

SPEECH RECOGNITION FOR CLEFT LIP AND PALATE VOICE AND STANDARD VOICE FOR CONSONANT WORDS /B/

Akhmad Anggoro^{*}, Samiadji Herdjunto, Risanuri Hidayat

Departemen Teknologi Informasi dan Teknik Elektro, Fakultas Teknik,
Universitas Gadjah Mada, Yogyakarta

Jl. Grafika No. 2 , Kampus UGM, Sinduadi, Mlati, Senolowo, Sinduadi, Kec. Mlati,
Kabupaten Sleman, Daerah Istimewa Yogyakarta

Email : ^{*}Akhmad.anggoro@mail.ugm.ac.id

Abstract

Advances in technology make speech recognition improve. But does speech recognition recognize the sounds of cleft lip and palate? this research uses the voice of cleft lip and palate to normal voice. With the letter / b / which is the letter for lip articulation. Words used include Clothes, Ash and Moist. The extraction method uses Mel Frequency Cepstral Coefficients (MFCC), the classification uses K-Nearest Neighbor (KNN) with K-Fold Cross-Validation as a test. The results show the accuracy above 70%. 75% in the word "Baju". 75% in the word "Abu". 83% in the word "Lembab".

Keywords : Cleft Lip Palate. MFCC, KNN, K-Fold Cross Validation..

1. Pendahuluan

Pengenalan otomatis telah banyak berkembang. Saat ini pengenalan mampu mengenali puluhan ribu kata dan algoritma dapat menguraikan dengan cepat sehingga dapat melakukan pengenalan ucapan secara terus-menerus. [1].

Persyaratan dasar dalam pengenalan ucapan adalah mengekstraksi serangkaian fitur untuk masing-masing ucapan. Pengaplikasian metode ekstraksi pada pengolahan ucapan untuk beberapa suku kata dan fonem telah dilakukan. Seperti menggunakan metode wavelet Daubechies [1], [2]. Namun tidak banyak penelitian mengenai fonem terhadap ucapan Bahasa Indonesia. Terutama ucapan penderita bibir sumbing dan langit-langit atau dikenal dengan *cleft lip and palate* (CLP)

Pengenalan ucapan tentu tidak lepas dari suara. Suara merupakan alat komunikasi penting pada manusia. Namun tidak semua manusia memiliki suara normal. Dalam kasus ini penderita bibir sumbing dan langit-langit (CLP) penderita CLP merupakan penderita yang lahir dengan kelainan bentuk bibir atas, langit-langit maupun gabungan keduanya. Yang menyebabkan suara terganggu, sehingga komunikasi pun terganggu. Penderita CLP umumnya mengalami gangguan pada artikulasi wicara. Dimana terjadi perubahan suara hingga sulit dimengerti (sengau). Penelitian pengenalan pola suara penderita CLP masih minim di Indonesia [3]. Oleh sebab itu pada penelitian ini akan menguji pengenalan ucapan penderita CLP dan Normal pada masalah artikulasi wicara untuk konsonan /b/ kata yang akan diuji antara lain baju, abu, dan lembab. Menggunakan metode ekstraksi MFCC dengan klasifikasi KNN serta validasi silang K-Fold

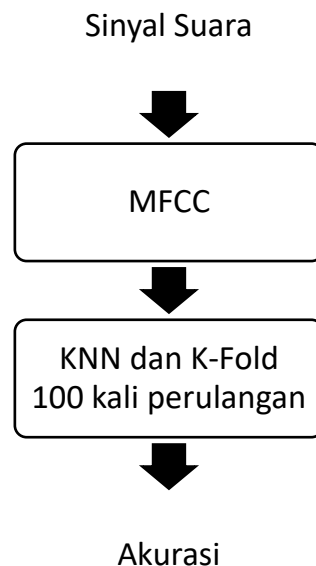
2. Metode Penelitian

2.1 Pengumpulan Data

Perekaman suara dilakukan terhadap 20 orang pembicara. Masing-masing 10 orang CLP dan 10 orang normal. Suara yang direkam berupa kalimat dari suara konsonan /b/ dalam Bahasa Indonesia, yaitu kalimat baju, abu dan lembab. Alat perekaman berupa mikrofon yang terintegrasi dengan *microphone*. Perangkat lunak perekaman adalah Audacity. Proses perekaman dilakukan dengan frekuensi sampling 8.000 Hz, mono, PCM 64 bit. Standar pengucapan suku kata mengikuti standar pengucapan pada International Phonetic Association (IPA) (International Phonetic Association n.d.). Perekaman dilakukan di dalam ruangan untuk dapat mengurangi pengaruh derau latar. Proses selanjutnya menggunakan matlab 2016.

