

MFCC dan KNN untuk Pengenalan Suara Artikulasi P

Akhmad Anggoro*, Samiadji Herdjunto, Risanuri Hidayat

Departemen Teknologi Informasi dan Teknik Elektro, Fakultas Teknik,

Universitas Gadjah Mada, Yogyakarta

*email: akhmad.anggoro@mail.ugm.ac.id

Abstract

Cleft lip and palate (CLP) is a term for patients who experience speech organ disorders, that disorder is caused by a gap found in the lip or palate. Patients will experience speech problems. Pattern recognition in CLP sound is still small in Indonesia. In this research in the language identification of CLP and standard sound patterns using the extraction of the Mel Frequency Cepstral Coefficients (MFCC) feature with K-Nearest Neighbor (KNN) classification and K-Fold cross-validation. By making words that have the letter /p/ as a reference, known as bilabial. The words used include Paku, Kapak, and Atap. The accuracy of recognition results reached more than 69%, with a minimum accuracy of 41%.

Keywords — Cleft Lip Palate. MFCC, KNN, K-Fold cross-validation

1. Pengantar

Salah satu gangguan bicara yang dikenal sebagai bibir sumbing dan langit-langit mulut disebut juga *Cleft Lip and Palate* (CLP) adalah cacat bawaan dimana bagian-bagian wajah yang membentuk bibir atas yang terpisah tidak melebur sebelum lahir. Kesenjangan serupa dapat terjadi di langit-langit mulut. Selain itu, penderita CLP mengalami keretakan pada bibir, langit-langit, dan keduanya. Menurut WHO, pada 2011, ada sekitar 1 per 7000 kelahiran hidup, dengan persentase penderita berkisar antara 7,94 hingga 9,92 per 10.000 kelahiran hidup. Dua dari setiap 1.000 kelahiran bayi di Indonesia mengalami bibir sumbing. Selain bibir sumbing sebagian besar juga memiliki kelainan wajah yang lebih luas [14].

Secara umum, banyak penderita CLP mengalami gangguan bicara pada hidung (Nasal) [10,12,15], dan ada juga yang tidak. Penderita bibir sumbing dapat diobati dengan melakukan beberapa operasi, baik operasi bibir dan langit-langit mulut, hingga terapi wicara. Semua dilakukan secara bertahap dan melalui proses yang panjang. Penderita CLP cenderung memiliki masalah komunikasi, terutama dalam artikulasi. Kelainan pada organ bicara menyebabkan penderita komunikasi mengalami gangguan sehingga sulit dikenali oleh orang lain. Artikulasi terganggu karena kelainan pada bibir pasien. Jadi huruf dan kata yang diucapkan juga mengalami gangguan. Studi ini akan terbatas pada artikulasi huruf /p/. Di mana huruf /p/ adalah huruf bibir (Bilabial).

Suara adalah salah satu alat komunikasi penting bagi manusia. Lebih lanjut, dengan CLP suara akan terganggu dan menyebabkan komunikasi menjadi lebih sulit. Kesulitan dalam berkomunikasi akan menimbulkan masalah dalam kehidupan seseorang dan sulitnya mendapatkan informasi, bersosialisasi hingga gangguan psikologis. Di zaman modern, komunikasi dan teknologi informasi tidak jauh dari penggunaan suara sebagai perintah. Sinyal suara analog diproses menjadi digital, maka perintah yang ada dalam sinyal suara diproses. Ini sangat umum. Penggunaan yang paling sering dengan “Google Voice”. Dengan CLP, tentu saja, proses menggunakan perintah suara tidak optimal. Berkomunikasi dengan sesama manusia itu sulit, terutama dengan mesin. Mesin tentu membutuhkan pengenalan pola dalam suara analog. Dan ini adalah tantangan untuk suara CLP.

