Karakterisasi Suara Vokal dan Aplikasinya Dalam Speaker Recognition

Siwi Setyabudi, Agus Purwanto dan Warsono
Laboratorium Getaran dan Gelombang, Jurdik Fisika, FMIPA, UNY

ABSTRAK

Penelitian ini bertujuan untuk mendapatkan karakter bunyi vokal pada aksen Jawa dan kemudian menggunakannya dalam program speaker recognition.

Metode penelitian yang digunakan adalah dengan merekam kata 'buka' yang diucapkan dalam aksen Jawa oleh dua orang laki-laki *native speaker* bahasa Jawa. Hasil rekaman tersebut kemudian dibagi ke dalam potongan-potongan sinyal sepanjang 16 ms. Empat potongan sinyal diambil sebagai sampel untuk masing-masing fonem vokal yaitu /u/ dan /a/ dan ditentukan komponan frekuensi dan rasio amplitudo yang menjadi karakteristik masing-masing fonem dengan DFT. Perekaman kedua dilakukan dan kemudian dibandingkan dengan masing-masing fonem acuan tadi dengan menggunakan fungsi cross-correlation.

Hasil penelitian menunjukkan bahwa masing-masing vokal memiliki karakteristik pada puncak-puncak domain frekuensi. Sedangkan cross-correlation untuk suara orang yang sama menghasilkan tingkat kecocokan relatif lebih tinggi dibandingkan dengan suara orang yang berbeda.

Kata kunci: *vokal, DFT, frekuensi, rasio amplitudo, cross-correlation*

PENDAHULUAN

Bayangkan jika terjadi suatu kasus di mana bukti yang ada hanya sebuah rekaman suara. Namun tidak ada seorangpun yang mengenal suara tersebut. Tapi jika pihak yang berwajib memiliki sebuah alat yang mampu mengenali identitas seseorang dari suara, kasus ini mungkin akan terpecahkan. Tapi bagaimana kita bisa membuat alat yang mampu mengenali identitas seseorang dari suaranya?

Suara manusia dihasilkan oleh pita suara yang kemudian diteruskan ke rongga suara yaitu mulut dan rongga hidung. Terutama di rongga mulut suara akan diubah menjadi bunyi-bunyian yang berbeda-beda tergantung dari posisi alat-alat seperti lidah, bibir, dan rahang. Karena bentuk dan ukuran rongga dan alat-alat tersebut berbeda-beda pada tiap orang dan juga perbedaan cara pengucapan suatu bunyi itulah yang menyebabkan karakter suara dari masing-masing orang berbeda-beda.

Vokal seperti /a/, /i/, /u/, /e/, dan /o/ merupakan suara manusia yang sesungguhnya karena sebagian besar suara orang yang kita dengar sebenarnya

Dipresentasikan dalam SEMINAR NASIONAL MIPA 2007 dengan tema "**Peningkatan Keprofesionalan Peneliti, Pendidik & Praktisi MIPA**" yang diselenggarakan oleh Fakultas Matematika dan Ilmu Pengetahuan Alam UNY, Yogyakarta pada tanggal 25 Agustus 2007.

adalah vokal dan oleh karenanya karakter suara seseorang dapat dilihat dari suara vokalnya. Untuk dapat memperoleh karakter suatu vokal terlebih dahulu sebuah sinyal suara vokal diubah ke dalam domain frekuensi. Sedangkan untuk dapat mengenali suara seseorang, data suara orang tersebut diperlukan sebagai acuan yang kemudian akan diverifikasi dengan suaranya yang lain menggunakan cross-correlation. Penelitian ini bertujuan untuk mengetahui domain frekuensi masing-masing vokal dan verifikasi suara. Pengetahuan tentang domain frekuensi dapat digunakan lebih lanjut dalam sintesis suara sedangkan verifikasi suara atau speaker recognition dapat digunakan dalam bidang keamanan sebagai tanda identitas seseorang.

