

# Aplikasi Metode Independent Component Analysis untuk Pemisahan Sinyal Wicara dengan Backsound pada Audio Movie

Titon Dutono<sup>1</sup>, Dinda Ayu Oktaviasari<sup>2</sup>, Tri Budi Santoso<sup>3</sup>

<sup>1,2,3</sup>Teknik Elektro, Politeknik Elektronika Negeri Surabaya

Jln. Raya ITS, Sukolilo, Surabaya, 60111

Email: <sup>1</sup> titon@pens.ac.id, <sup>3</sup>tribudi@pens.ac.id

*Abstract- Voice analysis in establishing speech and non-speech is carried out by audio signal processing from the sound source. This research focuses on mixed voice data in the form of actor voices (speech signal) and background voices (backsound) from a movie. In this study, sound signal separation has been carried out using the Blind Source Separation (BSS) method to separate mixed signals into a number of forming signals without information about the number of signal sources or the process of mixing these signals. The algorithm used for BSS in this study is the Independent Component Analysis (ICA) algorithm. The actor's voice signal that has been separated from the background sound in a film was tested using MSE and SIR analysis to determine the quality of the separation signal. From the Mean Square Error (MSE) test of the speech signal from the separation results, a test value of 5 seconds was obtained of 0.036, for a duration of 10 seconds the MSE value was 0.0432, and in the data of 15 seconds it produced an MSE value of 0.0558. Signal to Interference Ratio (SIR) analysis of speech signals resulting from separation, for data with a duration of 5 seconds of 14,485, for data measuring 10 seconds obtained SIR of 13,645, and for data with a duration of 15 seconds obtained SIR of 12,533. The output of this sorting process is then used as an input for other necessary signal processing processes.*

**Abstrak-** Analisis suara dalam menetapkan speech dan non-speech dilakukan dengan audio signal processing dari sumber suara. Penelitian ini berfokus pada data suara campuran berupa suara aktor (speech signal) dan suara latar (backsound) dari sebuah film. Pada penelitian ini telah dilakukan pemisahan sinyal suara dengan menggunakan metode Blind Source Separation (BSS) untuk memisahkan sinyal tercampur menjadi sejumlah sinyal pembentuknya tanpa informasi mengenai jumlah sumber sinyal atau proses tercampurnya sinyal-sinyal tersebut. Algoritma yang digunakan untuk BSS dalam penelitian ini adalah algoritma Independent Component Analysis (ICA). Sinyal suara aktor yang telah terpisah dari suara latar pada sebuah film diuji menggunakan analisa MSE dan SIR untuk mengetahui kualitas sinyal hasil pemisahan. Dari pengujian Mean Square Error (MSE) terhadap sinyal speech hasil pemisahan, diperoleh nilai pengujian pada durasi 5 detik sebesar 0.036, untuk durasi 10 detik nilai MSE yakni 0.0432, dan pada data sebesar 15 detik menghasilkan nilai MSE sebesar 0.0558. Analisa Signal to Interference Ratio (SIR) terhadap sinyal wicara hasil pemisahan, untuk data berdurasi 5 detik sebesar 14.485, untuk data berukuran 10 detik diperoleh SIR sebesar 13.645, dan untuk data dengan berdurasi 15 detik diperoleh SIR sebesar 12.533. Luaran dari proses pemilahan ini selanjutnya digunakan sebagai input bagi proses pengolahan sinyal lainnya yang diperlukan.

**Kata Kunci:** Sinyal Wicara, Backsound, BSS, ICA.

## I. PENDAHULUAN

Ucapan merupakan cara paling alami untuk pertukaran informasi. Dengan demikian, merancang mesin cerdas yang dapat mengenali informasi lisan telah menjadi topik penelitian bagi para ilmuwan dan insinyur selama lebih dari lima dekade [1]. Hal ini dapat digunakan secara efisien dalam berbagai keseharian untuk meningkatkan lingkungan kerja, atau untuk memecahkan masalah nyata dalam kehidupan seperti membuat teknologi modern yang dapat diakses dalam kegiatan sehari-hari bagi orang-orang yang menderita cacat fisik. Proses berbicara dapat terjadi jika manusia mampu mengeluarkan sinyal akustik yang disebut dengan suara. Suara manusia dihasilkan dengan adanya interaksi antara organ-organ pembentuk sinyal. Klasifikasi suara menjadi wicara dan non-wicara memberikan segmentasi akustik awal untuk aplikasi pengolahan ucapan seperti pengenalan wicara.

