

Frekuensi Dominan Dalam Vokal Bahasa Indonesia

Tjong Wan Sen^{#1}

[#]Fakultas Komputer, Universitas Presiden
Jln. Ki Hajar Dewantara, Jababeka, Cikarang

¹wansen@president.ac.id

Abstract— Pengenalan ucapan otomatis sudah dikembangkan selama lebih dari empat puluh tahun dan masih perlu untuk terus dikembangkan agar dapat memiliki kemampuan yang menyerupai manusia. Teknologi ini sangat bermanfaat tetapi hingga saat ini masih belum bisa dipakai secara umum dengan mudah dalam aktivitas manusia sehari-hari. Ketersediaan pilihan bahasa pun masih menjadi kendala, diperlukan waktu pelatihan yang cukup panjang untuk masing-masing bahasa yang diinginkan.

Salah satu bahasa yang banyak digunakan di dunia adalah Bahasa Indonesia. Meskipun demikian belum banyak sistem Pengenalan Ucapan Otomatis yang menggunakan Bahasa Indonesia. Beberapa sistem yang dikembangkan oleh perusahaan besar telah memiliki pilihan Bahasa Indonesia, tetapi basis data yang digunakan tidak dapat diakses oleh masyarakat umum. Oleh karena itu perlu dikembangkan basis data suara ucapan dalam Bahasa Indonesia.

Dalam artikel ini dilaporkan pengembangan data suara ucapan vokal Bahasa Indonesia. Pengumpulan data dilakukan dengan cara merekam suara vokal dari berbagai sumber. Sumber ucapan dibedakan secara etnisitas, jenis kelamin dan usia. Dari masing-masing kelompok diidentifikasi kelompok frekuensi yang dominan. Himpunan frekuensi dominan antar kelompok tersebut kemudian dibandingkan satu dengan lainnya untuk mengetahui persamaan dan perbedaannya.

Keywords— data suara ucapan, frekuensi *formant*, pengenalan ucapan otomatis, suara ucapan vokal Bahasa Indonesia

I. PENDAHULUAN

Salah satu cara berkomunikasi yang murah, cepat dan sangat natural bagi manusia adalah Komunikasi Suara. Manusia melalui alat ucap dan pendengaran serta bantuan otak melatih secara terus menerus kemampuan tersebut. Dengan cara menggunakannya setiap hari hampir semua manusia dapat melakukannya dengan baik. Dengan bantuan teknologi telekomunikasi, jangkauan komunikasi suara menjadi sangat jauh dibandingkan dengan jangkauan awalnya yang hanya beberapa meter saja. Di sisi lain ada perkembangan teknologi komputer yang kemampuannya meningkat secara sangat cepat. Hampir semua bidang aktivitas manusia sehari-hari telah menggunakan komputer. Tentunya akan sangat baik jika komputer bisa berinteraksi dengan manusia melalui komunikasi suara.

Oleh karena itu manusia sejak lebih dari empat puluh tahun yang lalu berusaha untuk mewujudkan hal tersebut yaitu sistem Pengenalan Ucapan Otomatis (*Automatic Speech Recognition* disingkat ASR). Sistem ini yang dapat secara otomatis mengenali suara yang diucapkan oleh manusia. Jika digabungkan dengan teknologi Bahasa Natural maka komunikasi suara antara manusia dengan komputer dapat dilakukan. Tetapi seperti juga pada komunikasi suara antar manusia, masalah perbedaan bahasa juga terjadi pada komunikasi suara antara manusia dengan komputer. Dalam hal ini manusia dan komputer perlu untuk menggunakan bahasa yang sama. Jika tidak maka akan diperlukan penterjemah. Dari sudut pandang ASR, hal tersebut disebabkan oleh karena basis data suara ucapan. Basis data suara ucapan ASR dihasilkan melalui proses pelatihan menggunakan bahasa tertentu. ASR yang dilatih menggunakan Bahasa Inggris misalnya sudah pasti tidak akan dapat mengenali suara ucapan dalam Bahasa Indonesia dengan baik.

