ALGORITMA PENGHAPUS DERAU/*SILENCE* DAN PENENTUAN ENDPOINT DENGAN NILAI AMBANG TERBOBOT UNTUK SINYAL SUARA

Syahroni Hidayat[1], Uswatun Hasanah[2], Ahmad Ashril Rizal[3]
(1,2,3) Jurusan Teknik Informatika, STMIK Bumigora Mataram
Jl. Ismail Marzuki, Cakranegara, Mataram, Nusa Tenggara Barat
(1)syahroni.hidayat@stmikbumigora.ac.id, (2)uswatun@stmikbumigora.ac.id,
(3)ashril.rizal@stmikbumigora.ac.id

Abstract

In automatic speech recognition, the accuracy of recognition depends on the accuracy of endpoint detection of speech signal. There are several methods commonly used, such as short time energy-zero crossing rate, statistics and hybrid. However, these methods have limitations in determining the threshold value, the range of methods and computational efficiency. Therefore, we need a method that can solve these problems one of them by modifying the threshold value. The threshold value is modified such that its value increase four time from its initial value after multiplied by the weight. The results shows this novel method provides high accuracy on silence remover and end point detection although some data were missing.

Key word : *Automatic Speeh Recognition,silence remover, endpoint detection, weighted value.*

1. Pendahuluan

Performa suatu sistem pengenal suara otomatis bergantung pada kualitas sinyal suara [1]. Sedangkan kualitas sinyal suara sangat bergantung pada akurasi pemrosesan awal (pre-processing) sinyal suara yang meliputi penghapusan derau, deteksi *endpoint*, preemphasis, segmentasi, *windowing* dan lain-lain [2]. Tahap terpenting dari permrosesan awal tersebut adalah penghapusan derau/*silence* dan penentuan *endpoint* agar dapat mendeteksi keberadaan suara dan mengekstraknya dari sinyal suara yang berderau. Selain itu, kedua proses tersebut akan berpengaruh pada efisiensi komputasi dikarenakan data sinyal yang diproses lebih kecil [3], [4].

Deteksi *endpoint* begitupula penghapusan derau/*silence* berdasarkan batas ambang energi atau *Short Time Energy* (STE) dan *Zero crossing Rate* (ZCR) sinyal suara dalam domain waktu merupakan dua buah teknik yang paling sering digunakan [2], [5]. Dimana log batas ambang energi (STE) merupakan profil energi sinyal suara yang berubah-ubah sedangkan ZCR merupakan banyaknya *zero crossing* pada sinyal suara yang diukur pada selang waktu tertentu [5], [6]. Meskipun telah banyak diaplikasikan pada proses sistem pengenalan suara otomatis, kedua metode ini memiliki kekurangan terutama pada penentuan nilai batas ambang.

Sebuah metode statistika untuk mendeteksi *endpoint* telah dikembangkan pada[2]. Fungsi *Mahalanobis Distance* yang merupakan pengembangan dari teori distribusi Normal/Gauss digunakan sebagai pengklasifikasi linier. Dengan mengasumsikan *background noise* yang terdapat pada sinyal suara terdistribusi secara normal. Kelebihan metode ini terletak pada penentuan batas ambang yang dapat ditentukan tanpa melakukan trial dan error seperti pada metode klasik. Akan tetapi, metode ini masih harus membutuhkan modifikasi ketika sinyal suara memiliki tipe derau yang berbeda, sehingga akan menyebabkan munculnya beberapa nilai ambang sekaligus.

Pada[3], telah dikembangkan metode deteksi *endpoint hybrid*. Dimana metode *hybrid* tersebut menggabungkan metode yang berbasis STE-ZCR dengan metode yang berbasis DTW. Oleh karena penggabungan metode ini, *output* sinyal suara akan memiliki banyak titik *endpoint* sehingga dibutuhkan proses umpan balik untuk mendapatkan nilai *endpoint* sebenarnya yang tentunya hal ini sangat berpengaruh pada efisiensi proses komputasi. Padahal salah satu tujuan dari deteksi *endpoint* adalah untuk mempercepat proses komputasi.

Sinyal suara merupakan sinyal yang bervariasi lambat sebagai fungsi waktu, dalam hal ini ketika diamati pada durasi yang sangat pendek (5 sampai 100 mili detik) karakteristiknya masih stasioner. Tetapi bilamana diamati dalam durasi yang lebih panjang (> 1/5 detik) karakteristik sinyalnya berubah untuk merefleksikan suara ucapan yang keluar dari pembicara.