2.2 Alur penelitian

Alur penelitian ini adalah memproses suara yang telah direkam. Kemudian suara diekstraksi menggunakan fitur MFCC. Lalu hasil ekstraksi dipasangkan berdasarkan CLP dan normal. Kemudian diklasifikasi menggunakan KNN dan validasi silang K-Fold menggunakan nilai 2 hingga 10 dengan perulangan 100 kali. [4]



Gambar. 1. Alur Penelitian

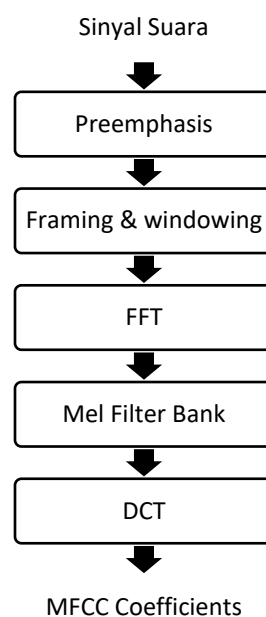
2.3 Bibir Sumbing

Bibir sumbing dan langit-langit adalah cacat bawaan sejak lahir. Kelainan ini terjadi akibat gangguan dalam proses penyatuan bibir atas pada masa embrio awal. Tidak jarang ada yang mengalami keduanya. Penderita bibir sumbing akan mengalami gangguan pada suara (sengau) yang mengakibatkan pengenalan ucapan bermasalah. Angka kejadian CLP di Indonesia mencapai 2 dari setiap 1.000 kelahiran bayi [5] Pembeda makna pada suara atau bunyi disebut fonem. Seperti dengan mudah dapat kita buktikan dengan pasangan paku dan selam yang merupakan fonem /p/ dan /s/.

Perbedaan antara vokal dan konsonan hanya terletak pada ada tidaknya halangan ketika bunyi-bunyi itu diucapkan. Bunyi bahasa yang dihasilkan dengan getaran pita suara tanpa adanya penyempitan dalam saluran suara di atas glotis disebut dengan vokal. Contohnya: a, i, u, e, o. Konsonan adalah bunyi bahasa yang dihasilkan dengan menghambat saluran udara. Contohnya: b, c, d, f, dst. [6].

2.4 Mel Frequency Cepstral Coefficients (MFCC)

Metode ekstraksi fitur yang akan digunakan dalam penelitian ini adalah *Mel Frequency Cepstral Coefficients* (MFCC). Yang sudah banyak digunakan dalam pengenalan ucapan pada penelitian sebelumnya. [7]–[11] mulai dari pengenalan pola hingga pengenalan ucapan. Berikut adalah urutan dasar metode MFCC.

