditions with interfering talkers and low signal-to-noise ratios. Recent approaches have adopted deep neural networks and successfully demonstrated speech intelligibility improvements. A selection of components may be responsible for the success with these state-of-the-art approaches: the system architecture, a time frame concatenation technique and the learning objective. The aim of this study was to explore the roles and the relative contributions of these components by measuring speech intelligibility in normal-hearing listeners. A substantial improvement of 25.4 percentage points in speech intelligibility scores was found going from a subband-based architecture, in which a Gaussian Mixture Model-based classifier predicts the distributions of speech and noise for each frequency channel, to a state-of-the-art deep neural network-based architecture. Another improvement of 13.9 percentage points was obtained by changing the learning objective from the ideal binary mask, in which individual time-frequency units are labeled as either speech- or noise-dominated, to the ideal ratio mask, where the units are assigned a continuous value between zero and one. Therefore, both components play significant roles and by combining them, speech intelligibility improvements were obtained in a six-talker condition at a low signal-to-noise ratio.
The impact of exploiting spectro-temporal context in computational speech segregation
Computational speech segregation aims to automatically segregate speech from interfering noise, often by employing ideal binary mask estimation. Several studies have tried to exploit contextual information in speech to improve mask estimation accuracy by using two frequently-used strategies that (1) incorporate delta features and (2) employ support vector machine (SVM) based integration. In this study, two experiments were conducted. In Experiment I, the impact of exploiting spectro-temporal context using these strategies was investigated in stationary and six-talker noise. In Experiment II, the delta features were explored in detail and tested in a setup that considered novel noise segments of the six-talker noise. Computing delta features led to higher intelligibility than employing SVM based integration and intelligibility increased with the amount of spectral information exploited via the delta features. The system did not, however, generalize well to novel segments of this noise type. Measured intelligibility was subsequently compared to extended short-term objective intelligibility, hit–false alarm rate, and the amount of mask clustering. None of these objective measures alone could account for measured intelligibility. The findings may have implications for the design of speech segregation systems, and for the selection of a cost function that correlates with intelligibility.

General information
State: Published
Organisations: Department of Electrical Engineering, Hearing Systems
Contributors: Bentsen, T., Kressner, A. A., Dau, T., May, T.
Pages: 248-259
Publication date: 2018
Peer-reviewed: Yes

Publication information
Journal: Journal of the Acoustical Society of America
Volume: 143
Issue number: 1
ISSN (Print): 0001-4966
Ratings:
BFI (2019): BFI-level 2
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 1.77 SJR 0.695 SNIP 1.224
Web of Science (2017): Impact factor 1.605
Web of Science (2017): Indexed yes
BFI (2016): BFI-level 2
Scopus rating (2016): CiteScore 1.83 SJR 0.819 SNIP 1.271
Web of Science (2016): Impact factor 1.547
Web of Science (2016): Indexed yes
BFI (2015): BFI-level 2
Scopus rating (2015): CiteScore 1.77 SJR 0.854 SNIP 1.416
Web of Science (2015): Impact factor 1.572
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 1.8 SJR 0.887 SNIP 1.402
Web of Science (2014): Impact factor 1.503
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 2 SJR 0.707 SNIP 1.937
Web of Science (2013): Impact factor 1.555
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 1.75 SJR 0.771 SNIP 1.619
Web of Science (2012): Impact factor 1.646
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 1.68 SJR 0.686 SNIP 1.624
Web of Science (2011): Impact factor 1.55
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 0.734 SNIP 1.511
Web of Science (2010): Impact factor 1.644
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 0.778 SNIP 1.692
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 0.83 SNIP 1.657
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 0.838 SNIP 1.635
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 0.739 SNIP 1.678
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 0.946 SNIP 1.728
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 0.77 SNIP 1.761
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 0.875 SNIP 1.695
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 0.785 SNIP 1.572
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 0.727 SNIP 1.483
Web of Science (2001): Indexed yes
Scopus rating (2000): SJR 0.639 SNIP 1.404
Web of Science (2000): Indexed yes
The impact of reverberation on speech intelligibility in cochlear implant recipients