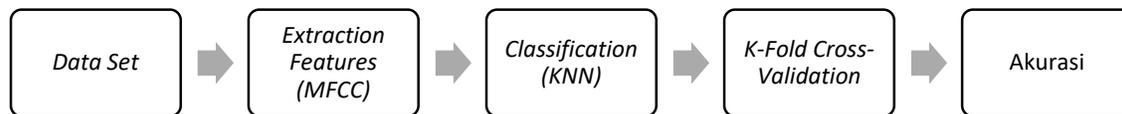
Dalam sinyal suara, empat fase pemrosesan terjadi untuk klasifikasi *preprocessing*, segmentasi, ekstraksi fitur, dan klasifikasi. *Preprocessing* bertujuan untuk menghilangkan komponen yang tidak diinginkan seperti pengurangan noise, peningkatan, dan normalisasi sinyal suara. Kemudian segmentasi adalah langkah selanjutnya setelah *preprocessing*. Tujuan segmentasi adalah untuk membagi sinyal menjadi segmen yang lebih kecil untuk efektivitas yang lebih baik. Ekstraksi fitur adalah salah satu langkah paling kritis untuk sistem diagnostik, dan fitur sinyal suara dapat diekstraksi dalam berbagai domain, seperti mengekstraksi domain waktu atau frekuensi, domain morfologi dan / atau statistik. Sedangkan untuk klasifikasi, secara teoritis, semua pengklasifikasi dapat digunakan untuk klasifikasi sinyal suara seperti Neuro-fuzzy[1], Hidden Markov Model[2,6,9], Support Vector Machine[5], Backpropagation[3.4], dan K-Nearest Neighbor[7].

Penelitian ini menyajikan penggunaan koefisien Mel Frequency Cepstral Koefisien (MFCC) sebagai ekstraktor fitur dan K-Nearest Neighbor (KNN) sebagai penggolong untuk mengidentifikasi dan membedakan sinyal suara normal dan abnormal. MFCC digunakan sebagai ekstraksi fitur suara. Selain itu, KNN akan digunakan untuk mengklasifikasikan suara penderita CLP dari suara normal. Benchmark meliputi pelafalan kata-kata yang memiliki huruf /p/ dalam kata-kata Paku, Kapak, dan Atap. Kemudian data gabungan dari tiga kata ini[13].

Makalah ini akan tersusun dari beberapa bagian yaitu metode, hasil, diskusi dan yang terakhir adalah kesimpulan.

2. Metode

Metodologi yang diusulkan dalam penelitian ini terdiri dari lima langkah. Yaitu Data set, Ekstraksi fitur MFCC, Klasifikasi, validasi silang K-Fold dan akurasi. Gambar 1 menjelaskan tentang metodologi yang digunakan dalam penelitian ini.



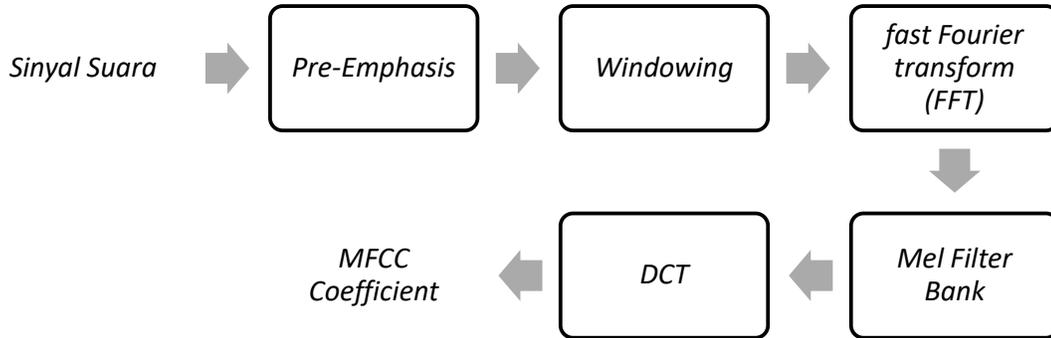
Gambar 1. Metodologi penelitian yang menggambarkan urutan pemrosesan sinyal

2.1 Data Set

Data Set diperoleh dengan merekam suara sebanyak 20 orang. Yang terdiri dari 8 wanita dan 12 pria. Terdiri dari 10 pasien dengan CLP dan sepuluh normal. Kumpulan data adalah tiga kata yang memiliki huruf /p/, baik di awal kata, tengah, dan akhir. Setiap orang mengatakan satu kata, jadi kami mendapatkan 60 set data. Setiap kata telah dikategorikan sebagai celah (1) dan normal (2). Kata-kata yang digunakan adalah Paku untuk /p/ diawal. Kapak untuk /p/ pusat dan Atap untuk /p/ end. /p/ adalah huruf *bilabial* (bibir), yang berperan dalam artikulasi pasien dengan CLP. /p/ dipilih karena bisa mewakili gangguan artikulasi yang diderita oleh CLP. Set data akan diuji untuk kata-kata, dan set data total itu sendiri[13].

2.2 *Mel Frequency Cepstral Coefficients* (MFCC)

Proses penelitian lebih lanjut menggunakan *Software* MATLAB dengan versi 2016. Proses ekstraksi fitur MFCC dapat dilihat pada Gambar 2.