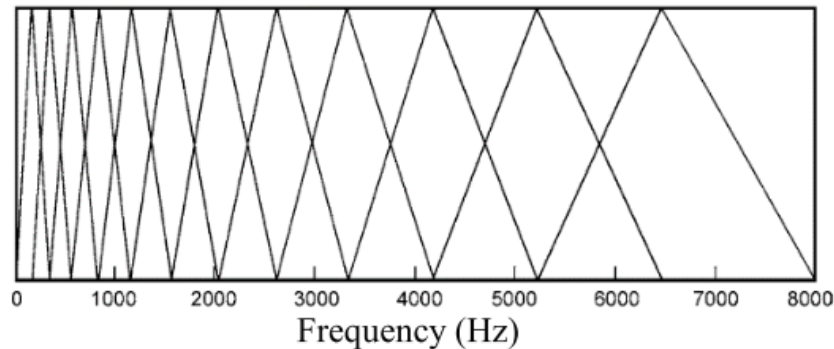


Gambar. 2. Blok Diagram ekstraksi fitur MFCC

- *Preemphasis*: *Preemphasis* digunakan untuk mengkompensasi bagian frekuensi tinggi dari sinyal bicara yang ditekan selama mekanisme produksi suara manusia [8].
- *Framing dan windowing*: Dalam proses framing, sinyal dibagi menjadi bagian yang lebih pendek. Panjang sinyal yang terbagi sekitar 20-40 ms. Sinyal yang dibagi 10 ms tumpang tindih dengan sinyal sebelumnya dan berikutnya [9]. *Windowing* digunakan untuk menghindari diskontinuitas antar sinyal. Jenis jendela yang paling banyak digunakan adalah *window hamming* [8]
- FFT: Dalam transformasi *Fourier*, ada algoritma yang memiliki perhitungan yang sangat cepat untuk melakukan transformasi *Fourier* dalam domain diskrit, biasanya disebut FFT (*Fast Fourier Transform*). FFT adalah algoritma dengan perhitungan cepat dari *Discrete Fourier Transform* (DFT) [10].
- *Mel Filter bank Processing*. Proses perhitungan nilai *Mel* dalam satuan Frekuensi. Filterbank adalah filter bandpass yang saling melalui (*overlaps*) yang berbasis skala *Mel* dengan frekuensi dibawah 1 kHz. Rumus skala *Mel* adalah sebagai berikut.

$$mel = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad 1)$$

- Dengan mel adalah output sedangkan f adalah input. Dengan angka 2595 dan 700 adalah standard pengolahan mel.



Gambar. 3. Frekuensi

- DCT (*Discrete Cosine Transform*) adalah proses terakhir dari MFCC dimana akan menghasilkan coefficients. Jumlah data atau fitur yang dihasilkan berkisar 13 fitur. Dengan dimensi data mencapai 26 fitur dimensi data

2.5 K-Nearest Neighbor (KNN).

Metode klasifikasi secara umum banyak yang telah digunakan dalam penelitian antara lain Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) [12], *Hidden markov models*[13][14] dan KNN [4], [15], [16].

Salah satu algoritma klasifikasi statistik yang digunakan untuk mengklasifikasikan objek berdasarkan contoh pelatihan terdekat adalah KNN hanya memerlukan beberapa parameter untuk menyempurnakan akurasi klasifikasi tinggi, yaitu K dan metrik jarak. Biasanya, K adalah angka ganjil sehingga dapat menghindari suara terikat [17]. Metode jarak yang digunakan dalam penelitian ini adalah jarak euclidean. Persamaan berikut [18] adalah bentuk matematika dari jarak *Euclidean*

$$d_{st} = \sqrt{\sum_{j=1}^n (x_{sj} - y_{tj})^2} \quad 2)$$

2.6 Validasi Silang K-Fold

Untuk data set rendah menggunakan validasi silang k-fold. Teknik ini menjadikan data dibagi menjadi beberapa K dataset. Data set digunakan sebagai data pengujian dan data pelatihan. Jika data 1 sebagai pengujian, maka data ke-2, ke-3 dan seterusnya akan menjadi pelatihan. Proses diulang sebanyak K kali dengan akurasi yang juga dibagi dengan nilai K. Total data dalam penelitian ini adalah 60 dengan pembagian 3 kata. Penelitian ini menggunakan KNN 2 hingga 10 dan K-Fold 2 hingga 10. Dengan 100 kali perulangan setiap pengujiannya.

2.7 Perhitungan Tingkat Pengenalan.

Evaluasi hasil penelitian terhadap tingkat akurasi pengenalan sistem akan menggunakan ROC Untuk setiap data yang diuji, akan dilihat apakah data tersebut mengklasifikasi dengan benar atau tidak. [9]

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad 3)$$

Ket :

True Positive (TP)

True Negative (TN)

False Positive (FP)

False Negative (FN)

3. Hasil Dan Diskusi

Model Model yang digunakan pada penelitian ini terdiri dari 60 sinyal suara dengan CLP dan normal. Data terbagi menjadi 4 bagian. Data yang terdiri 3 kata (Baju, Abu, Lembab) yang masing-masing 10 CLP dan 10 normal. Lalu data terakhir adalah total semua data yang berjumlah 60 data. Validasi Silang K-Fold diterapkan pada model yang diusulkan dengan jumlah K = 2 sampai 10. Demikian juga untuk Nilai K dari KNN yang diusulkan. Tujuannya adalah menemukan kombinasi KNN dan Validasi silang mana yang memiliki akurasi tertinggi. Untuk ekstrasi menggunakan 27 filter MFCC dengan 13 koefisien yang dijadikan sebagai fitur. [7]