Listening to speech in an environment with reverberation can be challenging for both the normal and impaired auditory system. However, it has been shown for both normal- and impaired-hearing listeners that it is the late reflections that are responsible for degrading intelligibility, whereas early reflections actually aid intelligibility by increasing the effective signal-to-noise ratio. Contrastingly, studies conducted with cochlear implant (CI) recipients have suggested that CI recipients have almost no tolerance for reverberation and that they are negatively impacted by both early and late reflections. The main objective of the current study is to re-evaluate the influence of reverberation on speech intelligibility in CI recipients using more authentic virtual auditory environments. Unlike previous studies in this area, this study was conducted using a loudspeaker-based auralization system rather than non-individualized binaural room simulations. Speech intelligibility was measured in simulations of a range of actual physical rooms with plausible source-receiver distances, both with and without late reflections. The results show that the effect of reverberation is much smaller than previously suggested, especially with short source-receiver distances. Furthermore, the results suggest that, in contrast to previous literature, early reflections may not actually be detrimental to CI recipients.
Investigating the effects of noise-estimation errors in simulated cochlear implant speech intelligibility

A recent study suggested that the most important factor for obtaining high speech intelligibility in noise with cochlear implant recipients is to preserve the low-frequency amplitude modulations of speech across time and frequency by, for example, minimizing the amount of noise in speech gaps. In contrast, other studies have argued that the transients provide the most information. Thus, the present study investigates the relative impact of these two factors in the
framework of noise reduction by systematically correcting noise-estimation errors within speech segments, speech gaps, and the transitions between them. Speech intelligibility in noise was measured using a cochlear implant simulation tested on normal-hearing listeners. The results suggest that minimizing noise in the speech gaps can substantially improve intelligibility, especially in modulated noise. However, significantly larger improvements were obtained when both the noise in the gaps was minimized and the speech transients were preserved. These results imply that the correct identification of the boundaries between speech segments and speech gaps is the most important factor in maintaining high intelligibility in cochlear implants. Knowing the boundaries will make it possible for algorithms to both minimize the noise in the gaps and enhance the low frequency amplitude modulations.

Cochlear implant speech intelligibility outcomes with structured and unstructured binary mask errors
It has been shown that intelligibility can be improved for cochlear implant (CI) recipients with the ideal binary mask (IBM). In realistic scenarios where prior information is unavailable, however, the IBM must be estimated, and these estimations will inevitably contain errors. Although the effects of both unstructured and structured binary mask errors have been investigated with normal-hearing (NH) listeners, they have not been investigated with CI recipients. This study assesses these effects with CI recipients using masks that have been generated systematically with a statistical model. The results demonstrate that clustering of mask errors substantially decreases the tolerance of errors, that incorrectly removing target-dominated regions can be as detrimental to intelligibility as incorrectly adding interferer-dominated regions, and that the individual tolerances of the different types of errors can change when both are present. These trends follow those of NH listeners. However, analysis with a mixed effects model suggests that CI recipients tend to be less tolerant than NH listeners to mask errors in most conditions, at least with respect to the testing methods in each of the studies. This study clearly demonstrates that structure influences the tolerance of errors and therefore should be considered when analyzing binary-masking algorithms.
Comparing the influence of spectro-temporal integration in computational speech segregation

The goal of computational speech segregation systems is to automatically segregate a target speaker from interfering maskers. Typically, these systems include a feature extraction stage in the front-end and a classification stage in the back-end. A spectrotemporal integration strategy can be applied in either the frontend, using the so-called delta features, or in the back-end, using a second classifier that exploits the posterior probability of speech from the first classifier across a spectro-temporal window. This study systematically analyzes the influence of such stages on segregation performance, the error distributions and intelligibility predictions. Results indicated that it could be problematic to exploit context in the back-end, even though such a spectro-temporal integration stage improves the segregation performance. Also, the results emphasized the potential need of a single metric that comprehensively predicts computational segregation performance and correlates well with intelligibility. The outcome of this study could help to identify the most effective spectro-temporal integration strategy for computational segregation systems.

Outcome measures based on classification performance fail to predict the intelligibility of binary-masked speech

To date, the most commonly used outcome measure for assessing ideal binary mask estimation algorithms is based on the difference between the hit rate and the false alarm rate (H-FA). Recently, the error distribution has been shown to substantially affect intelligibility. However, H-FA treats each mask unit independently and does not take into account how errors are distributed. Alternatively, algorithms can be evaluated with the short-time objective intelligibility (STOI) metric using the reconstructed speech. This study investigates the ability of H-FA and STOI to predict intelligibility for binary-masked speech using masks with different error distributions. The results demonstrate the inability of H-FA to predict the behavioral intelligibility and also illustrate the limitations of STOI. Since every estimation algorithm will make errors that are distributed in different ways, performance evaluations should not be made solely on the basis of these metrics.
Structure in time-frequency binary masking errors and its impact on speech intelligibility

Although requiring prior knowledge makes the ideal binary mask an impractical algorithm, substantial increases in measured intelligibility make it a desirable benchmark. While this benchmark has been studied extensively, many questions remain about the factors that influence the intelligibility of binary-masked speech with non-ideal masks. To date, researchers have used primarily uniformly random, uncorrelated mask errors and independently presented error types (i.e., false positives and negatives) to characterize the influence of estimation errors on intelligibility. However, practical estimation algorithms produce masks that contain errors of both types and with non-trivial amounts of structure. This paper introduces an investigation framework for binary masks and presents listener studies that use this framework to illustrate how interactions between error types and structure affect intelligibility. First, this study demonstrates that clustering (i.e., a form of structure) of mask errors reduces intelligibility. Furthermore, while previous research has suggested that false positives are more detrimental to intelligibility than false negatives, this study indicates that false negatives can be equally detrimental to intelligibility when they contain structure or when both error types are present. Finally, this study shows that listeners tolerate fewer mask errors when both types of errors are present, especially when the errors contain structure.
Web of Science (2015): Indexed yes
BFI (2014): BFI-level 2
Scopus rating (2014): CiteScore 1.8 SJR 0.887 SNIP 1.402
Web of Science (2014): Impact factor 1.503
Web of Science (2014): Indexed yes
BFI (2013): BFI-level 2
Scopus rating (2013): CiteScore 2 SJR 0.707 SNIP 1.937
Web of Science (2013): Impact factor 1.555
ISI indexed (2013): ISI indexed yes
Web of Science (2013): Indexed yes
BFI (2012): BFI-level 2
Scopus rating (2012): CiteScore 1.75 SJR 0.771 SNIP 1.619
Web of Science (2012): Impact factor 1.646
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
BFI (2011): BFI-level 2
Scopus rating (2011): CiteScore 1.68 SJR 0.686 SNIP 1.624
Web of Science (2011): Impact factor 1.55
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
BFI (2010): BFI-level 2
Scopus rating (2010): SJR 0.734 SNIP 1.511
Web of Science (2010): Impact factor 1.644
Web of Science (2010): Indexed yes
BFI (2009): BFI-level 2
Scopus rating (2009): SJR 0.778 SNIP 1.692
Web of Science (2009): Indexed yes
BFI (2008): BFI-level 2
Scopus rating (2008): SJR 0.83 SNIP 1.657
Web of Science (2008): Indexed yes
Scopus rating (2007): SJR 0.838 SNIP 1.635
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 0.739 SNIP 1.678
Web of Science (2006): Indexed yes
Scopus rating (2005): SJR 0.946 SNIP 1.728
Web of Science (2005): Indexed yes
Scopus rating (2004): SJR 0.77 SNIP 1.761
Web of Science (2004): Indexed yes
Scopus rating (2003): SJR 0.875 SNIP 1.695
Web of Science (2003): Indexed yes
Scopus rating (2002): SJR 0.785 SNIP 1.572
Web of Science (2002): Indexed yes
Scopus rating (2001): SJR 0.727 SNIP 1.483
Web of Science (2001): Indexed yes
Scopus rating (2000): SJR 0.639 SNIP 1.404
Web of Science (2000): Indexed yes
Scopus rating (1999): SJR 0.56 SNIP 1.306
Original language: English
DOIs:
10.1121/1.4916271
Source: PublicationPreSubmission
Source-ID: 119928554
Research output: Research - peer-review › Journal article – Annual report year: 2015
A novel binary mask estimator based on sparse approximation

While most single-channel noise reduction algorithms fail to improve speech intelligibility, the ideal binary mask (IBM) has demonstrated substantial intelligibility improvements. However, this approach exploits oracle knowledge. The main objective of this paper is to introduce a novel binary mask estimator based on a simple sparse approximation algorithm. Our approach does not require oracle knowledge and instead uses knowledge of speech structure. © 2013 IEEE.

General information
State: Published
Organisations: Georgia Institute of Technology
Contributors: Kressner, A. A., Anderson, D. V., Rozell, C. J.
Pages: 7497-7501
Publication date: 2013

Host publication information
Title of host publication: Proceedings of Ieee International Conference on Acoustics, Speech and Signal Processing
Publisher: IEEE
ISBN (Print): 978-1-4799-0356-6
Keywords: Signal Processing, Software, Electrical and Electronic Engineering, Ideal binary mask, intelligibility, noise reduction, sparse approximation, time-frequency masking, Binary masks, Ideal binary mask (IBM), Intelligibility improvements, Single-channel noise reductions, Sparse approximations, Time-Frequency Masking, Approximation algorithms, Noise abatement, Signal processing, Speech intelligibility, Acoustic noise, approximation theory, speech intelligibility, Signal Processing and Analysis, Approximation methods, binary mask estimator, Matching pursuit algorithms, oracle knowledge, Signal processing algorithms, Signal to noise ratio, single-channel noise reduction algorithm, Speech, Time-frequency analysis
DOIs:
10.1109/icassp.2013.6639120
Source: FindIt
Source-ID: 250704315
Research output: Research - peer-review › Article in proceedings – Annual report year: 2013

Causal binary mask estimation for speech enhancement using sparsity constraints

While most single-channel noise reduction algorithms fail to improve speech intelligibility, the ideal binary mask (IBM) has demonstrated substantial intelligibility improvements for both normal- and impaired-hearing listeners. However, this approach exploits oracle knowledge of the target and interferer signals to preserve only the time-frequency regions that are target-dominated. Single-channel noise suppression algorithms trying to approximate the IBM using locally estimated signal-to-noise ratios without oracle knowledge have had limited success. Thought of in another way, the IBM exploits the disjoint placement of the target and interferer in time and frequency to create a time-frequency signal representation that is more sparse (i.e., has fewer non-zeros). In recent work (submitted to ICASSP 2013) we have introduced a novel time-frequency masking algorithm based on a sparse approximation algorithm from the signal processing literature. However, the algorithm employs a non-causal estimator. The present work introduces an improved de-noising algorithm that uses more realistic frame-based (causal) computations to estimate a binary mask.

General information
State: Published
Organisations: Department of Electrical Engineering, Hearing Systems, Georgia Institute of Technology
Contributors: Kressner, A. A., Anderson, D. V., Rozell, C. J.
Number of pages: 9
Publication date: 2013

Host publication information
Title of host publication: Proceedings of Meetings on Acoustics
Article number: 055037
DOIs:
10.1121/1.4800862
Source: PublicationPreSubmission
Source-ID: 119928594
Research output: Research - peer-review › Article in proceedings – Annual report year: 2013

Evaluating the Generalization of the Hearing Aid Speech Quality Index (HASQI)

Many developers of audio signal processing strategies rely on objective measures of quality for initial evaluations of algorithms. As such, objective measures should be robust, and they should be able to predict quality accurately regardless of the dataset or testing conditions. Kates and Arehart have developed the Hearing Aid Speech Quality Index (HASQI) to predict the effects of noise, nonlinear distortion, and linear filtering on speech quality for both normal-hearing and hearing-
impaired listeners, and they report very high performance with their training and testing datasets [Kates, J. and Arehart, K., Audio Eng. Soc., 58(5), 363-381 (2010)]. In order to investigate the generalizability of HASQI, we test its ability to predict normal-hearing listeners' subjective quality ratings of a dataset on which it was not trained. This dataset is designed specifically to contain a wide range of distortions introduced by real-world noises which have been processed by some of the most common noise suppression algorithms in hearing aids. We show that HASQI achieves prediction performance comparable to the Perceptual Evaluation of Speech Quality (PESQ), the standard for objective measures of quality, as well as some of the other measures in the literature. Furthermore, we identify areas of weakness and show that training can improve quantitative prediction.

General information
State: Published
Organisations: Georgia Institute of Technology
Contributors: Kressner, A. A., Anderson, D. V., Rozell, C. J.
Pages: 407-415
Publication date: 2013
Peer-reviewed: Yes

Publication information
Journal: IEEE Transactions on Audio, Speech and Language Processing
Volume: 21
Issue number: 2
ISSN (Print): 1558-7916
Ratings:
BFI (2019): BFI-level 2
Web of Science (2019): Indexed yes
BFI (2018): BFI-level 2
Web of Science (2018): Indexed yes
BFI (2017): BFI-level 2
Scopus rating (2017): CiteScore 4.4 SJR 0.867 SNIP 4.2
Web of Science (2017): Impact factor 2.95
Web of Science (2017): Indexed yes
Scopus rating (2016): CiteScore 3.5 SJR 0.762 SNIP 3.429
Web of Science (2016): Impact factor 2.491
Scopus rating (2015): CiteScore 2.4 SJR 1.247 SNIP 2.4
Web of Science (2015): Impact factor 1.225
Web of Science (2015): Indexed yes
Scopus rating (2014): CiteScore 4.58 SJR 1.468 SNIP 2.936
Web of Science (2014): Impact factor
Scopus rating (2013): CiteScore 4.03 SJR 1.021 SNIP 3.242
Web of Science (2013): Impact factor 2.625
ISI indexed (2013): ISI indexed no
Web of Science (2013): Indexed yes
Scopus rating (2012): CiteScore 3.05 SJR 0.918 SNIP 2.672
Web of Science (2012): Impact factor 1.675
ISI indexed (2012): ISI indexed yes
Web of Science (2012): Indexed yes
Scopus rating (2011): CiteScore 2.56 SJR 0.892 SNIP 2.743
Web of Science (2011): Impact factor 1.498
ISI indexed (2011): ISI indexed yes
Web of Science (2011): Indexed yes
Scopus rating (2010): SJR 1.127 SNIP 2.502
Web of Science (2010): Impact factor 1.668
Scopus rating (2009): SJR 1.116 SNIP 2.798
Scopus rating (2008): SJR 0.931 SNIP 2.389
Scopus rating (2007): SJR 0.613 SNIP 2.745
Web of Science (2007): Indexed yes
Scopus rating (2006): SJR 1.256 SNIP 3.301
Web of Science (2006): Indexed yes
Speech understanding in noise provided by a simulated cochlear implant processor based on matching pursuit

Speech reception is poor for cochlear implant recipients in listening environments with interfering noise. This study investigates the speech understanding provided in interfering noise by a coding strategy based on the sparse approximation algorithm matching pursuit (MP) and additionally proposes two modifications to the strategy. The levels of spectral information provided by the MP strategy and the modified MP strategy are compared to that of continuous interleaved sampling (CIS) and a strategy based on the ideal binary mask (IBM) using vocoded speech and the normalized covariance metric (NCM). We demonstrate objective intelligibility improvements in quiet, and total and partial objective intelligibility restoration in steady-state and fluctuating noise, respectively.

General information
State: Published
Organisations: Georgia Institute of Technology
Contributors: Kressner, A. A., Rozell, C. J.
Number of pages: 4
Publication date: 2013

Host publication information
Title of host publication: Proceedings of 2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics
Publisher: IEEE
ISBN (Print): 978-1-4799-0972-8/
Keywords: Cochlear implant, Speech coding, Sparse coding, Matching pursuit (MP), Continuous interleaved sampling (CIS), Normalized covariance metric (NCM)
Source: PublicationPreSubmission
Source-ID: 119928584
Research output: Research - peer-review › Article in proceedings – Annual report year: 2013

A Causal Locally Competitive Algorithm for the Sparse Decomposition of Audio Signals

While current inference methods can decompose audio signals, they require the entire signal upfront and are therefore ill-suited for real-time applications requiring causal processing. We propose a neurally-inspired, causal, sparse inference scheme based on the Locally Competitive Algorithm (LCA) over a temporal-spectral neighborhood. We demonstrate that this causal inference scheme can achieve lower sparsity levels and better signal fidelity than current filter and threshold approaches. Additionally, for some regimes, the sparsity level approaches those of Matching Pursuit while still maintaining signal integrity.

General information
ROBUSTNESS OF THE HEARING AID SPEECH QUALITY INDEX (HASQI)

Objective measures of speech quality have been the subject of significant prior work, particularly in the areas of speech codecs and communication channels for normal-hearing listeners. One of the primary concerns of researchers in this area is how these metrics generalize to datasets or listener studies which are "unknown" to the measures. Another growing concern is how these metrics perform for the hearing-impaired community. Researchers working with the this community need to be able to predict how hearing-impaired listeners will perceive the quality of speech, as well as how they will perceive the quality of speech processed specifically by hearing aids. A relatively recent metric, the Hearing Aid Speech Quality Index (HASQI), is a model-based objective measure of quality developed in the context of hearing aids for normal-hearing and hearing-impaired listeners (Kates & Arehart, Journal of the Audio Engineering Society, 2010). As such, HASQI makes substantial progress on some of the generalization issues. However, HASQI has not been tested thus far on any datasets other than the one on which it was trained. The objective of this study is to demonstrate the robustness of HASQI in predicting subjective quality. We use an "unknown" dataset of noisy speech processed by noise suppression algorithms, along with a corresponding set of subjective quality scores from normal-hearing listeners, to demonstrate HASQI's prediction performance. Furthermore, we compare HASQI's performance with that of several other objective measures in order to provide a point of reference.
15/06/2014 → 07/03/2018
Award relations: Computational speech segregation inspired by principles of auditory processing
Project: PhD