Gambar 2. Ekstraksi Fitur MFCC

- *Pre-emphasis*, proses dimana sinyal akan melewati sebuah *filter* yang dapat melewatkan frekuensi tinggi. Pada penelitian ini frekuensi *cut-off* yang digunakan antara 600-1200 Hz, dikarenakan adanya suara CLP yang cukup menantang, baik dari kejelasan, *noise*, hingga kualitas suara sendiri.
- *Windowing* Tahap untuk memilimalisir lemahnya sinyal ada suara dari bagian awal sampai akhir. Dengan mempertimbangkan blok. Lebar *windowing* yang digunakan pada penelitian ini adalah 512.
- *Fast fourier Transform*. N dan *frame* diubah menjadi domain frekuensi dari domain waktu. Pada penelitian ini menggunakan nilai N sebesar 512.
- *Mel Filter bank Processing*. Proses perhitungan nilai Mel dalam satuan Frekuensi. *Filter Bank* yang digunakan adalah 21 *filter bank*.
- DCT dalam tahap ini adalah perubahan *log Mel Spectrum* menjadi domain waktu lagi dengan menggunakan *Discrete Cosine Transform* kemudian hasilnya disebut MFCC dengan Jumlah akhir *Coefficient* yang digunakan adalah 13 fitur.

2.3 *K-Nearest Neighbor* (KNN).

Klasifikasi *K-Nearest Neighbor* (KNN) dilakukan berdasarkan pembelajaran dengan membandingkan data tes dengan data pelatihan mengikuti Persamaan (1). Latih (n) data atribut, dengan setiap data yang menggambarkan titik tambah ruang dengan dimensi n. Ketika suatu data tidak diketahui di kelasnya, KNN akan mencari pola ruang untuk data pelatihan terdekat, sehingga bisa mengklasifikasikan data sesuai kebutuhan[16].

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

Keterangan :

d_{st} = jarak *Euclidean*

x_{sj} = data latih

y_{tj} = data uji

n = jumlah data

2.4 *K-fold cross-validation*

K-fold cross-validation diterapkan ketika memiliki set data yang rendah. Teknik ini membuat data dibagi menjadi beberapa bagian K dari dataset. Dataset pertama digunakan sebagai data uji, sedangkan dataset kedua, ketiga ... K digunakan sebagai data pelatihan. Kemudian, proses ini diulang di mana dataset dua digunakan sebagai data uji, sementara 1, 3 ... K digunakan sebagai data pelatihan. Proses ini diulang K kali dengan akurasi dibagi dengan K. Karena total data dalam penelitian ini adalah 60, dengan pembagian 3 kata. Karena penelitian ini adalah studi pertama, penelitian ini menggunakan KNN 2 hingga 10 dengan K-Fold 2 hingga 10 dengan 100 pengulangan untuk setiap KNN dan K-Fold. Diharapkan dengan tes ini, hasil yang lebih bervariasi akan diperoleh [17].

2.5 *Acuration*

Mengevaluasi kinerja sistem sangatlah diperlukan, itulah sebabnya untuk mengevaluasinya, metrik kinerja terkenal seperti akurasi (Acc), digunakan. Metrik tersebut diperoleh dari matriks kebingungan yang membentuk empat indeks yang *True Positive* (TP), *True Negative* (TN), *False Positive* (FP) dan *False Negative* (FN). TP menunjukkan jumlah sampel abnormal terdeteksi dengan benar, TN menunjukkan jumlah sampel normal terdeteksi dengan benar, FP menunjukkan jumlah sampel abnormal terdeteksi dengan salah, dan FN menunjukkan jumlah sampel normal terdeteksi dengan tidak benar. Berikut adalah rumus matematikanya. Persamaan 2 menyatakan bentuk-bentuk matematika dari fitur-fitur tersebut. dari K yang digunakan.

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

Keterangan :

True Positive (TP)

True Negative (TN)

False Positive (FP)

False Negative (FN)

3. Hasil dan Diskusi

Model yang diusulkan adalah database yang terdiri dari 60 sinyal suara dengan CLP dan Normal. Data dibagi menjadi empat bagian. Data tiga kata di mana sepuluh CLP dan sepuluh standar dari jumlah total data di mana ada 60 sinyal. Untuk mencapai hasil keseluruhan, validasi silang K-fold diterapkan pada model yang diusulkan dengan beberapa K = 2 hingga 10. Setelah langkah *preprocessing* menggunakan median filter, panjang sinyal disesuaikan ke angka tertentu untuk mengikuti kekuatan dua. Kemudian, setiap sinyal diekstraksi menggunakan MFCC. Dua puluh tujuh filter segitiga diterapkan pada sinyal, kemudian 13 koefisien dari 27 disimpan sebagai fitur [8].