Masalah pemisahan sumber sinyal audio dari campuran suara merupakan masalah penting yang muncul di berbagai aplikasi industry. Sebagai contoh, permasalahan tersebut dapat menjadi pembahasan pada sound based fault detection and diagnosis di dalam perangkat industry atau suatu multi-speaker recognition didalam suatu cocktail party problem. Satu hal yang sangat penting untuk kreasi pada generasi baru teknologi hearing aid yang mengisolasi dan menguatkan suatu sinyal wicara tertentu di dalam suasana bising. Pada paper ini, disusulkan suatu pendekatan baru untuk menyelesaikan permasalahan, didasarkan pada ansamble jaringan syaraf tiruan. Penyelesaian masalah dibagi dalam dua tahapan. Pada tahapan pertama, ensemble pada convolutional neural network menentukan keberadaan atau tidaknya sinyal wicara di dalam suatu lingkungan bernois, dengan menggunakan suatu set sampel-sampel sinyal wicara, dipersiapkan secara benar. Pada tahapan kedua, ensemble jaringan syaraf yang lain memfilter sinyal wicara, yang telah ditetapkan pada tahap pertama, dan memotong sisa sinyal lainnya dan diperlakukan sebagai noise. Ansambel jaringan saraf convolutional, yang digunakan pada tahap pertama, terdiri dari jaringan saraf, yang masing-masing mencakup tiga lapisan convolutional dan satu lapisan yang terhubung penuh. Analisis suara dilakukan berdasarkan spektogramnya, yang diperoleh dengan menggunakan FFT. Jaringan saraf ini diimplementasikan dengan Python dengan menggunakan pustaka perangkat lunak TensorFlow dan Keras. Hasil eksperimen komputasi dalam menggunakan jaringan saraf yang dirancang dan dilatih untuk menganalisis dan

memfilter aliran audio, yang berisi beberapa suara pria dan wanita yang dilapiskan dengan musik di latar belakang. [2]

Pemisahan sinyal wicara dari music background sangat penting di dalam berbagai aplikasi seperti speaker identification, speaker specific information retrieval, word recognition, dsb., yang mana dalam hal ini background music dipertimbangkan sebagai noise. Walaupun pemisahan sinyal wicara telah dipelajari secara luas dalam beberapa tahun ini, tetapi belum mampu menunjukkan kinerja yang cukup di dalam pemisahan sinyal voice ayau sinyal wicara dari background musical noise. Pada penelitian ini disajikan suatu usulan sebuah system untuk memisahkan sinyal wicara dari background music. Sistem yang disusun terdiri dari dua tahapan, yang memodifikasi nonnegative matrix factorization (NMF) untuk melakukan dekomposisi masukan ke mixture spectrogram. Diskontinuitas thresholding diaplikasikan pada mixture spectrogram untuk memilih komponen-komponen luaran NMF. Diskontinuitas ini dipertimbangkan dalam arah temporal (waktu) dan spectral (frekuensi) [3].

Paper ini menyajikan suatu metode baru untuk mengekstraksi vocal tract dari suatu campuran music, yang tersusun dari suatu suara penyanyi dan sebuah back track yang berasal dari beragam instrument. Di sini metode convolutional network digunakan dengan skip and residual connections sebaik model dilated convolutions untuk mengestimasi parameter-parameter vocoder, dan memberikan spectrogram pada suatu campuran masukan. Parameter-parameter yang diestimasi selanjutnya digunakan untuk mensintesa vocal track, tanpa suatu interferensi dari backing track (suara latar). Sistem telah dievaluasi melalui metrik objective yang relevan untuk kualitas audio dan interferensi dari sumber-sumber background, dan melalui suatu evaluasi comparative subjective. Open-source source separation yang didasarkan pada metode Non-negative Matrix Factorization (NMFs) dan Deep Learning methods sebagai pembandingan system ini, dan membahas aplikasi ke depan untuk algoritma-algoritma tersebut. [4]

Sebuah penelitian untuk menghasilkan perbaikan kinerja sinyal wicara sehingga dapat didengarkan secara jelas bagi penderita gangguan pendengaran telah disajikan [2-baru]. Sebagai Langkah awal telah disusun melalui algoritma DUET Blind Source Separation (BSS) yang telah dimodifikasi agar secara otomatis mampu mengekstrak sinyal wicara dalam suasana auditory scene [5]. Tetapi system ini belum dapat diwujudkan secara real time, dan masih dalam tahapan awal sebuah simulasi di mana proses mixing dilakukan secara sempurna.

Identifikasi sumber akustik secara akurat, untuk mendapatkan informasi tentang noise secara tepat telah dipaparkan dengan memanfaatkan blind source separation [6]. Dengan pengkondisian bahwa mesin sebagai sumber suara dalam kondisi constant, dan mendapatkan suatu model campuran sinyal dengan multi-source noise masih memungkinkan untuk teridentifikasi. Penelitian ini telah menunjukkan hasil bahwa luaran model BSS yang disajikan mampu memberikan kelayakan suara hasil pemisahan dengan bagus.

Sebuah pemisahan sinyal akustik dari noise non Gaussian telah diujikan dengan model data pengukuran akustik bawah air [7]. Proses pemisahan pemisahan sinyal akustik dilakukan dengan menggunakan Natural Gradient ICA berdasarkan

Generalized Gaussian Model yang didapat dari karakteristik distribusi sinyal akustik non-gaussian yakni ship radiated noise dan sea ambient. Pemisahan sinyal akustik dilakukan sebanyak tiga kali yakni dengan simulasi toolbox ICALABS V3 dan menggunakan pemisahan sinyal akustik dari data riil pengukuran. Dari hasil simulasi menunjukkan pemisahan dengan algoritma Natural Gradient ICA berdasarkan Generalized Gaussian Model berjalan dengan baik dengan nilai SIR yakni shrimp.wav 48,995 dB dan SIR ferry.wav 46,931 dB.