Basis data suara ucapan dalam Bahasa Indonesia adalah salah satu basis data suara ucapan yang belum dikembangkan secara lengkap dan terpublikasi secara luas. Padahal Bahasa Indonesia merupakan salah satu bahasa yang paling banyak dipakai di dunia pada saat ini. Ada beberapa sistem ASR yang dikembangkan oleh perusahaan besar yang dapat mengenali ucapan dalam Bahasa Indonesia, tetapi basis data suara ucapan yang digunakan tidak bisa dipakai tanpa sistem mereka. Oleh karena itu menjadi penting untuk mengembangkan basis data suara ucapan dalam Bahasa Indonesia.

Basis data suara ucapan adalah kumpulan ciri khusus dari semua suara ucapan yang ada dalam suatu bahasa. Ciri-ciri tersebut diperoleh dari banyak rekaman suara ucapan yang telah melalui proses ekstraksi ciri. Rekaman suara dihasilkan dari banyak orang yang berbeda-beda yang mengucapkan banyak ucapan yang berbeda-beda pula. Ciri-ciri tersebut merupakan kekhasan yang dipengaruhi oleh perbedaan jenis kelamin, usia, gaya bicara dan aksen.

I. TEORI

Pada bagian ini dipaparkan teori-teori yang mendasari langkah-langkah yang digunakan dalam penelitian yang dilakukan.

A. Pengenalan Ucapan Otomatis

Input bagi pengenalan ucapan otomatis adalah suara melalui mikrofon sedangkan outputnya pada umumnya adalah berupa tampilan teks di layar monitor sehingga terkadang disebut sebagai teknologi 'speech to text'. Ada tiga komponen utama dalam pengenalan ucapan otomatis yaitu pengekstrak ciri, pengenal dan basis data ucapan [1]. Basis data ucapan terdiri atas tiga sub komponen yaitu model akustik, model kamus dan model bahasa. Tugas dari pengekstrak ciri adalah untuk mengubah suara yang diterima menjadi sekelompok ciri yang secara unik mewakili masukan yang bersangkutan. Sedangkan tugas dari pengenal adalah untuk melakukan pencarian kemiripan yang paling tinggi atau relevan dari kelompok ciri yang dihasilkan oleh pengekstrak ciri dengan kelompok ciri yang tersimpan didalam basis data ucapan. Tugas basis data suara ucapan adalah untuk menyimpan semua kelompok ciri unik dari semua suara yang berbeda yang ada dalam suatu bahasa. Data yang disimpan oleh basis data model akustik adalah kelompok ciri unik dari satuan terkecil bunyi suara yang membedakan atau fonem. Sedangkan data yang disimpan oleh model kamus adalah informasi mengenai semua kata yang ada dalam suatu bahasa. Dan data yang disimpan oleh model bahasa adalah informasi mengenai aturan tata bahasa yang berlaku dalam suatu bahasa.

B. Data Suara Ucapan

Basis data suara ucapan pengenalan ucapan otomatis yang terdiri atas tiga sub komponen yaitu model akustik, model kamus dan model bahasa, dihasilkan pada tahap pelatihan [2]. Dalam tahap ini, model akustik pengenalan ucapan dilatih dengan semua kelompok ciri unik yang dihasilkan oleh pengekstrak ciri. Masing-masing kelompok ciri diperoleh dari potongan suara hasil perekaman dan penyuntingan. Satu data rekaman menghasilkan satu kelompok ciri yang unik. Kelompok-kelompok ciri dari fonem yang sama dikumpulkan dan membentuk kelompok yang lebih besar atau disebut kelas. Sehingga pada umumnya jumlah kelas tersebut adalah mengikuti jumlah fonem yang ada di dalam suatu bahasa.