Tabel 1 tabel spesifikasi data

no	data record	baju, abu, lembab
1	Jumlah sampel	20 orang
2	Pengulangan Perekaman	1
3	Tempo perekaman	Sesuai IPA
4	Frequency Sampling	8.000 Hz
5	Estimasi durasi	1 detik
6	Lingkungan Perekaman	Tertutup
7	Format Penyimpanan Data	*.wav

Tabel 2. Hasil uji coba dengan hasil maksimal

Maks Result			
Baju	Knn	2	75%
	K-Fold	7	
Abu	Knn	2	75%
	K-Fold	7	
Lembab	Knn	3	89%
	K-Fold	8	
B Semua	Knn	8	73%
	K-Fold	10	

Tabel menunjukkan nilai akurasi maksimal adalah 89% dengan akurasi terendah untuk semua data kata adalah 73%. Nilai KNN dan K-Fold juga berbanding terbalik. Pada hasil maksimal nilai KNN berada pada Nilai 2 dan 3 Namun K-Fold pada nilai tertinggi 7 – 10.

Tabel 3. Hasil uji coba dengan hasil minimal

Min Result			
Baju	Knn	9	42%
	K-Fold	2	
Abu	Knn	10	41%
	K-Fold	2	
Lembab	Knn	10	39%
	K-Fold	2	
B Semua	Knn	2	58%
	K-Fold	9	

Tabel menunjukkan nilai akurasi minimal adalah 39% dengan akurasi terendah untuk semua data kata adalah 58%. Nilai KNN dan K-Fold juga berbanding terbalik. Pada hasil minimal nilai KNN berada pada Nilai 10 dan 9. Namun K-Fold pada nilai terendah, yaitu 2.

Metode MFCC dengan Klasifikasi KNN dan diuji dengan validasi silang K-Fold menunjukkan akurasi diatas 70%. 75% pada kata Baju, 75% pada kata Abu, 83% pada kata Lembab dan 73% pada total data. Hasil ini terbilang memuaskan. Dengan presentase minimal akurasi 39% pada kata lembab.

Hasil yang lain adalah hubungan KNN dan validasi silang K-Fold dalam klasifikasi fitur MFCC. KNN yang berbanding kebalik. Nilai KNN kecil dengan K-Fold besar menunjukkan hasil akurasi maksimal. Sedangkan nilai KNN Besar dengan K-Fold Kecil justru menunjukkan nilai akurasi kecil

4. Kesimpulan

Metode ekstraksi MFCC dengan Klasifikasi KNN menggunakan validasi silang K-Fold dapat mengidentifikasi suara CLP terhadap suara normal pada huruf B yang merupakan huruf artikulasi bibir atas. Tingkat akurasi diatas 80% dengan terendah pada 41%.

Hasil penelitian membuktikan metode MFCC dengan klasifikasi menggunakan KNN menggunakan validasi silang K-Fold dapat mengklasifikasi suara CLP dan suara normal terhadap huruf bibir /b/ pada kata Baju, Abu, dan Lembab. Penelitian ini membuktikan Ekstraksi fitur MFCC dapat mengenali suara penderita CLP dengan akurasi yang baik dan tinggi.

5. Saran

- Penelitian berikutnya dapat mencoba pada ekstraksi fitur yang berbeda seperti *Wavelet* dan *Linear Predictif Coding*
- Penelitian berikutnya dapat mencoba klasifikasi yang masih satu keluarga seperti *Linear Regresi* dan *Centroid*.
- Diharapkan pada penelitian berikutnya sudah dapat menghasilkan produk yang bisa dipakai oleh orang banyak.

6. Ucapan Terimakasih

Dalam pembuatan paper ini banyak pihak yang membantu penulis sehingga dapat menyelesaikan paper ini, untuk itu Penulis mengucapkan terimakasih kepada :

- Dosen Pembimbing bapak Samiadji Herdjunto dan bapak Risanuri Hidayat.
- Teman dan guru Siti Agrippina Alodia Yusuf
- Terapis wicara dan semua teman yang sudah membantu penelitian ini
- Para partisipan baik bibir sumbing dan normal.

