

SEGMENTASI SUARA BAHASA INDONESIA MENJADI UNIT-UNIT FONEM SEGMENTATION OF INDONESIAN SPEECH INTO PHONEMIC UNITS

Guruh Susanto¹, Suyanto², -³

¹Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

Abstrak

Segmentasi suara memiliki peranan yang penting dalam proses pengenalan suara. Dengan menghasilkan segmen yang tepat dari suara yang akan dianalisa akan mengurangi kebutuhan terhadap kapasitas memori yang besar dan akan meminimalkan kompleksitas perhitungan pada vokabulari yang besar. Secara umum terdapat dua jenis segmentasi terhadap suara. Yang pertama adalah segmentasi berdasarkan unit-unit fonem, dimana suara akan disegmentasi menjadi abjad atau huruf. Jenis segmentasi yang lain adalah segmentasi berdasarkan unit-unit silabel.

Pada umumnya jenis segmentasi suara yang ada memerlukan pengetahuan atau informasi linguistik dari suara yang akan diprosesnya. Sebuah pendekatan dalam proses segmentasi suara yang disebut sebagai metode blind diperkenalkan. Pendekatan jenis ini memungkinkan sampel suara untuk dapat disegmentasi menjadi unit-unit fonem tanpa adanya pengetahuan apapun mengenai informasi linguistik. Oleh sebab itu, pendekatan ini melibatkan pencarian jumlah optimal segmen dalam unit-unit fonem pada sampel suara yang diberikan, sebelum menentukan batas-batas segmen.

Kata Kunci : segmentasi suara, unit fonem, unit silabel, jumlah optimal segmen

Abstract

Speech segmentation plays an important role in speech recognition in reducing the requirement for large memory and in minimizing the computation complexity in large vocabulary continuous speech recognition systems. In general, there are two kinds of segmentation. One is phonemic unit segmentation, which segments speech into fonemes based on the features of the phonemes. The other is syllabic unit segmentation.

In general, most of the approaches are based on the linguistic information used to segment the speech data. A new automatic speech segmentation procedure, called the Blind Method speech segmentation, is presented. This procedure allows a speech sample to be segmented into phonemic units without the knowledge of any linguistic information. Hence, this procedure involves finding the optimal number of fonemic unit segments in the given speech sample, before locating the segment boundaries.

Keywords : speech segmentation, phonemic unit, syllabic unit, optimal number segmen

1. Pendahuluan

1.1 Latar belakang

Segmentasi dan pelabelan pada data suara menurut fonetik atau aturan linguistik yang sama merupakan tugas fundamental dalam pemrosesan suara. Mayoritas sistem pemrosesan suara yang tersedia sekarang, mencakup sistem pengenalan suara dengan vokabulari medium sampai besar, sistem pengenalan pembicara, dan sistem pengidentifikasi bahasa, didesain berdasarkan pada unit-unit akustik fonem. Secara tradisional, segmentasi dan pelabelan pada data suara dilakukan secara manual dengan pelatihan fonetik yang menggunakan pendengaran dan isyarat visual. Akan tetapi melalui pendekatan ini terdapat beberapa kelemahan. Yang pertama, prosedur manual seperti ini dapat sangat melelahkan dan memerlukan banyak waktu. Yang kedua, hanya sebagian kecil orang memiliki kemampuan dan pengetahuan tentang hal ini sehingga prosedur manual lebih bersifat subjektif dan cenderung dipengaruhi oleh kesalahan manusia (*human error*). Karena itu, prosedur segmentasi suara otomatis menjadi lebih dipilih dan digunakan secara ekstensif pada sistem pemrosesan suara. Kebanyakan prosedur segmentasi otomatis dan pelabelan menggunakan keterkaitan pengetahuan linguistik, seperti teks yang diucapkan dan *string* fonem.

Permasalahan mendasar pada segmentasi suara otomatis adalah kebanyakan algoritma segmentasi yang digunakan membutuhkan pengetahuan linguistik pada data suara yang diberikan untuk dapat memprosesnya lebih lanjut, sehingga apabila pada data suara yang diberikan tidak terdapat pengetahuan linguistik maka kebanyakan algoritma segmentasi tidak dapat menanganinya. Oleh karena itu, algoritma segmentasi bukan hanya diperlukan untuk menentukan lokasi batas-batas segmennya saja, tetapi juga sekaligus menentukan jumlah optimal dari unit fonem yang dihasilkan pada sampel suara yang diamati. Tipe segmentasi suara yang seperti ini disebut sebagai prosedur segmentasi suara *blind*.

Pada prosedur segmentasi *blind* dimungkinkan sebuah sampel suara untuk disegmentasi menjadi unit-unit fonem tanpa adanya pengetahuan apapun mengenai informasi linguistik, seperti: pencatatan fonetik. Pada prosedur ini terdiri dari algoritma *Convex Hull*, *Spectral Variation Function*, dan *Normal Decomposition* untuk memperkirakan jumlah optimal fonem dari sampel suara yang diucapkan.

1.2 Perumusan masalah

Rumusan masalah dalam penelitian ini adalah :

1. Bagaimana menentukan range jumlah segmen dari sampel suara yang diinputkan.
2. Bagaimana menentukan jumlah segmen sebenarnya dari sampel suara yang diinputkan.