Di dalam kelas yang sama (mewakili sebuah fonem yang tertentu) kelompok ciri yang satu tidak sama persis dengan kelompok ciri yang lainnya. Hal tersebut dikarenakan hampir tidak ada suara manusia sebagai sumber ucapan yang direkam yang memiliki suara seratus persen sama. Bahkan dari satu sumber ucapan yang sama pun umumnya tidak dapat menghasilkan suara yang seratus persen sama (meskipun mengucapkan kata yang sama. Oleh karena itu dalam basis data model akustik pada umumnya memiliki toleransi. Toleransi diperoleh setelah memproses semua suara rekaman secara keseluruhan dengan statistik.

C. Vokal Bahasa Indonesia

Jumlah fonem di dalam suatu bahasa adalah tertentu. Fonem adalah unsur ucapan terkecil yang berfungsi sebagai pembeda arti. Oleh karena itu bunyi suara satu fonem umumnya berbeda secara signifikan satu dengan lainnya. Kata dibentuk dari satu atau lebih fonem. Sebuah kata memiliki kombinasi fonem yang berbeda dengan kata yang lainnya. Sehingga suatu kata akan memiliki arti yang berbeda dengan kata lainnya. Fonem dalam suatu bahasa pada umumnya terbagi ke dalam dua kelompok yaitu vokal dan konsonan namun beberapa bahasa dapat memiliki kelompok khusus yang lain. Salah satu ciri yang dimiliki oleh fonem dalam kelompok vokal adalah memiliki suara yang lebih keras atau jelas dan memiliki durasi yang lebih lama jika dibandingkan dengan fonem-fonem yang berada di dalam kelompok konsonan. Hal ini disebabkan oleh karena perbedaan gabungan frekuensi dan besarnya energi yang terkandung di dalam kelompok fonem tersebut. Karakteristik tersebut membuat fonem-fonem dalam kelompok vokal menjadi lebih mudah untuk dikenali.

Jumlah fonem dalam suatu bahasa biasanya diatur oleh tata bahasanya. Misalnya jumlah fonem dalam Bahasa Indonesia diatur oleh Tata Bahasa Indonesia [3]. Bahasa Indonesia memiliki lima buah fonem vokal utama yaitu A, E, I, O, dan U. Kombinasi satu atau lebih fonem vokal tersebut merupakan unsur pembentuk utama dari seluruh kata yang terdapat dalam kamus Bahasa Indonesia. Fonem-fonem dari kelompok konsonan pada umumnya membuat perbedaan yang melengkap. Posisi fonem-fonem konsonan umumnya ditempatkan di awal, akhir dan atau antara fonem-fonem vokal tersebut.

D. Frekuensi Suara Ucapan

Perbedaan bunyi suara pada umumnya ditentukan oleh perbedaan frekuensi yang dikandungnya. Kombinasi dari frekuensi-frekuensi yang berbeda menghasilkan bunyi suara yang berbeda pula. Oleh karena itu untuk mengenali bunyi suara ucapan dapat didasarkan kepada komponen-komponen frekuensi penyusunnya. Untuk keperluan tersebut diperlukan transformasi yang dapat membawa informasi dari domain waktu ke domain frekuensi. Transformasi Fourier dapat mengubah informasi suatu sinyal dalam domain waktu ke domain frekuensi [4]. Pada domain frekuensi sinyal rekaman suara $F(u)$, yang merupakan hasil dari Transformasi Fourier sinyal $f(x)$ dari domain waktu, direpresentasikan sebagai penjumlahan tidak terhingga dari gelombang sinus dan cosinus dengan semua frekuensi seperti ditunjukkan oleh (1) dan (2).

$$F(u) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi ux} dx \quad (1)$$

Atau

$$F(u) = \int_{-\infty}^{\infty} f(x) (\cos 2\pi ux - i \sin 2\pi ux) dx \quad (2)$$

Fungsi kontinu $F(u)$ adalah fungsi dalam domain frekuensi dan merupakan bilangan kompleks. Fungsi kontinu $f(x)$ adalah fungsi dalam domain waktu dengan variabel riil x yang dalam hal ini merupakan representasi dari sinyal rekaman suara. Variabel u adalah variabel frekuensi dalam domain frekuensi.