KAJIAN PUSTAKA

Untuk menganalisis suatu sinyal kita dapat melihatnya dari domain waktu maupun domain frekuensi. Salah satu alat yang sangat penting untuk melakukan tugas tersebut adalah transformasi Fourier yang dinyatakan dalam persamaan sebagai berikut

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-i\omega t} dt$$

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) e^{i\omega t} d\omega$$

Kedua persamaan tadi merupakan suatu pasangan, maksudnya adalah bahwa persamaan yang satu merupakan transformasi dari persamaan yang lain.

Sementara itu untuk membandingkan antara suatu sinyal dengan sinyal lain dapat dinyatakan dalam persamaan cross-correlation berikut ini

$$R_{xy}(\tau) = \lim_{T\to\infty} \int_{0}^{T} x(t) y(t+\tau) dt$$

di mana $x(t)$ adalah suatu sinyal acuan dan $y(t)$ adalah sinyal lain yang dibandingkan. Namun perhitungan langsung dengan persamaan ini memakan waktu terlalu lama. Oleh karena itu, diperlukan persamaan yang dapat dikerjakan

dengan lebih cepat. Salah satunya adalah dengan transformasi forier sehingga persamaan tersebut menjadi

$$\int_{-\infty}^{\infty} x(t)y(t+\tau)dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)\int_{-\infty}^{\infty} Y(\omega)e^{i\omega(t+\tau)}d\omega dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} Y(\omega)\left[\int_{-\infty}^{\infty} x(t)e^{i\omega(t)}dt\right]e^{i\omega\tau}d\omega$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} X*(\omega)Y(\omega)e^{i\omega\tau}d\omega$$

di mana $X*(\omega)$ merupakan kompleks konjugat dari $X(\omega)$. Persamaan ini dapat digunakan untuk menentukan tingkat kesamaan atau kemiripan suatu sinyal terhadap sinyal yang lain.

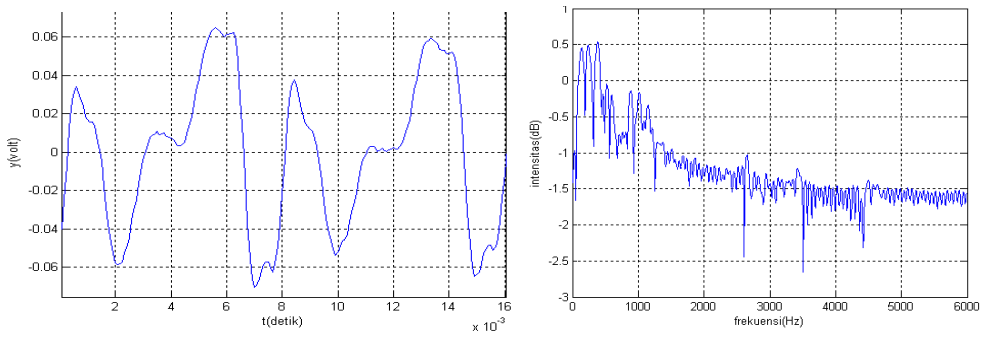
METODOLOGI PENELITIAN

Penelitian ini dilakukan di Laboratorium Getaran dan Gelombang, Jurdik Fisika, FMIPA UNY dengan menggunakan sampel suara dua orang laki-laki berusia 22 dan 23 tahun dan menggunakan aksen Jawa. Sampel suara direkam ke dalam komputer menggunakan microphone condensor yang dihubungkan dengan ADC (*soundcard*) dengan software MATLAB® 6.5.1.

Data yang diambil merupakan potongan vokal /u/ dan /a/ sepanjang 16 ms dari kata 'buka' yang diucapkan dengan nada yang diusahakan sama dan direkam pada sampling rate 12000 Hz. Analisis dan verifikasi suara (speaker recognition) dilakukan dengan menggunakan program pada MATLAB® 6.5.1.