Penelitian tentang pemisahan sinyal suara juga telah dilakukan pada berbagai topik, diantaranya pemisahan sinyal akustik bawah laut dan pemisahan sinyal pada seni geguntangan [8]. Penelitian tersebut bertujuan memisahkan suara vokal dengan gamelan pada geguntangan yang merupakan pesantian dalam upacara keagamaan. Algoritma SCA dan ICA digunakan untuk proses pemisahan suara dengan parameter nilai yang digunakan adalah Mean Square Error (MSE) dan Signal to Interference Ratio (SIR). Dari hasil simulasi menunjukkan Hasil perhitungan MSE dan SIR didapatkan hasil untuk metode ICA, nilai MSE adalah  $3.60 \times 10^{-5}$  untuk vokal dan  $1.71 \times 10^{-6}$  untuk instrumen, memiliki SIR 44.848 dB untuk vokal dan 59.897 dB untuk instrumen. Untuk metode SCA nilai MSE adalah  $1.25 \times 10^{-4}$  untuk vokalnya dan  $4.21 \times 10^{-6}$  untuk instrumennya, memiliki SIR 40.009 dB untuk vokal dan 54.708 dB untuk instrumen.

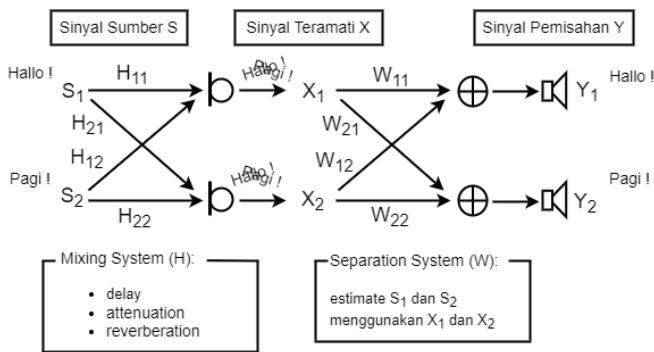
Sebuah kerangka kerja untuk mengaplikasikan ICA pada proses denoise side-channel measurements dengan tujuan untuk mereduksi kompleksitas proses dalam mengatasi key recovery attacks. Hasil penelitian ini menunjukkan bahwa dalam beberapa kasus yang telah dilakukan, dapat ditunjukkan bahwa system ICA memiliki kelebihan dibandingkan dengan berbagai system yang sudah umum digunakan, yaitu singular spectrum analysis [9]. Hal ini menunjukkan bahwa ICA merupakan model yang menarik dan masih terbuka di dalam proses penyelesaian masalah noise. Sementara itu di sisi lain pengembangan model untuk ekstraksi ciri dengan memanfaatkan model principal component Analysis juga sudah dikembangkan [12]. Dan penelitian lain yang menarik tentang pengenalan audio berbasis hardware (Arduino) juga telah dikembangkan [13], sedangkan untuk model pengenalan suara berbasis metode HMM juga telah dipaparkan pada [14]. Ketiga penelitan ini semakin menguatkan bahwa model independent komponen analysis merupakan Teknik yang menantang untuk pengembangan di bidang audio, khususnya pengolahan sinyal wicara.

Sinyal suara yang ada pada audio film merupakan gabungan dari beberapa sumber suara yang tidak diketahui jumlahnya. Untuk melakukan penelitian pengolahan suara seperti speech recognition dari sebuah film, data pengujian yang dibutuhkan ialah sinyal wicara dari audio film tersebut. Terlepas dari kebutuhan, permasalahan yang dihadapi dari data audio film ialah terdapatnya suara latar seperti suara musik ataupun efek-efek yang ditambahkan dalam sebuah audio film. Pada paper ini akan disajikan suatu metode Blind Source Separation, pemisahan suara dari suatu sinyal blind pada audio film dapat diatasi. Metode BSS selanjutnya dikombinasikan dengan algoritma Independent Component Analysis (ICA) yang memisahkan sejumlah sinyal independen dari sinyal campuran.

II. KONSEP PEMISAHAN SINYAL DARI NOISE

A. Blind Source Separation

Metode *Blind Source Separation* digunakan untuk memperkirakan sinyal asli dengan cara melakukan pemisahan satu set sinyal asli dari satu set sinyal campuran (sinyal observasi), dengan tanpa adanya informasi (atau dengan sangat sedikit informasi) tentang sumber sinyal ataupun proses pencampurannya. Pemisahan sinyal bergantung pada asumsi bahwa sumber sinyal tidak berkorelasi satu dengan yang lainnya. Secara sederhana metode BSS dapat diasjikan seperti pada Gambar 1 berikut ini:



Gambar. 1 Diagram blok model BSS

Pada kondisi ini diasumsikan bahwa sumber sinyal  $S_1$  dan  $S_2$  bersifat saling independen. Asumsi ini biasanya berlaku di dunia nyata. Sedangkan sinyal suara yang didapat dari hasil observasi berupa sinyal suara  $X_1$  dan  $X_2$ . Untuk mendapatkan sinyal output  $Y_1$  dan  $Y_2$  yang saling independen dilakukan dengan cara mengekstraksi sinyal observasi menggunakan unmixing matrix  $W_{ij}$ . Operasi seperti ini tidak memerlukan informasi apapun pada posisi atau periode keberadaan sumber, serta tidak memerlukan informasi tentang proses pencampurannya (mixing) yang dinotasikan dengan mixing matriks  $H_{ij}$ . Proses seperti ini dinamakan *Blind Source Separation* (BSS) [10].