Dalam prakteknya dengan menggunakan kartu suara sebuah komputer sinyal yang digunakan adalah sinyal diskrit. Transformasi Fourier Diskrit dapat dimanfaatkan untuk menggantikan Transformasi Fourier dalam menangani sinyal diskrit. Dengan mengambil sampel dari fungsi kontinu $f(x)$ secara periodik maka hasil Transformasi Fourier untuk sinyal diskrit yang bersesuaian dapat diperoleh melalui (3).

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x)e^{-i2\pi ux/N} \quad (3)$$

Variabel N adalah banyaknya sampel diskrit yang diambil dari fungsi $f(x)$ dengan cara mengambil sampel secara periodik. Pada penelitian ini digunakan salah satu bentuk khusus dari Transformasi Fourier Diskrit yaitu Cooley-Tukey Fast Fourier Transform (FFT) yang dapat mereduksi kebutuhan komputasi dari N^2 menjadi $N \log_2 N$ [5] [6].

E. Frekuensi Dominan

Di dalam domain frekuensi sinyal yang direpresentasikan oleh jumlah tak terhingga dari semua komponen frekuensi dapat di kelompokkan ke dalam dua bagian yaitu kelompok yang dominan dan yang tidak. Besarnya amplitudo masing-masing komponen frekuensi dapat digunakan untuk maksud tersebut. Komponen frekuensi yang memiliki nilai amplitudo besar dapat dimasukkan ke dalam kelompok dominan dan sebaliknya. Dalam penelitian ini akan diidentifikasi jumlah komponen frekuensi yang dominan berdasarkan besarnya nilai amplitudo dari semua suara rekaman secara keseluruhan.

II. METODA PENELITIAN

Penelitian ini dimulai dengan melakukan perekaman suara vokal dari berbagai sumber pengucap. Dari satu sumber ucapan akan direkam beberapa vokal acak yang terpisah secara otomatis dengan bantuan perangkat lunak. Masing-masing suara rekaman, yang berdurasi kurang lebih dua detik, disimpan dalam satu file. Format rekaman adalah mono dengan frekuensi sampling 44.100 Hz. Bit kuantisasi yang digunakan adalah 16 bit. Kondisi ruangan saat perekaman adalah tidak steril dari derau meskipun terbatas pada derau yang umum untuk ruangan kerja biasa di sebuah kantor.

Tahap perekaman dilanjutkan dengan tahap verifikasi dan penyuntingan. Proses verifikasi hasil rekaman dilakukan untuk memastikan suara rekaman adalah sama dengan nama file. Hal ini ditujukan untuk menghindari kesalahan data untuk fonem tertentu dikelompokkan ke dalam kelompok fonem yang lain. Proses verifikasi dilakukan secara *manual* dengan bantuan manusia. Pada tahap penyuntingan dilakukan

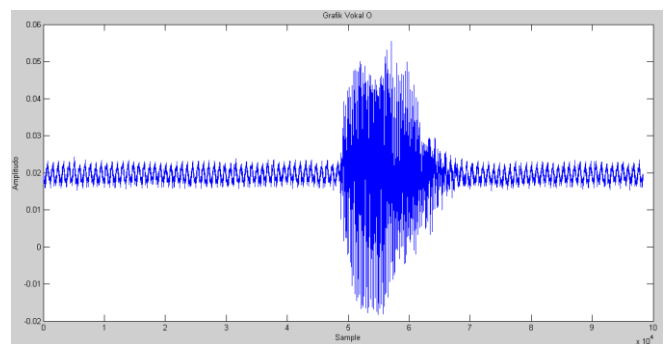
pemotongan hasil rekaman dari kurang lebih dua detik menjadi sekitar 0,128 detik atau setara dengan 5.632 sampel yang memiliki total energi yang paling besar. Jumlah sampel tersebut (5.632) diambil berdasarkan 4.096 sampel disesuaikan dengan satu masukan untuk FFT ditambah dengan tujuh kali 128 sampel di sebelah kanan dan 7 kali sampel di sebelah kiri kumpulan sampel (4.096) yang memiliki total energi yang paling besar.