Daftar Pustaka

- [1] Farooq, O., & Datta, S. (2003). Phoneme recognition using wavelet based features. *Information Sciences*, 150(1-2), 5-15.
- [2] Hidayat, R., & Ikawijaya, W. (2015, November). Wavelet based feature extraction for the vowel sound. In 2015 International Conference on Information Technology Systems and Innovation (ICITSI) (pp. 1-4). IEEE.
- [3] Sardjono, Terapi Wicara. Jakarta: Departemen Pendidikan Nasional, 2005.
- [4] Winursito, A., Hidayat, R., & Bejo, A. (2018, March). Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition. In 2018 International Conference on Information and Communications Technology (ICOIACT) (pp. 379-383). IEEE.
- [5] Supandi, A. (2014). Angka Kejadian Sumbing Bibir Di Rsup Prof. Dr. Rd Kandou Manado Periode 2011-2013. *e-CliniC*, 2(2).
- [6] Yeni, F. (2008). Interpretasi Lafal Fonem Penderita Bibir Sumbing.
- [7] Hidayat, S., Hidayat, R., & Adji, T. B. (2015). Sistem pengenalan tutur bahasa Indonesia berbasis suku kata menggunakan mfcc, wavelet dan hmm. In Conference on Information Technology and Electrical Engineering (CITEE) (pp. 246-251).
- [8] A. Setiawan, A. Hidayatno, R. R. Isnanto, J. T. Elektro, F. Teknik, and U. Diponegoro, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients

- (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer,” vol. 13, no. 3, pp. 82–86, 2011.
- [9] Susanti, E., & Al Sasongko, S. M. (2017). Klasifikasi Suara Berdasarkan Usia Menggunakan Mel Frequency Cepstral Coefficient (Mfcc) Dan K-Nearest Neighbour (K-NN). *DIELEKTRIKA*, 4(2), 120-126.
- [10] Orozco-Aroyave, J. R., Vargas-Bonilla, J. F., Vásquez-Correa, J. C., Castellanos-Domínguez, C. G., & Nöth, E. (2016). Automatic detection of hypernasal speech of children with cleft lip and palate from spanish vowels and words using classical measures and nonlinear analysis. *Revista Facultad de Ingeniería Universidad de Antioquia*, (80), 109-123.
- [11] Nieto, R. G., Marín-Hurtado, J. I., Capacho-Valbuena, L. M., Suarez, A. A., & Bolaños, E. A. B. (2014, September). Pattern recognition of hypernasality in voice of patients with Cleft and Lip Palate. In 2014 XIX Symposium on Image, Signal Processing and Artificial Vision (pp. 1-5). IEEE.
- [12] Setiawan, A., Hidayatno, A., & Isnanto, R. R. (2012). Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer (Doctoral dissertation, Diponegoro University).
- [13] B. Darmawan and S. Ariessaputra, “Sistem Pengenalan Dan Verifikasi Pembicara HMM,” *Semastikom 2016*, pp. 1–6, 2016
- [14] Kusumawati, R. (2016). Metode Linear Predictive Coding (LPC) Pada klasifikasi Hidden Markov Model (HMM) Untuk Kata Arabic pada penutur Indonesia. *MATICS*, 8(1), 32-35.
- [15] Satrya, R., Hidayat, I. B., & Budiman, G. (2010). Sistem Identifikasi Suara Pria dan Suara Wanita Berdasarkan Usia Menggunakan Mel Frequency Cepstral Coefficient dan K-Mens Clustering.
- [16] He, L., Zhang, J., Liu, Q., Yin, H., Lech, M., & Huang, Y. (2015). Automatic evaluation of hypernasality based on a cleft palate speech database. *Journal of medical systems*, 39(5), 61.
- [17] Saini, I., Singh, D., & Khosla, A. (2013). QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases. *Journal of advanced research*, 4(4), 331-344.
- [18] Diker, A., Cömert, Z., Avci, E., & Velappan, S. (2018, May). Intelligent system based on Genetic Algorithm and support vector machine for detection of myocardial infarction from ECG signals. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.