B. Independent Component Analysis

Permasalahan yang sering ditemui pada BSS adalah bagaimana menemukan representasi linier yang komponen-komponennya independen secara statistic. Pada kondisi real, hal ini akan sulit ditemukan komponen yang benar-benar saling independent, tetapi minimal kita dapat mengupayakan bagaimana sebuah komponen mendekati kondisi independent seperti yang diperlukan.

Independent Component Analysis (ICA), diperkenalkan pada tahun 1986 oleh Jeanny Hérault dan Christian Jutten sebagai Neural Network berdasarkan Hebb learning law yang mampu melakukan pemisahan blind signal. Secara khusus, Algoritma ICA mencoba untuk memisahkan sejumlah sinyal independen dari jumlah input sinyal yang sama adalah jumlah linier yang pertama [11]. Dua asumsi dasar agar system dapat diberi perlakuan dengan system ICA adalah bahwa Sebagian besar pada subcomponent sinyal memiliki pola distribusi non-Gaussian dan secara statistic ssinyal sinyal tersebut memiliki sifat saling independent satu dengan yang lain.

Pada bagian awal proses ICA, perlu dilakukan pengolahan pendahuluan (*pre-processing*) agar dapat berjalan dengan baik, yaitu *centering* dan *whitening*.

*Centering* adalah proses pemusatan data yang membuat nilai sinyal campuran ( $X$ ) menjadi sinyal campuran yang memiliki rerata nol atau *zero Mean* ( $D$ ). Hal ini ditujukan untuk memusatkan data dengan mengurangi rerataan dari semua sinyal. Diberikan  $n$  sinyal campuran ( $X$ ), rata-ratanya adalah  $1$  dan langkah pemusatan dapat dihitung sebagai berikut:

$$D = X - \mu = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} = \begin{pmatrix} x_1 - \mu \\ \vdots \\ x_n - \mu \end{pmatrix} \quad (1)$$

Di mana  $D$  adalah sinyal campuran setelah langkah pemusatan dan  $\mu \in R^{1 \times N}$  adalah rata-rata dari semua sinyal campuran. Vektor rata-rata dapat ditambahkan kembali ke komponen independen setelah menerapkan ICA.

*Whitening* adalah proses untuk memutihkan variabel yang diamati, dalam hal ini adalah sinyal campuran. Dari proses *whitening* didapatkan vektor sinyal campuran baru yang memiliki varian sama dengan satu ( $z$ ). Hal ini ditujukan untuk memutihkan data yang berarti mengubah sinyal menjadi sinyal yang tidak berkorelasi dan kemudian mengubah skala setiap sinyal menjadi dengan varian satuan.

Proses estimasi dilakukan untuk mengetahui sejauh mana hasil pemisahan sinyal bisa berjalan dengan baik. Dua cara yang cukup populer dalam hal ini adalah dengan memanfaatkan *Mean Square Error* (MSE) dan melihat *Signal to Interference Ratio* (SIR).

MSE merupakan suatu metode untuk mengukur perbedaan antara estimator (sinyal rekonstruksi) dan nilai sebenarnya (sinyal *baseline*) dari kuatitas yang diperkirakan. Secara garis besar dengan menghitung nilai MSE, maka akan didapatkan selisih pergeseran yang diperoleh antara sinyal asli dan sinyal rekonstruksi., yang dapat ditunjukkan pada persamaan berikut. Variabel yang dibutuhkan untuk pengujian MSE adalah jumlah data, sinyal asli, dan sinyal estimasi.

$$MSE = \frac{1}{n} \int_1^t (S - S_i)^2 dt \quad (2)$$

Variabel yang dibutuhkan untuk pengujian MSE adalah jumlah data, sinyal asli, dan sinyal estimasi.

SIR adalah perbandingan daya sinyal terhadap daya interferensi dan digunakan untuk menilai kualitas sinyal terhadap gangguan interferensi akibat penggabungan sinyal. Interferensi adalah gangguan selain *noise* yang dapat menyebabkan kualitas suatu sinyal menurun. Semakin tinggi nilai SIR, maka kualitas sinyal semakin baik, begitu juga sebaliknya. Variabel yang dibutuhkan untuk pengujian SIR adalah nilai dari MSE. Rumus untuk mencari nilai SIR adalah sebagai berikut:

$$SIR = -10 \log_{10}(MSE) \quad (3)$$

III. PEMODELAN SISTEM

Sebelum memulai dengan proses pemisahan sinyal dari background suara, perlu dilakukan langkah pendahuluan berupa pengumpulan data suara yang akan diproses, dan melakukan perubahan format dari file MP4 menjadi file WAV.