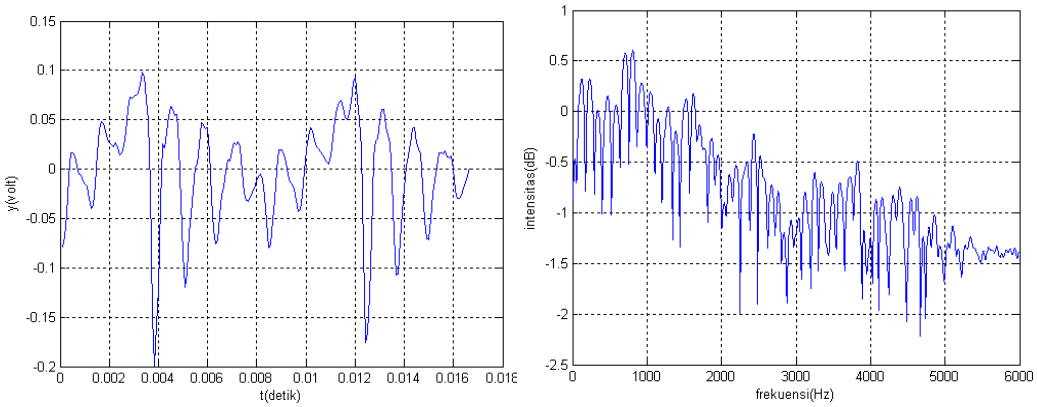
HASIL DAN PEMBAHASAN

Berikut ini adalah gambar potongan sinyal suara vokal /u/ dan /a/ dari kata 'buka' oleh orang pertama dan domain frekuensi serta respon frekuensinya. Domain frekuensi diperoleh dengan program DFT menggunakan MATLAB® 6.5.1.



(a) (b)

Gambar 1. (a) Domain waktu sinyal suara vokal /u/. (b) Domain frekuensi dan respon frekuensi sistem vokal /u/.



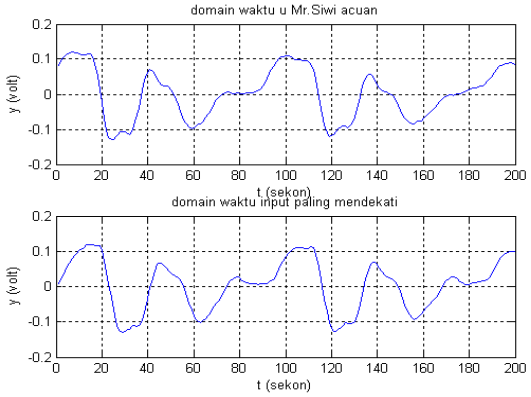
(a) (b)

Gambar 2. (a) Domain waktu sinyal suara vokal /a/. (b) Domain frekuensi dan respon frekuensi sistem vokal /a/.

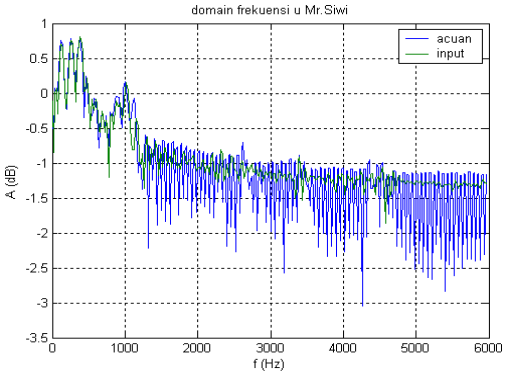
Dari gambar di atas dapat dilihat bahwa puncak-puncak domain frekuensi pada fonem /u/ berada pada sekitar frekuensi 350 Hz dengan intensitas (relatif) 0.5 dB pada puncak pertama dan sekitar 1000 Hz dengan intensitas (relatif) -0.1 dB pada puncak kedua. Sedangkan frekuensi fundamentalnya adalah 140 Hz dengan intensitas 0.45 dB. Secara relatif perbandingan amplitudo puncak pertama terhadap puncak kedua adalah 1 : 0.25. sedangkan pada vokal /a/ puncak pertama pada frekuensi 120 Hz dengan intensitas (relatif) 0.32 dB yang juga merupakan frekuensi fundamentalnya. Puncak kedua sekitar 820 Hz dengan intensitas(relatif) 0.60 dB yang merupakan frekuensi dengan intensitas tertinggi, puncak ketiga 1650 Hz dengan intensitas (relatif) 0.23 dB, puncak keempat 2500 Hz sebesar -0.23 dB, puncak kelima 3800 Hz -0.50 dB, dan puncak keenam 4500 Hz -0.75 dB.