Fitur diumpankan ke pengklasifikasi KNN. Untuk menemukan klasifikasi tertinggi, tentukan nilai K atau lingkungan KNN yang dibutuhkan karena itu mempengaruhi keakuratan sistem. Klasifikasi KNN dikombinasikan dengan validasi silang dan mencoba berbagai nilai ganjil K dalam klasifikasi KNN untuk menghindari ikatan. KNN dari nilai 2 hingga 10 dan validasi silang dari 2 hingga 10. Setiap data akan diulang 100 kali. Tujuannya adalah untuk menemukan kombinasi KNN dan *Cross-Validation*, yang memiliki akurasi tertinggi. Studi ini menemukan dua nilai, yakni: nilai maksimum dan nilai minimum. Hasil penelitian ditampilkan pada Tabel 1.

Tabel 1. Hasil akurasi pengenalan suara artikulasi /p/

Maksimal				Minimal		
Paku	Knn	3	74%	Knn	10	41%
	K-Fold	8		K-Fold	2	
Kapak	Knn	2	75%	Knn	10	42%
	K-Fold	9		K-Fold	2	
Atap	Knn	2	76%	Knn	10	41%
	K-Fold	7		K-Fold	2	
All P	Knn	5	69%	Knn	10	61%
	K-Fold	9		K-Fold	2	

Tabel 1 menunjukkan nilai akurasi maksimum di atas 69%. Nilai KNN dan K-Fold berbanding terbalik dengan hasil maksimum. Itu adalah nilai KNN kecil 2,3 dan 5. Sementara K-Fold berada pada nilai 8 dan 9.

Selain itu, tabel ini juga menunjukkan nilai akurasi minimum adalah 41%, dengan akurasi terendah untuk semua data kata adalah 61%. Nilai KNN dan K-Fold juga berbanding terbalik. Pada hasil minimum, nilai KNN berada pada Nilai 10. Namun, K-Fold berada pada nilai terendah, yaitu 2.

Hasilnya menunjukkan pengenalan suara untuk kata-kata bilabial dalam huruf konsonan /p/ berada pada nilai di atas 69%, dengan kata-kata itu sendiri berada di atas 74%. Dengan nilai KNN rendah tetapi *K-Fold* tinggi. Berbanding terbalik dengan akurasi minimum. Yang pada nilai *K-Fold* tinggi dengan nilai KNN rendah. Jika dibandingkan dengan penelitian lain seperti *Automation in Hypernasality* [11], penelitian pada paper ini ke depannya diharapkan dapat lebih baik dengan adanya penelitian lebih lanjut, dan dataset yang lebih luas. Pada penelitian di Indonesia dalam bidang yang sama [18] dapat dibandingkan bahwa penelitian ini memiliki lebih banyak kumpulan data dan hasil yang bervariasi untuk penelitian lebih lanjut.

4. Kesimpulan

Hasil penelitian dari 13 fitur MFCC pada 4 data, yaitu paku, kapak, atap, dan total /p/, memiliki tingkat akurasi yang berbeda. Penelitian ini menggunakan data pertama, oleh sebab itu penelitian ini masih bisa dikembangkan lebih lanjut. Tingkat akurasi terbaik adalah kata Atap, dengan akurasi hingga 76%. Akurasi terbaik adalah pada k sama dengan 2 dan k-lipat sama dengan 7. Penggunaan validasi silang dan KNN berbeda di setiap tes. Studi ini menemukan bahwa hasil ekstraksi fitur MFCC menggunakan klasifikasi KNN dengan validasi silang untuk CLP dan suara normal dalam masalah artikulasi ucapan untuk huruf-huruf bilabial, terutama huruf /p/ memiliki akurasi yang bervariasi. Akurasi di atas 70%. Dengan nilai terburuk pada 41% dan 42%, sedangkan untuk data keseluruhan, akurasi terburuk adalah 61%.

Penelitian ini menunjukkan bahwa nilai KNN 10 dengan validasi silang K-Fold 2 adalah akurasi minimal. Ini berbanding terbalik dengan akurasi maksimum, meskipun tidak pada angka yang sama. Namun, hal ini memiliki karakteristik nilai KNN kecil dengan nilai validasi silang *K-Fold* yang cukup besar.

Metode MFCC dengan klasifikasi menggunakan KNN menggunakan validasi silang *K-Fold* dapat mengklasifikasikan suara CLP dan suara normal menjadi lip /p/ huruf pada Paku, