

Myung-jae Kim, Il-ho Yang, Min-seok Kim, Ha-jin Yu, 2017. Histogram equalization using a reduced feature set of background speakers' utterances for speaker recognition. *Frontiers of Information Technology & Electronic Engineering*, **18**(5):738-750. <http://dx.doi.org/10.1631/FITEE.1500380>

Histogram equalization using a reduced feature set of background speakers' utterances for speaker recognition

Key words: Speaker recognition; Histogram equalization; i-vector

Corresponding author: Ha-jin Yu

E-mail: hjyu@uos.ac.kr

 ORCID: <http://orcid.org/0000-0003-3657-0665>



Motivation

The length of input utterances to perform normalization using histogram equalization affects the performance of speaker recognition system.

In particular, histogram equalization gives low performance in speaker recognition when the length of an utterance is very short.



Main idea

Speech features of the background set is used to estimate robust CDFs

To increase the system performance and to decrease the system processing time, the large background set is made into a small background set which represents distribution of the large background set.

Method

1. Histogram equalization as a feature normalization method (Segura *et al.*, 2004).
2. Greedy selection algorithm for greedy kernel PCA (Franc, 2005).
3. We propose a novel approach for histogram equalization using a reduced background speaker feature set (Eq. 25 ~ Eq. 32).

Major results

Proposed method improves the speaker recognition system.

Table 2 Results of the male set evaluated on the dets 6 and 7 conditions of the NIST 2008 SRE short2-short3 condition

Method	EER		DCF	
	det 6	det 7	det 6	det 7
Baseline	5.95%	2.93%	0.0353	0.0180
+ EFR	5.61%	2.79%	0.0329	0.0154
+ SphNorm	5.85%	2.91%	0.0332	0.0158
Rank norm	5.97%	2.95%	0.0351	0.0169
+ EFR	5.25%	2.31%	0.0295	0.0158
+ SphNorm	5.39%	2.31%	0.0301	0.0157
Proposed ($K=500$)	5.50%	2.91%	0.0326	0.0161
+ EFR	5.14%	2.71%	0.0279	0.0151
+ SphNorm	5.11%	2.30%	0.0282	0.0154
Proposed ($K=1000$)	5.68%	2.75%	0.0329	0.0159
+ EFR	5.03%	2.55%	0.0276	0.0154
+ SphNorm	5.13%	2.48%	0.0281	0.0158

The numbers in bold represent the best performance achieved by the combining methods

Major results (Cont'd)

Proposed method reduces the processing time

Table 3 Processing time of 100 utterances of the rank normalization and the proposed method

Method	Processing time (s)
Rank norm	8.378
Proposed ($K=500$)	0.621
Proposed ($K=1000$)	0.691

Conclusions

A novel approach for histogram equalization method

Normalization method for short speech

Improve speaker recognition system performance

Reducing the background feature set

Reducing processing time

Various methods are applied

Greedy selection method shows the best performance