Salah satu cara dalam menyajikan sebuah sinyal suara adalah dengan menampilkannya dalam tiga kondisi dasar, yaitu *silence* (S) atau keadaan tenang dimana sinyal wicara tidak diproduksi, *unvoiced* (U) dimana *chord* vokal tidak bervibrasi, dan yang ketiga adalah *voiced* (V) dimana *chord* vokal bervibrasi secara periodik sehingga menggerakkan udara ke kerongkongan melalui mekanisme akustik sampai keluar mulut dan menghasilkan sinyal suara[7].

Pada sinyal suara, sinyal stasioner yang terjadi antar 5-100 ms pertama dapat dianggap sebagai derau dan dinyatakan dalam persamaan:

$$x(t) = A\sin(2\pi t + \phi) \quad (1)$$

Dimana A merupakan amplitudo/magnitudo sinyal dengan ϕ sebagai fasenya. Energi sinyal suara pada rentang tertentu dinyatakan oleh [8] sebagai berikut:

$$E_i = x_i^2 - x_{i+1}x_{i-1} \quad (2)$$

Sehingga,

$$E = A^2 sin^2(\Omega) \quad (3)$$

Dari persamaan (3), energi amplitudo dinyatakan sebagai batas ambang derau yang kan dikalikan dengan bobot. Dimana besar bobot adalah satu seperempat kali besar nilai ambang.

Pada penelitian ini, dilakukan pendeteksian terhadap sinyal suara berupa *silence* dengan menggunakan metode nilai ambang terbobot. Nilai ambang sebenarnya diperoleh dengan mencari nilai amplitudo maksimum sinyal pada durasi 10-30 ms pertama. Kemudian bobot nilai ambang ditentukan sedemikian rupa sehingga nilai ambang sebenarnya akan bertambah sebesar seperempat kali dirinya. Ini dilakukan agar nilai ambang dapat mencakup seluruh derau sinyal yang memiliki perbedaan sangat kecil. Selanjutnya sinyal suara diekstrak dari sinyal berderau dengan cara menghapus sinyal yang dianggap *silence*. Eksperimen ini telah dilakukan pada sinyal suara yang berupa sebuah kata yang terdiri dari dua suku kata dan sinyal suara yang berupa suku kata dari kata tersebut. Metode ini menunjukkan performa hasil yang lebih baik, dengan algoritma yang lebih sederhana dan komputasi yang ringan.

2. Metodologi

Sinyal suara yang digunakan pada penelitian ini direkam dengan frekuensi sampling sebesar 16 kHz dengan durasi perekaman selama 1 detik[4], [9],[10]. Pemilihan frekuensi sampling tersebut karena frekuensi sampling suara rekaman dengan kualitas rendah sampai kualitas terbaik berkisar antara 8 kHz – 16 kHz. Rekaman berupa sebuah kata dalam bahasa Indonesia dan suku kata dari kata tersebut . Kata yang dipilih hanya terdiri dari dua suku kata dan memiliki fonem yang mirip[4], [9].

Selanjutnya dari sinyal suara yang telah direkam dicari nilai maksimum sebagai ambang batasnya. Pencarian dilakukan hanya pada durasi stasioner yaitu pada 400 sampel pertama. Nilai ambang yang diperoleh kemudian dikalikan dengan nilai bobot.

Proses selanjutnya adalah membagi sinyal suara menjadi beberapa frame. Panjang masing-masing frame adalah 0.025 detik. Kemudian dari frame-frame tersebut diekstrak sinyal yang mengandung sinyal suara tanpa *overlapping*. Berikut adalah algoritma dari proses penghapusan derau/silence:

Langkah 1. Baca file suara, nyatakan sebagai *x* kemudian hitung jumlah sampelnya.

$$x = (x_1, x_2, x_3, \dots, x_n) \quad (7)$$

Langkah 2. Cari nilai maksimum untuk 400 sampel pertama dari absolute sinyal *x* lalu tetapkan sebagai nilai ambang.

$$TH = \max(x(1{:}\,400)) \quad (8)$$

Langkah 3. Kalikan nilai ambang dengan bobot. Besar bobot adalah 1.25 yaitu satu seperempat kali besar nilai ambang.

$$TH_w = TH * 1.25 \quad (9)$$

Langkah 4. Bagi sinyal suara menjadi beberapa frame dimana sebuah frame masing-masing terdiri dari 400 data. Proses dilakukan tanpa *overlapping* sampel frame.

$$frame = 0.025 * fsampling \quad (10)$$

Langkah 5. Identifikasi frame-frame *non silent* dengan mencari frame yang memiliki nilai amplitudo maksimum yang lebih besar dari ambang (TH).

$$ampli_{maks} = \max(frame)$$

$$if\ ampli_{maks} > TH_w \quad (11)$$

Langkah 6. Buat sebuah sinyal baru yang tidak mengandung frame *silence*.

$$sinyal_{baru} = frame\_nonsilent \quad (12)$$

Sinyal baru yang terbentuk masih memiliki panjang yang sama dengan sinyal yang lama. Akan tetapi, sinyal baru tersebut memiliki pola seperti sinyal yang telah diberi *zero padding*. Oleh karena itu, untuk menentukan *endpoint*, data sinyal yang bernilai nol tersebut harus dihilangkan. Adapun langkah-langkahnya adalah sebagai berikut.

