

# A novel method for estimating the number of speakers based on generalized eigenvalue-vector decomposition and adaptive wavelet transform by using K-means clustering

Por: [Dehghan Firoozabadi, A](#) (Dehghan Firoozabadi, Ali)<sup>[1]</sup>; [Irarrazaval, P](#) (Irarrazaval, Pablo)<sup>[2,3,4]</sup>; [Adasme, P](#) (Adasme, Pablo)<sup>[5]</sup>; [Zabala-Blanco, D](#) (Zabala-Blanco, David)<sup>[6]</sup>; [Azurdia-Meza, C](#) (Azurdia-Meza, Cesar)<sup>[7]</sup>

[Ver número de ResearcherID y ORCID de Web of Science](#)

## SIGNAL IMAGE AND VIDEO PROCESSING

DOI: 10.1007/s11760-020-01634-2

Acceso anticipado: JAN 2020

Tipo de documento: Article; Early Access

[Ver impacto de la revista](#)

### Abstract

The aim of this article is estimating the number of simultaneous speakers from the overlapped speech signals. The percentage of correct number of speakers is an important factor for the proposed algorithm. The proposed method in this article is based on spectrum estimation by using the adaptive wavelet transform in combination with generalized eigenvalue-vector decomposition (GEVD) and K-means clustering. Firstly, the speech signals are obtained by a uniform circular array, and each adjacent microphone pairs are considered for the processing. Then, the spectral estimation method is implemented on all microphone signals to select the best part of the speech spectrum. Next, the microphone signals are divided into different subbands by using adaptive wavelet transform. The GEVD algorithm is implemented on each microphone pairs in different subbands and time frames to estimate the room impulse response and time difference of arrival (TDOA). Finally, the K-means clustering with silhouette criteria is used to estimate the number of speakers (K value). The proposed algorithm is implemented on simulated and real data to show the superiority of proposed method in comparison with PENS, Bessel, i-vector PLDA, Hilbert envelope and DNN-based method. The proposed scheme outperforms the other evaluated schemes by 18% in terms of correct estimations in noisy-reverberant conditions for five simultaneous speakers.

### Palabras clave

Palabras clave de autor: [Adaptive filters](#); [Eigenvalue-vector decomposition](#); [Speaker counting](#); [Spectral estimation](#); [Wavelet transform](#)

### Información del autor

Dirección para petición de copias: Dehghan Firoozabadi, A (autor para petición de copias)

Univ Tecnol Metropolitana, Dept Elect, Ave Jose Pedro Alessandri 1242, Santiago 7800002, Chile.

#### Direcciones:

- [ 1 ] Univ Tecnol Metropolitana, Dept Elect, Ave Jose Pedro Alessandri 1242, Santiago 7800002, Chile
- [ 2 ] Pontificia Univ Catolica Chile, Dept Elect Engn, Santiago 7820436, Chile
- [ 3 ] Pontificia Univ Catolica Chile, Biomed Imaging Ctr, Santiago 7820436, Chile
- [ 4 ] Pontificia Univ Catolica Chile, Inst Biol & Med Engn, Santiago 7820436, Chile
- [ 5 ] Univ Santiago Chile, Dept Elect Engn, Ave Ecuador 3519, Santiago 9170124, Chile
- [ 6 ] Univ Catolica Maule, Dept Comp & Ind, Talca 3466706, Chile
- [ 7 ] Univ Chile, Dept Elect Engn, Santiago 8370451, Chile

**Direcciones de correo electrónico:** [adehghanfirouzabadi@utem.cl](mailto:adehghanfirouzabadi@utem.cl)

#### Financiación

Entidad financiadora <a href="#">Mostrar más información</a>	Número de concesión
Comision Nacional de Investigacion Cientifica y Tecnologica (CONICYT) CONICYT FONDECYT	3190147 11180107 11160517

[Ver texto de financiación](#)

#### Editorial

SPRINGER LONDON LTD, 236 GRAYS INN RD, 6TH FLOOR, LONDON WC1X 8HL, ENGLAND

#### Información de la revista

- **Impact Factor:** [Journal Citation Reports](#)

#### Categorías / Clasificación

**Áreas de investigación:** Engineering; Imaging Science & Photographic Technology

**Categorías de Web of Science:** Engineering, Electrical & Electronic; Imaging Science & Photographic Technology

#### Información del documento

**Idioma:** English

**Número de acceso:** WOS:000510294800002

**ISSN:** 1863-1703

**eISSN:** 1863-1711