A. Pengumpulan Data

Pengumpulan data dimulai dari ekstraksi audio dari sebuah film hingga pemilahan sinyal campuran. Data yang digunakan yakni file suara dengan durasi masing-masing 5 detik, 10 detik, dan 15 detik. Dalam hal ini, film yang

digunakan yakni sebuah film Indonesia yang telah ditentukan dengan judul “Bumi Manusia”. Film Bumi Manusia dipilih karena dari sisi cerita dan latar yang menarik serta terdapat banyak adegan menggunakan percakapan dan disertai background suara pendukung, hal tersebut memudahkan dalam melakukan pengambilan data.

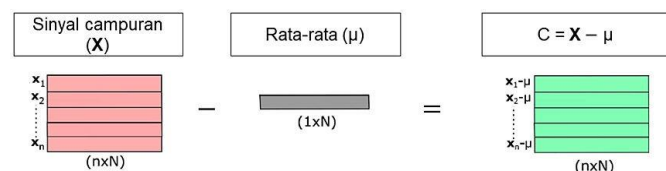
Proses pertama yang dilakukan dalam mengumpulkan data adalah mengambil suara dari file video dengan format MP4 menjadi file suara dengan format WAV. Proses ini menggunakan software AoA Audio Extractor dalam mengambil suara pada sebuah film. Sebelum dilakukan proses konversi dari film menjadi audio dengan format WAV file, terlebih dahulu mengubah format file film menjadi MP4, karena software AoA Audio Extractor tidak dapat mendukung proses konversi dalam format MKV. Berikut merupakan proses pengubahan format MKV menjadi MP4 menggunakan tools Wondershare UniConverter yang telah terinstall pada PC. Setelah file suara telah menjadi format MP4, kemudian proses konversi menggunakan AoA Audio Extractor dapat dilakukan. Dalam hal ini durasi yang dipilih untuk data pertama yakni dari 00:23:51 hingga 00:24:25, untuk Audio Bitrate digunakan 128kbps, Audio SampleRate sebesar 48000, dan dengan Channel Stereo.

Proses kedua adalah pemilahan sinyal campuran. Sinyal suara yang didapatkan dari proses sebelumnya untuk selanjutnya dipilah dengan cara diambil hanya pada bagian-bagian yang mempunyai campuran antara sinyal wicara dan backsound. Proses konversi dilanjutkan untuk mengubah sinyal dengan sample rate sebesar 8000Hz, sample encoding Lin32, serta mengubah kanal menjadi mono.

**B. Pre-Processing ICA**

Seperti telah dipaparkan pada Bab 2, bahwa ada dua pre-processing ICA, yaitu centering dan whitening. Kedua proses ini dilakukan menggunakan pemrograman python yang terdapat pada fungsi centering dan whitening.

Hal mendasar yang perlu dilakukan dalam proses Centering adalah dengan memusatkan data x. Prinsipnya mengurangi data tersebut dengan rerata data dari vektor yang ada, sehingga x menjadi variabel bererata nol (*zero-mean*). Pemusatan data ini semata-mata untuk membuat algoritma ICA menjadi mudah dan cepat. Setelah melakukan taksiran matriks pencampur A dengan data memusat, kita melengkapi taksiran tersebut dengan menambahkan vektor rerata kepada hasil taksiran yang memusat. Secara sederhana dapat diberikan pada Gambar 2 berikut ini.



Gambar 2. Ilustrasi proses Centering

Whitening berfungsi untuk me-mutih-kan variabel yang diamati. Sehingga didapatkan sebuah vektor baru yang variansnya sama dengan satu serta komponen nyata dan imajinerinya tidak terkorrelasi dengan nilai varians yang sama. Secara singkat, matriks kovarians dari  $\check{x}$  sama dengan matriks identitas (orthogonal) dan digambarkan dalam

persamaan berikut:

$$E\{\check{X}\check{X}^T\} = I \tag{4}$$

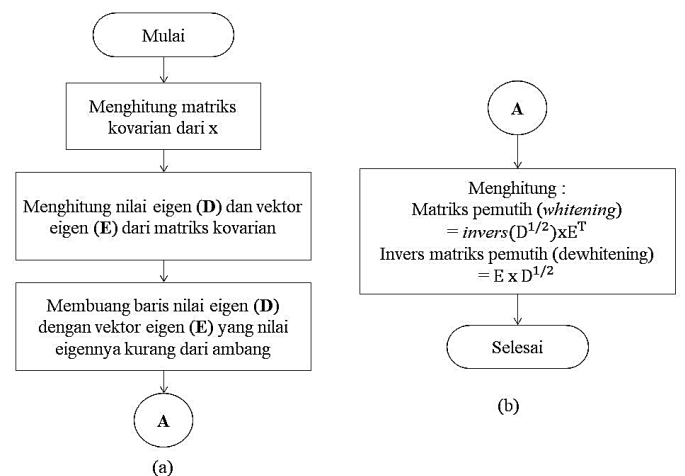
Metode yang sering digunakan adalah *eigenvalue decomposition* (EVD) dari matriks kovarians  $E\{xx^T\}$ . Pada tahap ini, terlebih dahulu mengamati nilai dan vektor *eigen* dari  $E\{xx^T\}$ , seperti yang dilakukan pada *Principal Component Analysis* (PCA). Persamaan untuk mengamati nilai dan vektor eigen dirumuskan sbb:

$$x\lambda = D\lambda \tag{5}$$

Dengan x menunjukkan matriks yang akan dicari nilai dan vektor *eigennya*. Dalam hal ini D merupakan matriks nilai eigen berbentuk diagonal dan  $\lambda$  adalah vektor *eigen* dari matriks x. Untuk mendapatkan Persamaan 4, dilakukan pemutihan isyarat yang dapat dilakukan dengan persamaan berikut.

$$\check{X} = D^{-1/2}\lambda^T x \tag{6}$$

Di mana nilai  $D^{-1/2}\lambda^T$  disebut sebagai matriks *whitening* dan untuk mengembalikan proses disebut *invers matriks whitening (dewhitening)*. Proses whitening ini ditunjukkan seperti pada gambar 3 berikut.



Gambar 3. Diagram alir proses whitening (a) Fungsi PCA, (b) Fungsi whitening

**C. Proses ICA**

Setelah data mengalami praproses yaitu pemusatan dan pemutihan. Kemudian data diolah dengan suatu metode yang efisien yang disebut *FastICA*. Algoritma *FastICA* dilakukan setiap iterasi yang ditunjukkan sebagai berikut:

*Langkah pertama* adalah memperkirakan satu unit komponen bebas, yang dilakukan dengan langkah sbb:

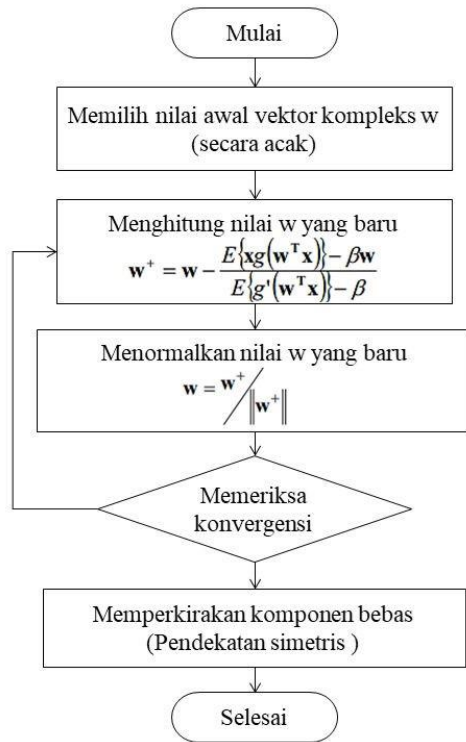
- Memilih sebuah nilai awal vektor kompleks w, dapat secara acak
- Menghitung nilai w yang baru dengan persamaan berikut:

$$w^+ = w - \frac{E\{xg(w^T x)\} - \beta w}{E\{g'(w^T x)\} - \beta} \tag{6}$$

di mana fungsi g pada langkah ini adalah fungsi nonlinearitas.

- Menormalkan nilai w yang baru  $w = w^+ / \|w^+\|$

- Memeriksa konvergensi, bila tidak konvergen maka kembali ke langkah b. Konvergen pada langkah d berarti bahwa nilai  $w$  yang baru dan yang lama mempunyai arah yang sama, yaitu perkalian titik diantaranya sama dengan 1. Konvergensi dapat tercapai jika galat yang terbentuk bernilai di bawah ambang batas nilai yang dimasukkan. Untuk mempermudah maka didefinisikan *epsilon* yang merupakan nilai kesalahan terbesar (toleransi) dari nilai tersebut, pada tugas akhir ini sebesar  $10^{-4}$ .



Gambar 4. Diagram alir algoritma FastICA

Langkah Kedua adalah memperkirakan beberapa komponen bebas. Hal ini dilakukan dengan menjalankan algoritma FastICA untuk satu unit di muka untuk masing-masing komponen bebas dengan vektor  $w_1, \dots, w_n$ . Setelah itu dilakukan dekorelasi untuk masing-masing output  $w$  setiap iterasi, untuk menjaga agar vektor yang berbeda tidak konvergen ke besaran yang sama. Pendekatan ini berarti memperkirakan komponen bebas yang bersangkutan satu persatu. Ketika telah diperkirakan 2 komponen bebas, atau 2 vektor  $w(w_1, w_2)$  lalu menjalankan algoritma FastICA maka setiap akhir proses iterasi dilakukan proses penghitungan seperti persamaan berikut.

$$w^+ = w + \text{diag}(\alpha_i)[\text{diag}(\beta_i) + E\{g(y)y^T\}]w \quad (8)$$

Di mana :

$$y = Wx, \beta_i = E\{y_i(y_i)\}$$

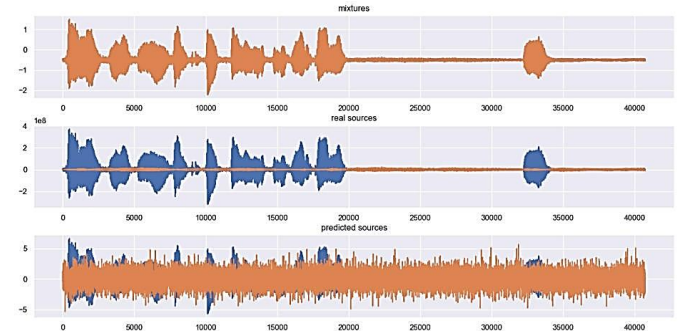
$$\alpha_i = -\frac{1}{\beta_i - E\{g^t(y_i)\}}$$

Matriks  $W$  selalu ortogonal pada akhir setiap langkah. Secara sederhana algoritma FastICA ditunjukkan pada diagram alir pada Gambar 4.

#### IV. HASIL PENGUJIAN DAN ANALISA

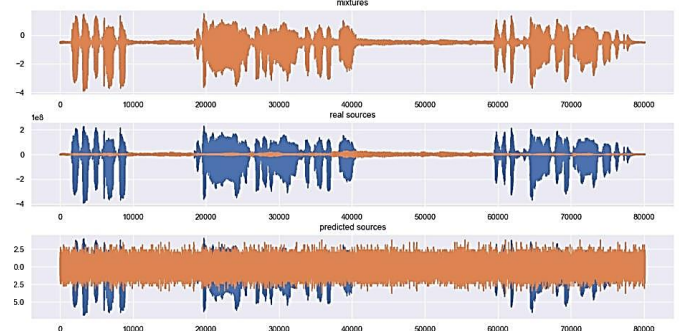
##### A. Pemisahan Sinyal dari Backsound

Hasil pemisahan sinyal dengan menggunakan simulasi python ditunjukkan pada gambar 5 dengan durasi data 5 detik. Sinyal dengan warna merah menunjukkan backsound, sedangkan sinyal dengan warna biru menunjukkan sinyal wicara.



Gambar 5. Pemisahan sinyal durasi 5 detik

Pada gambar 6 menunjukkan hasil pemisahan sinyal dengan durasi 10 detik, yang mana diawali dari proses pencampuran sinyal hingga sinyal wicara terpisah dari backsound



Gambar 6. Pemisahan sinyal durasi 10 detik

Hasil pemisahan sinyal dengan durasi 15 detik menunjukkan bahwa backsound hasil pemisahan berukuran besar karena noise. Hal ini bukan menjadi masalah karena yang dibutuhkan adalah hasil pemisahan dari sinyal wicara.

##### B. Pengujian MSE dan SIR

Dalam pengujian ini, sinyal rekonstruksi merupakan sinyal wicara hasil pemisahan dengan menggunakan metode ICA, sedangkan untuk sinyal estimator yaitu sinyal wicara hasil pemisahan dengan menggunakan Spleeter GUI. Signal to Interference Ratio (SIR) adalah perbandingan daya sinyal terhadap daya interferensi dan digunakan untuk menilai kualitas sinyal terhadap gangguan interferensi akibat penggabungan sinyal.

##### Pengujian Mean Square Error

Pada pengujian nilai mean square error dilakukan terhadap hasil pemisahan sinyal dengan metode ICA. Secara garis besar dengan menghitung nilai MSE, maka akan didapatkan selisih pergeseran yang diperoleh antara sinyal asli dan sinyal rekonstruksi (sinyal hasil pemisahan). Variabel yang dibutuhkan untuk pengujian MSE adalah jumlah data, sinyal asli, dan sinyal hasil pemisahan.



Pada pengujian ini digunakan fungsi matlab yaitu *immse* yang digunakan untuk menghitung nilai *MSE* terhadap kedua array *X* dan *Y*. *X* yang merupakan sinyal asli atau sinyal pemisahan *spleeter* dan *Y* merupakan sinyal hasil pemisahan dengan metode *ICA*. Hasil pengujian nilai *MSE* ditunjukkan sebagai berikut.

```
The mean-squared error (5 second) is 0.0356
The mean-squared error (10 second) is 0.0432
The mean-squared error (15 second) is 0.0558
```

Gambar 8. Screen shoot hasil pengujian MMSE

*Perhitungan Signal to Interference Ratio (SIR)*

Pada proses perhitungan SIR digunakan nilai MSE yang didapatkan dari proses sebelumnya. Hasil perhitungan SIR akan menunjukkan semakin tinggi nilai SIR, maka kualitas sinyal semakin baik, begitu juga sebaliknya. Hasil perhitungan dari nilai *signal to interference ratio* ditunjukkan pada tabel 1.