Secara relatif perbandingan amplitudo puncak-puncak tersebut adalah 0.52 : 1 : 0. 43 : 0.15 : 0.08 : 0.04.

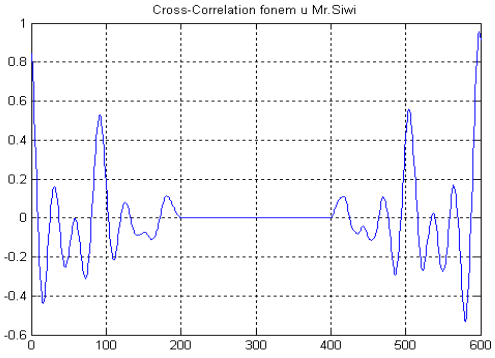
Penelitian selanjutnya mengenai speaker recognition dilakukan dengan data acuan yaitu berupa potongan sinyal vokal /u/ dan /a/ dari kata 'buka' oleh orang pertama yang direkam pada tanggal 27 Juli 2007 pukul 13:53 WIB yang telah dianalisis di atas dan orang kedua yang direkam pada tanggal 28 Juli 2007 pukul 23:30 WIB. Pengujian verifikasi identitas pertama dilakukan dengan kata 'buka' oleh orang pertama yang direkam pada tanggal 28 Juli 2007 pukul 15:26. Hasilnya adalah sebagai berikut:



(a)



(b)



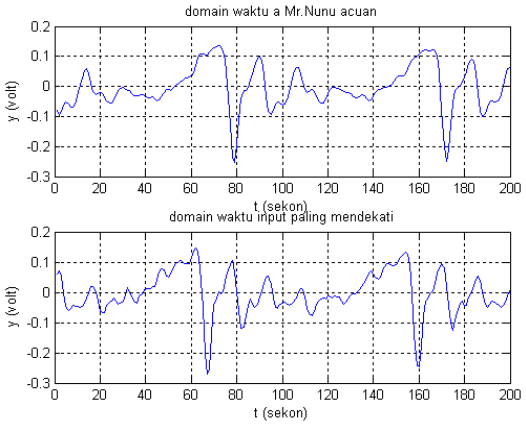
(c)

Hasil keluaran program speaker recognition:
Gmax = 9.5563e-001
To = 1
Identitas = Mr. Siwi (orang pertama)

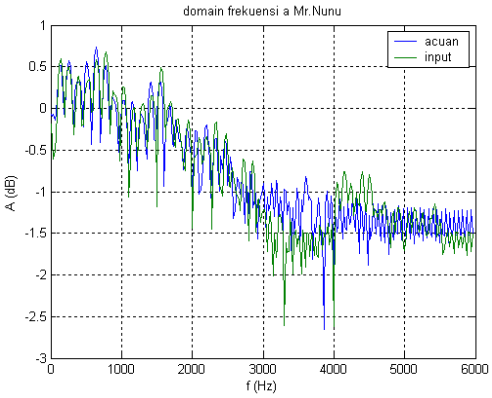
Gambar 3. (a) domain waktu sinyal acuan dan sinyal teruji maksimum.
(b) domain frekuensi sinyal acuan dan sinyal teruji maksimum.
(c) Cross-Correlation sinyal acuan dan sinyal teruji maksimum.