Langkah 7. Vektor data ***x'*** merupakan nilai absolut dari data sinyal baru.

$$x' = abs(sinyal_{baru}) \quad (13)$$

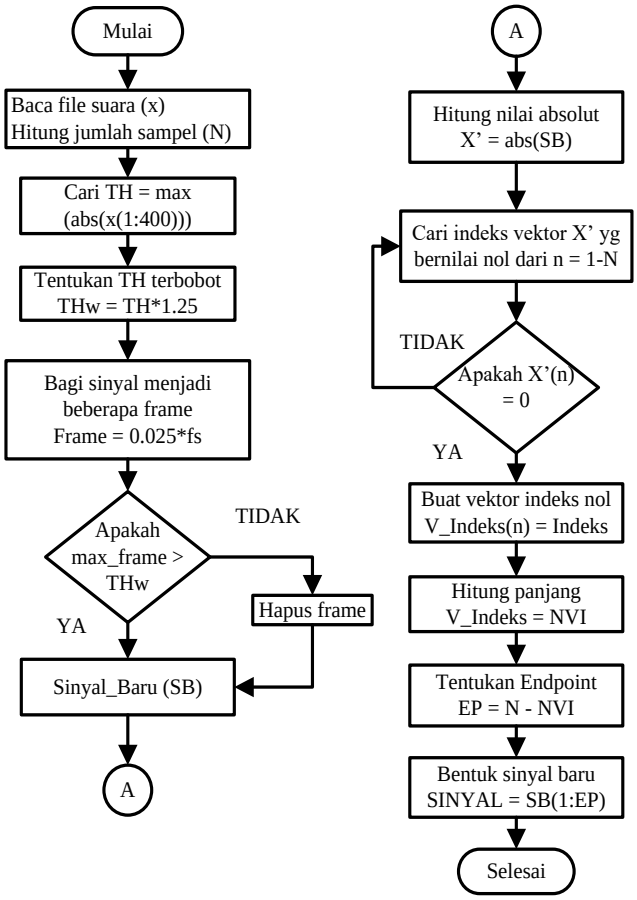
Langkah 8. Cari indeks nilai dari vektor data sinyal ***x'*** yang bernilai nol. Kemudian bentuk sebuah vektor indeks nilai.

$$Indeks(n) = \min\left(\sum_{n=1}^{N} xi\right) \quad (14)$$

Langkah 9. Setelah terbentuk vektor indeks nol, hitunglah panjang vektor tersebut.

Langkah 10. Kurangi panjang vektor sinyal baru dengan panjang vektor indeks nol. Selanjutnya bentuklah sinyal baru dengan panjang vektor sebesar hasil pengurangan kedua vektor tersebut.

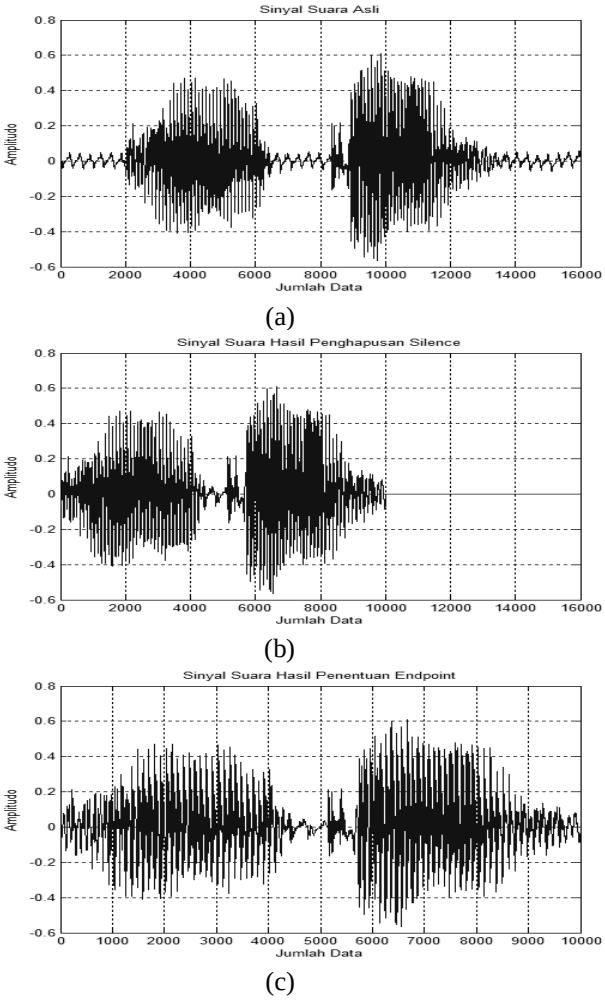
Secara garis besar algoritma penghapus derau/silence dan pendeteksi *endpoint* ditunjukkan oleh *flowchart* pada Gambar 1 berikut ini.