Setelah tahap penyuntingan penelitian dilanjutkan ke tahap transformasi ke domain frekuensi dengan bantuan FFT. Jumlah sampel bagi satu proses FFT diambil sebesar 4.096 sampel yang bersesuaian dengan kurang lebih 93 milidetik. Hasil transformasi yang diperoleh adalah 2.048 kelompok frekuensi (yang direpresentasikan oleh bilangan kompleks) yang bersesuaian dengan respon frekuensi 0 sampai dengan 22.050 Hz (setengah dari frekuensi sampling 44.100 Hz). Dari kumpulan respon frekuensi inilah frekuensi yang dominan bagi masing-masing fonem akan ditentukan.

Penentuan jumlah frekuensi yang akan dipakai sebagai basis data suara ucapan vokal Bahasa Indonesia dalam penelitian ini dilakukan dengan cara a) mengurutkan nilai amplitudo komponen frekuensi mulai dari yang paling besar sampai yang paling kecil dan b) menghitung nilai penjumlahan secara berulang mulai dari urutan teratas, kedua, ketiga dan seterusnya sampai dicapai tingkat persentase tertentu (dibandingkan terhadap nilai penjumlahan seluruh komponen frekuensi).

III. HASIL DAN PEMBAHASAN

Hasil dari tahap pertama yaitu perekaman adalah berupa file dengan format wav dan masing-masing file berisikan satu vokal dari seorang pengucap. Gbr 1 menunjukkan tampilan isi file tersebut dalam bentuk grafik dengan bantuan perangkat lunak MATLAB.



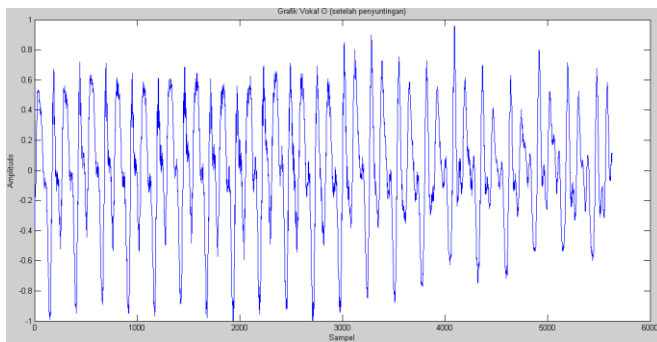
Gbr. 1 Bentuk gelombang salah satu vokal O yang direkam.

Seperti terlihat dalam Gbr. 1, dari sebuah file rekaman berdurasi 2 detik dengan frekuensi sampling 44.100 Hz terdapat sekitar 98.000 sampel dengan besar amplitudonya masing-masing. Bagian awal dan akhir dari rekaman yang tidak mengandung informasi mengenai suara vokal yang diucapkan merupakan bagian yang akan dihilangkan pada

