

Corpus ID: 16157310

Share This Paper



FILTER-BANK ENERGIES FOR ROBUST SPEECH RECOGNITION

K. Paliwal · Published 1999 · Computer Science

TLDR The FBEs are physically meaningful quantities and amenable for applying human auditory processing such as masking and perform at least as good as (and sometimes even better than) the MFCCs for robust speech recognition.

3 Citations

Highly Influential Citations 1
Methods Citations 1

[View All](#)

Abstract Though Mel frequency cepstral coefficients (MFCCs) have been very successful in speech recognition, they have the following two problems: 1) They do not have any physical interpretation, and 2) Liftering of cepstral coefficients, found to be highly useful in the earlier dynamic warping-based speech recognition systems, has no effect in the recognition process when used with continuous observation Gaussian density hidden Markov models. In this paper, we propose to use the filter-bank energies (FBEs) as features. The FBEs are physically meaningful quantities and amenable for applying human auditory processing such as masking. We describe procedures to decorrelate and lifter the FBEs and show that the FBEs perform at least as good as (and sometimes even better than) the MFCCs for robust speech recognition.

[Collapse](#)

[PDF](#) [maxwell.ict.griffith.edu.au](#) [Save to Library](#) [Create Alert](#) [Cite](#)

[Tables](#)[3 Citations](#)[19 References](#)[Related Papers](#)

Tables from this paper

DTW-based speech recognizer.

Type of lifter	∞	30	25	20	15
No lifter	93.0	70.6	55.9	37.2	24.0
Lin. lifter	94.0	90.8	86.6	80.1	70.6
Stat. lifter	93.9	86.7	78.3	68.3	55.3
Sin. lifter	94.3	85.3	78.9	68.7	51.5
Exp. lifter	94.3	90.1	85.1	78.9	68.1

Table 1

speech recognition performance.

Feature set	∞	30	25	20	15
Set 1	74.3	71.5	66.2	57.4	42.1
Set 2	74.5	73.5	71.5	67.6	57.7
Set 3	65.3	63.5	61.0	56.7	47.5
Set 4	76.0	75.0	71.2	65.6	55.3
Set 5	75.6	74.7	72.8	68.2	59.5
Set 6	75.8	75.6	73.8	70.1	61.7
Set 7	76.5	75.4	72.6	69.1	62.2
Set 8	69.3	69.8	69.3	66.4	58.7

Table 2

features for speech recognition.

Feature set	∞	30	25	20	15
MFCC:					
c	76.4	74.3	71.4	66.2	55.5
$c + \Delta c$	86.0	83.6	81.8	78.6	71.6
$c + \Delta c + \Delta \Delta c$	89.0	86.5	85.0	82.2	76.6
FBE:					
e	76.5	75.4	72.6	69.1	62.2
$e + \Delta e$	86.6	83.6	84.2	81.3	74.7
$e + \Delta e + \Delta \Delta e$	88.9	87.7	86.5	83.7	77.3

Table 3

3 Citations

Search authors, publication Date Range Citation Type Has PDF Author More Filters Sort by Relevance

A fuzzy approach for the equalization of cepstral variances

W.-W. Hung, Hsiao-Chuan Wang · Engineering · IEEE International Conference on Acoustics... · 2000

TLDR Experimental results for recognition of continuous Mandarin telephone speech showed that the proposed fuzzy filter bank analysis (FFBA) offers a significant improvement in the discrimination capability of cepstral features while maintaining a low computation cost. [Expand](#)

4 · [Save](#) [Alert](#)

Classification of normal and dysphagia in patients with GERD using swallowing sound analysis

Babak Basiri, M. Yali, S. Agah · Medicine · Artificial Intelligence and Signal Processing... · 2017

TLDR The experimental results show that the proposed acoustic-based method to separate dysphagia patients with reflux disorder from normal persons can provide concrete features for clinicians to diagnose dysphagia in reflux patients. [Expand](#)

2 · [Highly Influenced](#) · [View 4 excerpts, cites methods](#) [Save](#) [Alert](#)

Content-based texture image retrieval using cepstral features

Xue-Wen Chen, D. Casasent, M. Karim, M. Alam · Computer Science · SPIE Defense + Commercial Sensing · 2003

TLDR The results show that the cepstral based texture features provide an efficient and effective tool for content-based image retrieval. [Expand](#)

1 · [Save](#) [Alert](#)

19 References

Search authors, publication Date Range Citation Type Has PDF Author More Filters Sort by Relevance

On the decorrelation of filter-bank energies in speech recognition

C. Nadeu, J. Hernando, M. Gorricho · Computer Science · EUROSPEECH · 1995

TLDR A new representation is proposed that significantly outperforms both mel-cepstrum and LPC-cePstrum techniques in both recognition rate and computational cost and consists of filtering the frequency sequence of filter-bank energies with an extremely simple filter that equalizes the variance of the cepstral coefficients.[Expand](#)

👤 114 (PDF) · ≡ View 1 excerpt, references background 📌 Save 🔔 Alert

On the use of bandpass liftering in speech recognition

B. Juang, L. Rabiner, J. Wilpon · Physics · ICASSP '86. IEEE International Conference on... · 1986

TLDR This paper has found that a bandpass "liftering" process reduces the variability of the statistical components of LPC-based spectral measurements and hence it is desirable to use such a liftering process in a speech recognizer.[Expand](#)

👤 319 (PDF) · ≡ View 1 excerpt, references methods 📌 Save 🔔 Alert

Dimensionality reduction of the enhanced feature set for the HMM-based speech recognizer

K. Paliwal · Computer Science · Digit. Signal Process. · 1992

👤 45 (PDF) · ≡ View 1 excerpt, references background 📌 Save 🔔 Alert

Signal modeling techniques in speech recognition

J. Picone · Computer Science · Proc. IEEE · 1993

TLDR A tutorial on signal processing in state-of-the-art speech recognition systems is presented, reviewing those techniques most commonly used, and three important trends that have developed in the last five years in speech recognition are examined.[Expand](#)

👤 870 (PDF) · ≡ View 1 excerpt, references background 📌 Save 🔔 Alert

Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Se

S. Davis, P. Mermelstein · Computer Science · 1980

TLDR Several parametric representations of the acoustic signal were compared with regard to word recognition performance in a syllable-oriented continuous speech recognition system and the superior performance of the mel-frequency cepstrum coefficients may be attributed to the fact that they better represent the perceptually relevant aspects of the short-term speech spectrum.[Expand](#)

👤 3,259 🌟 Highly Influential (PDF) · ≡ View 1 excerpt, references methods 📌 Save 🔔 Alert

Distance measure for speech recognition based on the smoothed group delay spectrum

F. Itakura, T. Umezaki · Computer Science · ICASSP '87. IEEE International Conference on... · 1987

TLDR A novel spectral distance measure based on the smoothed LPC group delay spectrum which gives a stable recognition performance under variable frequency transfer characteristics and additive noise and gives a robust recognition rate in spite of variation in frequency characteristics and signal to noise ratio.[Expand](#)

👤 66 · ≡ View 2 excerpts, references results and methods 📌 Save 🔔 Alert

Feature decorrelation methods in speech recognition. a comparative study

E. Batlle, C. Nadeu, José A. R. Fonollosa · Computer Science · ICSLP · 1998

TLDR In this paper, various decorrelation methods for the features used in speech recognition are studied and the performance of each one is compared by running several tests with a speech database.[Expand](#)

👤 35 (PDF) · ≡ View 1 excerpt, references background 📌 Save 🔔 Alert

On the performance of the quefrency-weighted cepstral coefficients in vowel recognition

K. Paliwal · Engineering · Speech Commun. · 1982

👤 65 (PDF) · ≡ View 1 excerpt, references methods 📌 Save 🔔 Alert

Speech recognition experiments with linear predication, bandpass filtering, and dynamic programming

G. White, R. Neely · Computer Science · 1976

TLDR Automatic speech recognition experiments are described in which several popular preprocessing and classification strategies are compared and it is shown that dynamic programming is of major importance for recognition of polysyllabic words.[Expand](#)

👤 144 (PDF) · ≡ View 1 excerpt, references methods 📌 Save 🔔 Alert

Automatic Recognition of Phonetic Patterns in Speech

H. Dudley, S. Balashek · Physics · 1958

TLDR A working model of a phonetic pattern recognizer built at Bell Telephone Laboratories based on a detailed power-frequency spectrum analysis is described, which is proven adequate for limited word recognition, at least in the case of a single talker.[Expand](#)

👤 74 · ≡ View 1 excerpt, references methods 📌 Save 🔔 Alert

← 1 2 →

Related Papers

●●●●●●●●

Evaluating Gammatone Frequency Cepstral Coefficients with Neural Networks for Emotion Recognition from Speech

Gabrielle K. Liu · Computer Science · arXiv.org · 2018

Current approaches to speech emotion recognition focus on speech features that can capture the emotional content of a sp...

👤 29 📌 Save 🔔 Alert

On the inversion of Mel-frequency cepstral coefficients for speech enhancement applications

L. Boucheron, P. de Leon · Computer Science · International Conference on Signals and... · 2008

The use of Mel-frequency cepstral coefficients (MFCCs) is well established in the fields of speech processing, particula...

👤 22 📌 Save 🔔 Alert

Product of power spectrum and group delay function for speech recognition

Donglai Zhu, K. Paliwal · Computer Science · IEEE International Conference on Acoustics... · 2004

Mel-frequency cepstral coefficients (MFCCs) are the most widely used features for speech recognition. These are derived ...

👤 101 📌 Save 🔔 Alert

What Is Semantic Scholar?

Semantic Scholar is a free, AI-powered research tool for scientific literature, based at the Allen Institute for AI.

[Learn More](#)

About

[About Us](#)
[Meet the Team](#)
[Publishers](#)
[Blog](#)
[AI2 Careers](#)

Product

[Product Overview](#)
[Beta Program](#)
[Semantic Scholar API](#)
[Semantic Reader](#)

Research

[Publications](#)
[Researchers](#)
[Research Careers](#)
[Resources](#)

Help

[FAQ](#)
[Librarians](#)
[Tutorials](#)
[Contact](#)

Proudly built by AI2

[Collaborators & Attributions](#) • [Terms of Service](#) • [Privacy Policy](#) • [API License Agreement](#)

AI2 Allen Institute for AI