Batasan masalah dalam penelitian ini adalah :

1. Penelitian hanya sebatas pada proses segmentasi suara, tidak sampai pada proses mengenalannya.
2. Sistem yang dihasilkan digunakan sebagai alat simulasi.
3. Data suara yang akan diteliti merupakan suara angka yang dihasilkan dari pembicara laki-laki dan wanita, dimana jenisnya berupa *isolated digit*.

1.3 Tujuan

Tujuan dari penelitian ini adalah dihasilkannya sebuah sistem sebagai alat simulasi proses segmentasi suara Bahasa Indonesia menjadi unit-unit fonem, dimana parameter yang akan dianalisa adalah performansi sistem ditinjau dari pergeseran waktu segmen yang dihasilkan dan keakuratan jumlah segmen yang dihasilkan dari tiap sampel suara yang diproses oleh sistem.

1.4 Metodologi penyelesaian masalah

1. Studi Literatur : Dilakukan studi literature tentang dasar-dasar pengolahan sinyal suara dan segmentasi suara.
2. Pendefinisian Masalah : Melakukan pengamatan tentang bagaimana segmentasi suara secara *blind*.
3. Analisa Kebutuhan Sistem dan Perancangan Perangkat Lunak : Melakukan analisa terhadap model implementasi yang akan dibangun dengan tujuan memahami secara jelas proses yang dilakukan pada sistem tersebut, serta perancangan proses dan aliran data yang terjadi dalam sistem.
4. Implementasi dan Pengujian Perancangan Perangkat Lunak : Implementasi terhadap kebutuhan dan perancangan perangkat lunak sehingga sistem dapat dibangun dan kemudian akan dilakukan pengujian terhadap sistem yang telah dibangun. Dari data output yang dihasilkan akan dilakukan analisa terhadap performansi sistem.
5. Penyusunan Laporan : Laporan yang akan dihasilkan berupa buku Tugas Akhir. Penyusunan laporan dengan mengikuti kaidah penulisan Tugas Akhir yang berlaku di Departemen Teknik Informatika STT Telkom.

Telkom
University

5. Kesimpulan Dan Saran

5.1. Kesimpulan

Dari hasil penelitian yang telah dilakukan dapat diambil beberapa kesimpulan sebagai berikut:

1. Performansi sistem yang diperoleh dari analisis hasil pengujian 40 sampel suara adalah tingkat penambahan (*insertion*) segmen sebesar 2.90%, tingkat pengurangan (*deletion*) segmen sebesar 17.65%, dan tingkat kebenaran (*correctnes*) sebesar 79.30%.
2. Pergeseran waktu (*shift time*) segmen yang dihasilkan oleh sistem dengan data pengujian 40 sampel suara adalah rata-rata sebesar 20.47 ms dengan standar deviasi sebesar 21.43 ms.
3. Pasangan nilai threshold dan orde LPC yang dipergunakan oleh sistem, yaitu 2 dan 20, dapat melakukan proses segmentasi secara baik untuk ucapan suara *isolated digits*.
4. Proses segmentasi suara secara otomatis tanpa adanya pengetahuan atau informasi linguistik sebelumnya, masih tetap dapat dilakukan.

5.2. Saran

Untuk dapat meningkatkan performansi sistem segmentasi suara secara otomatis masih perlu dilakukan banyak perbaikan. Hal yang dapat kami anjurkan untuk dapat meningkatkan akurasi sistem adalah sebagai berikut:

1. Parameter threshold yaitu perbedaan intensitas energi, untuk proses penentuan jumlah segmen minimum perlu ditambahkan. Parameter lain yang bisa dikombinasikan dengan threshold yang ada, misal: *zero crossing rate* dan panjang sub-interval di kiri dan kanan batas segmen.
2. Proses segmentasi manual dapat diganti dengan menggunakan tools segmentasi otomatis sehingga hasil yang diperoleh akan menjadi lebih akurat, karena hasil dari segmentasi tersebut akan dijadikan sebagai faktor pembandingan hasil dari sistem yang dibangun.

Telkom
University

Daftar Pustaka

- [1] Juang, B-H and Rabiner, L.R. *Fundamentals of Speech Recognition*. Prentice-Hall, NJ, 1992.
- [2] Li, Bavy N.L. and Liu, James N.K. *A Comparative Study of Speech Segmentation and Preprocessing for Automatic Multi-lingual Recognition*. Department of Computing Hong Kong Polytechnic University.
- [3] Mermelstein, Paul. *Automatic segmentation of speech into syllabic units*. Journal of Acoustical Society of America, 58(4):880-883, October 1975.
- [4] Myers, C.S. and Rabiner, L.R. *A Level Building Dynamic Time Warping algorithm for Connected Word Recognition*. IEEE Transactions on Acoustics, Speech and Signal Processing, 29:284-297, April 1981.
- [5] Sharma, M. and Mammone, R. *Subword-based text dependent speaker verification system with user-selectable passwords*. In Proceedings of ICASSP, pages 93-96, 1996.
- [6] Sharma, M. and Mammone, R. *“Blind” Speech Segmentation: Automatic Segmentation Of Speech Without Linguistic Knowledge*.
- [7] Villing, R., Timoney, J., Ward, T., and Costello, J. *Automatic Blind Syllable Segmentation for Continuous Speech*. ISSC June 30 – July 2.
- [8] http://id.wikisource.org/wiki/Pedoman_Umum_Ejaan_Bahasa_Indonesia_yang_Disempurnakan