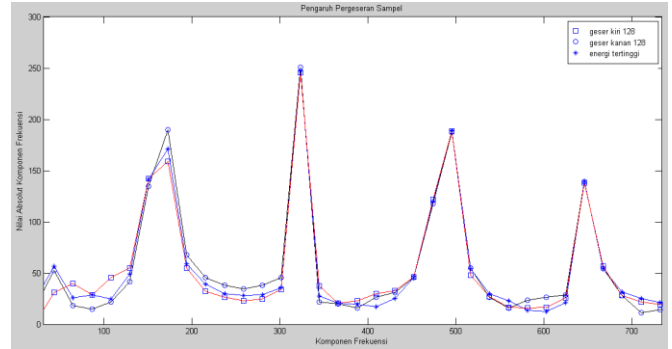
tahap penyuntingan. Dua sumber perbedaan utama dari rekaman suara yang diperoleh adalah keras-pelannya suara yang dihasilkan pengucap dan karakteristik mikrofon dan kartu suara yang digunakan (seperti terlihat dalam Gbr. 1, titik tengah amplitudo yang tidak berada di titik nol merupakan pengaruh dari kartu suara atau konfigurasi komputer yang digunakan untuk merekam). Pengolahan hasil rekaman lebih lanjut dengan mengurangi rata-rata sinyal dan normalisasi amplitudo ke dalam range -1 dan 1 adalah ditujukan untuk maksud tersebut.

Hasil dari tahap selanjutnya yaitu tahap penyuntingan adalah potongan rekaman suara ucapan yang memiliki kandungan energi paling besar. Salah satu contoh potongan hasil penyuntingan dari sebuah suara vokal hasil rekaman (yaitu vokal O yang ditunjukkan oleh Gbr. 1) ditunjukkan oleh Gbr. 2. Seperti ditunjukkan oleh Gbr. 2, terdapat 5.632 sampel beserta amplitudonya masing-masing.

Hasil dari tahap selanjutnya yaitu tahap transformasi informasi dari domain waktu ke domain frekuensi dengan bantuan Transformasi Fourier Diskrit (dalam hal ini menggunakan FFT) adalah respon frekuensi dari sampel-sampel yang berada di dalam potongan rekaman suara yang bersangkutan. Gbr. 3 menunjukkan tiga respon frekuensi dari tiga potongan rekaman suara dari vokal O yang ditunjukkan oleh Gbr. 2. Himpunan sampel potongan suara yang memiliki energi paling tinggi (bintang), himpunan sampel potongan suara yang digeser 128 sampel ke kiri (lingkaran) dan himpunan sampel potongan suara yang digeser ke kanan 128 sampel (kotak).



Gbr. 2 Bentuk gelombang vokal O dalam Gbr. 1 setelah melalui penyuntingan.

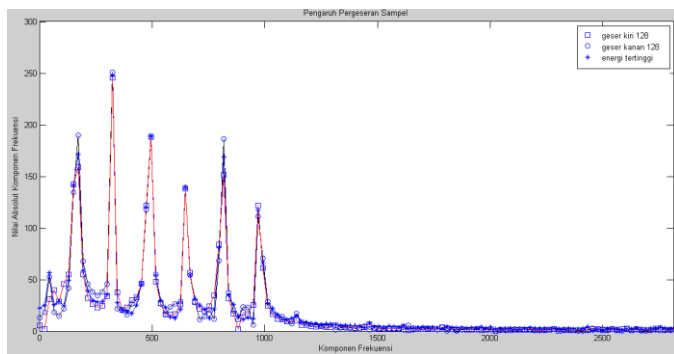


Gbr. 3 Respon frekuensi dari potongan vokal O dalam Gbr. 2.

Seperti ditunjukkan oleh Gbr. 3, pergeseran himpunan sampel potongan suara mengubah respon frekuensi. Nilai amplitudo dari komponen frekuensi (sumbu y) yang digeser berbeda dengan yang tidak digeser untuk komponen frekuensi yang sama (sumbu x). Hal ini menunjukkan bahwa dari potongan rekaman suara vokal yang sama yang diucapkan oleh pengucap yang sama dapat memiliki respon frekuensi yang berbeda jika diambil dari himpunan sampel yang berbeda (dalam hal ini berupa pergeseran sebanyak 128 sampel ke kiri dan kanan).