Sedangkan untuk orang kedua diverifikasi juga dengan kata 'buka' yang direkam pada tanggal 28 Juli 2007 pukul 22:51 WIB. Verifikasi tidak dilakukan

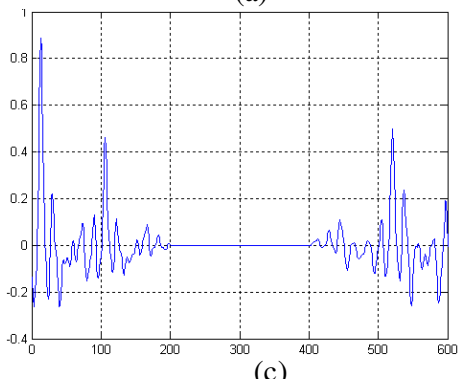
seara lifetime tetapi terlebih dahulu merekam kata ‘buka’ sebanyak-banyaknya kemudian memilih secara acak untuk dijadikan sampel. Nilai minimal Gmax tiap pengujian sehingga diterima (dikenali) adalah 0.85. Hasilnya adalah sebagai berikut:



(a)



(b)



(c)

Hasil keluaran program speaker recognition:
Gmax = 8.8894e-001
To = 4
Identitas = Mr. Nunu (orang kedua)

Gambar 4. (a) domain waktu sinyal acuan dan sinyal teruji maksimum.
(b) domain frekuensi sinyal acuan dan sinyal teruji maksimum.
(c) Cross-Correlation sinyal acuan dan sinyal teruji maksimum.

Gmax adalah nilai cross-correlation maksimum dari semua pengujian, To adalah nomor data teruji maksimum. Pada verifikasi identitas pertama Gmax sebesar 9.5563e-001 adalah pada To = 1 yang berarti pada data pertama atau data vokal /u/ oleh orang pertama. Sedangkan nilai cross-correlation yang lain adalah Gm2 = 8.4682e-001, Gm3 = 7.6024e-001, Gm4 = 7.3146e-001. Dari hasil ini dapat dilihat bahwa Gm2 yang merupakan cross-correlation terhadap data kedua atau vokal /a/ orang pertama juga memiliki nilai lebih besar dari Gm3 dan Gm4 yang merupakan cross-correlation terhadap data vokal /u/ dan /a/ orang kedua.

Pada verifikasi identitas kedua Gmax = 8.8894e-001 berada pada To=4 atau terhadap data vokal /a/ orang kedua. Sedangkan hasil cross-correlation lain yaitu Gm1 = 7.6763e-001, Gm2 = 7.8942e-001, Gm3 = 6.0413e-001 tampaknya menunjukkan bahwa pada pengucapan vokal /u/ orang kedua lebih menyerupai orang pertama. Hal ini bisa saja terjadi karena memang pada saat perekaman digunakan kata 'buka' orang pertama sebagai contoh yang harus ditirukan. Namun hasil ini belum melampaui nilai minimum sebesar 0.85 sehingga belum dapat diterima atau dikenali sebagi suara yang sama.

KESIMPULAN

1. Setiap vokal memiliki karakteristik yang dapat dilihat dari puncak-puncak domain frekuensi yaitu pada vokal /u/ pada 350 Hz dan 1000 Hz dengan rasio 1 : 0.25 sedangkan pada vokal /a/ pada 120 Hz, 820 Hz, 1650 Hz, 2500 Hz, 3800 Hz, dan 4500 Hz dengan rasioamplitudo 52 : 1: 0. 43 : 0.15 : 0.08 : 0.04.
2. Cross-correlation untuk suara orang yang sama menghasilkan tingkat kecocokan yang lebih relatif tinggi dibandingkan dengan suara orang lain sehingga dapat dimanfaatkan sebagai alat verifikasi identitas atau speaker recognition.

DAFTAR PUSTAKA

Karris, Steven T**.** (2003). *Signals and Systems with MATLAB® Applications, Second Edition.*California: Orchard Publications.