Gambar 1. Diagram alir penghapus derau/silence dan deteksi *endpoint.*

3. Pembahasan

Dari percobaan yang telah dilakukan diperoleh hasil yang memuaskan. Pada proses ekstraksi suara dari derau, algoritma dapat bekerja secara umum baik pada sinyal suara berupa kata maupun suku kata. Dimana ketepatannya baik dalam penghapusan derau sinyal untuk yang berada di awal, di akhir, maupun di antara sinyal suara gambar 2(b). Adapun pada gambar 2(c) terlihat akurasi algoritma sudah cukup baik dalam menentukan *endpoint* sinyal suara.



Gambar 2. Hasil metode ambang terbobot, (a). Sinyal suara asli,
(b). Penghapus derau/silence (c). Deteksi *endpoint*.

Terjadi kehilangan beberapa data sinyal setelah menentukan panjang sinyal suara baru. Hal ini diakibatkan adanya nilai 0 yang berupa *zero crossing* sinyal yang terletak diantara sinyal suara. Jumlah *zero crossing* yang muncul di antara sinyal suara tidaklah terlalu signifikan sehingga kualitas sinyal suara masih dapat dikategorikan baik.

4. Kesimpulan

Pada paper ini telah dibuat sebuah algoritma untuk menghapus derau dan menentukan *endpoint* sinyal suara secara tepat dan akurat. Metode yang digunakan adalah dengan memberikan bobot pada nilai ambang yang diperoleh dari nilai maksimum sinyal stasioner pada 0-0.25 detik pertama sinyal suara. Algoritma umumnya

dapat bekerja baik pada sinyal suara yang terdiri dari kata dan kosa kata. Namun dalam menentukan *endpoint* dibutuhkan sedikit modifikasi agar dapat mangabaikan sinyal berupa *zero crossing* dan dianggap *zero padding*.

5. Saran

Algoritma ini dapat dikembangkan agar dapat menentukan jumlah suku kata pada suatu sinyal suara serta dapat menentukan *endpoint* dari masing-masing suku kata tersebut.

Daftar Pustaka

[1] J. Ramirez, J. Górriz, and J. Segura, "Voice activity detection. fundamentals and speech recognition system robustness," *Robust Speech Recognit. …*, no. June, pp. 1–22, 2007.

[2] S. S. G. Saha, Sandipan Chakroborty, "A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications," in *IEEE 11th National Conference on Communications (NCC)*, 2005, pp. 291–195.

[3] L. F. Lamel, L. R. Rabiner, A. E. Rosenberg, and J. G. Wilpon, "Improved endpoint detector for isolated word recognition," *IEEE Trans. Acoust.*, vol. ASSP-29, no. 4, pp. 777–785, 1981.

[4] R. Asliyan, "Syllable Based Speech Recognition," in *Speech Technology*, I. Ipsic, Ed. InTech, 2011, pp. 263–284.

[5] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice Hall, 2001.

[6] L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances," *Bell Syst. Tech. J.*, vol. 54, no. 2, pp. 297–315, 1975.

[7] D. Jurafsky and J. H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, 2008.

[8] L. Gu and S. A. Zahorian, "A new robust algorithm for isolated word endpoint detection," *Energy*, vol. 1, no. 1. pp. 1–4, 2002.

[9] Abriyono and A. Harjoko, "Pengenalan Ucapan Suku Kata Bahasa Lisan Menggunakan Ciri LPC, MFCC, dan JST," *IJCCS*, vol. 6, no. 2, pp. 23–34, 2012.

[10] I. McLoughlin, *Applied Speech And Audio Processing : With Matlab Examples*, 1st ed. United Kingdom: Cambridge University Press, 2009.