Tabel 1. Hasil pengukuran MSE dan SIR

Durasi	MSE	SIR
5 detik	0.0356	14.4855000203
10 detik	0.0432	13.6451625319
15 detik	0.0558	12.5336580106

V. KESIMPULAN

Dari paparan yang telah disajikan pada paper ini, dengan mengacu pada hasil pengujian pada model yang telah disusun, selanjutnya diperoleh gambaran singkat sebagai berikut:

- Proses iterasi pada *FastICA* dengan pendekatan *negentropy* lebih banyak daripada kurtosis yakni sebanyak 6 kali dengan nilai akhir sebesar 6.388e-08, sedangkan pada pendekatan dengan kurtosis sebanyak 4 kali dan nilai akhir iterasi didapatkan sebesar 6.398e-07.
- Pengujian *Mean Square Error* (MSE) terhadap sinyal *wicara* hasil pemisahan, diperoleh nilai sebesar 0.0334 pada pemisahan dengan pendekatan nilai kurtosis dan *negentropy*, sedangkan untuk pendekatan pada maksimum kurtosis didapatkan MSE sebesar 0.0001.
- Perhitungan *Signal to Interference Ratio* (SIR) terhadap sinyal *wicara* hasil pemisahan, diperoleh nilai sebesar 1.4762535332 untuk pemisahan dengan pendekatan nilai kurtosis dan *negentropy*, sedangkan untuk pendekatan pada maksimum kurtosis didapatkan nilai SIR sebesar 4.
- Hasil pemisahan sinyal pada ketiga pendekatan, yakni *negentropy*, kurtosis, maupun maksimum kurtosis menunjukkan kualitas yang baik, yakni nilai MSE yang dihasilkan telah mendekati 0 dan nilai SIR lebih dari 1.

UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada Manajemen PENS yang telah memberikan dukungan dalam bentuk fasilitas laboratorium dan penyediaan semua perangkat untuk pelaksanaan penelitian ini.

REFERENSI

- [1] C.Y. Fook, M. Hariharan, S. Yaacob, and A. Ah, "A Review : Malay Speech Recognition and Audio Visual Speech Recognition," no. February, pp. 27-28, Universiti Malaysia Perlis (UniMAP), 2012.
- [2] M.V. Gubin, "Using Convolutional Neural Networks to Classify Audio Signal in Noisy Sound Scenes", 2018 Global Smart Industry Conference (GloSIC).
- [3] Snehal S. Gaikwad, Pallavi P. Ingale, Dr. S. L. Nalbalwar, "Separation of singing voice from background musical noise using modified NMF and Filtering", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016 .
- [4] Pritish Chandna, Merlijn Blaauw, Jordi Bonada, Emilia Gomez, , "A Vocoder based Method for Singing Voice Extraction", ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
- [5] Kenneth John Faller, Jason Riddley, and Elijah Grubbs, "Automatic Blind Source Separation of Speech Source in an Auditory Scene ", 2017 51-st Asilomar Conference on Signals, Systems, and Computers.
- [6] Yang Yang, Zuoli Li, Xiuqin Wang and Di Zhang, "Noise Source Separation based on the Blind Source Separation", 2011 Chinese Control and Decision Conference (CCDC), Mianyang, China ,23-25 May 2011.
- [7] R.Farkhan, "Application of blind source separation technique for separating noise from non gaussian acoustic signal", Institut Teknologi Sepuluh Nopember Surabaya, 2013.
- [8] P. Angga Pramana, W. Ni Wayan, M. Ni Putu, B. I Dewa Made, "Independent Component Analysis (ICA) dan Sparse Component Analysis (SCA) dalam Pemisahan Vokal dan Instrumen pada Seni Geguntangan", Jurnal Elektronik Ilmu Komputer Udayana p-ISSN: 2301-5373 Volume 8, No 1, Universitas Udayana, 2019.
- [9] Housseem Maghrebi and Emmanuel Prouff, "On the Use of Independent Component Analysis to Denoise Side-Channel Measurements", International Workshop on Constructive Side-Channel Analysis and Secure Design, COSADE 2018. In the book of Constructive Side-Channel Analysis and Secure Design pp 61–81.
- [10] M. Shoji, S. Hiroshi, M. Ryo, A. Shoko, "Blind Source Separation of Convolutive Mixtures of Speech in Frequency Domain", IEICE Trans. Fundamental, VOL.E88-A, NO.7 July 2005.
- [11] T. W. Lee. "Independent Component Analysis", Kluwer Academic Publishers, Boston, 2000.
- [12] T.Sihotang, A.Asni, dan Anwar Fattah, " Ekstraksi Ciri Menggunakan Algoritma Discrete Wavelet Transform (DWT) dan Principal Component Analysis Pada Warna Kulit wajah", JTE UNIBA, Vol. 3, No. 2, April 2019.
- [13] Saftiadi, A. Asni B, Aswadul Fitri Saiful Rahmant, " Perancangan Sistem Kontroler Alat Elektronik Rumah Tangga berbasis Mikrokontroler Arduino dengan Perintah Suara", JTE UNIBA. Vol. 3. No. 2, April 2019.
- [14] A.Asni B, Diah Patriana Setianingsih, " Pengenalan IdentitasPenutur MenggunakanAlgoritmaDiscreteWavelet Transform (DWT)dan Hidden Markov Modesls (HMM)", JTE UNIBA, Vol. 05, No 1. September 2018.