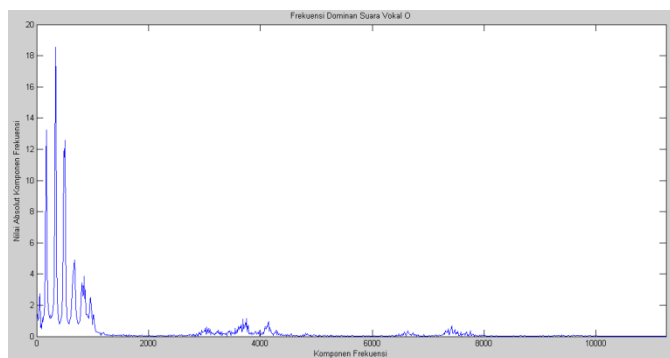
Seperti ditunjukkan oleh Gbr. 3, terdapat nilai amplitudo dari empat komponen frekuensi yang memiliki besar berbeda secara signifikan jika dibandingkan dengan nilai amplitudo komponen frekuensi lainnya. Nilai-nilai tersebut merupakan kandidat untuk dimasukkan ke dalam kelompok frekuensi yang dominan di tahap selanjutnya. Gbr. 3 adalah gambar yang diperbesar sehingga hanya menunjukkan sebagian dari frekuensi dominan tersebut. Seluruh frekuensi dominan yang dimiliki oleh potongan suara tersebut secara lebih lengkap ditunjukkan oleh Gbr. 4.

Setelah semua potongan rekaman ditransformasikan ke domain frekuensi dengan bantuan Transformasi Fourier Diskrit, tahap selanjutnya adalah mengidentifikasi frekuensi-frekuensi dominan dari setiap respon frekuensi yang diperoleh. Salah satu contoh identifikasi dari respon frekuensi yang diperoleh ditunjukkan oleh Gbr. 4. Terdapat sembilan frekuensi yang dominan yaitu yang terletak di kelompok frekuensi di sekitar 40 Hz, 175 Hz, 325 Hz, 490 Hz, 640 Hz, 820 Hz, 970 Hz, 1140 Hz, dan 1465 Hz. Potongan suara yang frekuensi dominannya ditunjukkan oleh Gbr. 4 adalah suara vokal O yang ditunjukkan oleh Gbr 2.



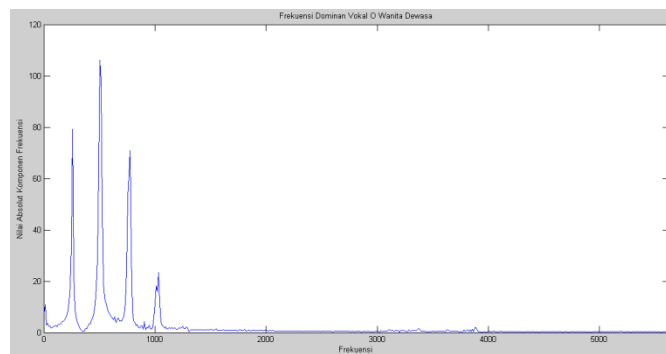
Gbr. 4 Enam Frekuensi Dominan dari vokal O seorang pria dewasa.

Jumlah frekuensi dominan yang dimiliki oleh masing-masing potongan rekaman memiliki perbedaan yang cukup signifikan. Gbr. 5 menunjukkan frekuensi dominan dari respon frekuensi suara ucapan vokal O yang diucapkan oleh seorang pria dewasa yang berbeda dengan Gbr. 2. Terlihat frekuensi dominan yang dimiliki adalah berjumlah tiga belas yaitu pada kelompok frekuensi yang bersesuaian dengan frekuensi 50 Hz, 170 Hz, 335 Hz, 500 Hz, 785 Hz, 850 Hz, 970 Hz, 3695 Hz, 3750 Hz, 4140 Hz, 4285 Hz, 7415 Hz, dan 7750 Hz.



Gbr. 5 Frekuensi Dominan Vokal O Pria Dewasa yang berbeda.

Sebagai perbandingan lebih jauh, seperti ditunjukkan oleh Gbr. 6 suara vokal O yang diucapkan oleh seorang wanita dewasa memiliki delapan buah frekuensi dominan yaitu pada kelompok frekuensi yang bersesuaian dengan frekuensi 150 Hz, 265 Hz, 510 Hz, 775 Hz, 1030 Hz, 3035 Hz, 3375 Hz dan 3880 Hz.



Gbr. 6 Frekuensi Dominan Vokal O Wanita Dewasa.

Kumpulan informasi frekuensi dominan dari seluruh hasil rekaman disimpan ke dalam sebuah basis data sebagai data suara ucapan vokal Bahasa Indonesia. Empat informasi utama yang disimpan adalah fonem yang diucapkan, jumlah frekuensi dominan yang dikandung, frekuensi yang bersesuaian dan nilai absolut atau amplitudo pada frekuensi tersebut. Informasi juga dilengkapi keterangan mengenai kategori sumber ucapan (pria-wanita, dewasa-anak-anak dan etnis), komputer yang digunakan dan waktu perekaman.

IV. PENUTUP

Kumpulan data suara vokal Bahasa Indonesia yang berisikan informasi mengenai frekuensi yang dominan dari suara ucapan lima vokal utama dalam Bahasa Indonesia telah dihasilkan. Sumber ucapan diambil dari banyak pengucap yang berbeda jenis kelamin, usia dan etnis. Informasi mengenai frekuensi dominan tersebut bisa dimanfaatkan untuk data latih sistem pengenalan ucapan otomatis. Data ini bisa dijadikan sebagai masukan untuk modul pengenal ASR Bahasa Indonesia. Selain itu, berdasarkan perbedaan frekuensi dominan yang signifikan diantara pengucap, data ini dapat juga dimanfaatkan untuk sistem pengenalan pengucap otomatis (*Automatic Speaker Recognition*) dalam Bahasa Indonesia.

Pengembangan lebih lanjut dari penelitian ini adalah dengan mengatur ulang pengelompokan data untuk meningkatkan kinerja. Dilakukan pemetaan dari banyak ke satu untuk kelompok frekuensi dominan yang mirip walaupun bersumber dari fonem yang berbeda. Hal ini dapat membuat keluaran modul pengenal menjadi lebih toleran terhadap perubahan. Penelitian ini dapat juga dilanjutkan dengan mencari himpunan frekuensi dominan yang minimal yang masih dapat membedakan antar kelompok dengan baik. Hal tersebut ditujukan untuk menurunkan kebutuhan komputasi sistem ASR secara keseluruhan.

UCAPAN TERIMA KASIH

Penelitian ini didukung oleh Lembaga Riset dan Pengabdian Masyarakat Universitas Presiden. Terima kasih diucapkan atas bantuan dan dukungan yang telah diberikan

selama ini. Selain itu juga terima kasih dan penghargaan disampaikan kepada semua pihak yang telah terlibat secara langsung maupun tidak langsung yang tidak dapat disebutkan satu persatu disini.

REFERENSI

- [1] S. Furui, "Automatic speech recognition and its application to information extraction", Association for Computational Linguistics, hal. 11–20, 1999.
- [2] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2. hal. 257–286, 1989.
- [3] H. Alwi, S. Dardjowidjojo, H. Lapoliwa, and A. M. Moeliono, Tata Bahasa Baku Bahasa Indonesia, Edisi Ketiga, Jakarta: Balai Pustaka, 2014.
- [4] Y. Yoo, Tutorial on Fourier theory, (2001).
- [5] P. Duhamel and M. Vetterli, "Fast Fourier transforms: a tutorial review and a state of the art", Signal Processing, vol. 19, hal. 259–299, 1990.
- [6] (2016) The MathWorks Documentation website. [Online], <http://www.mathworks.com/help/matlab/ref/fft.html> tanggal akses: 12 Juni 2016.

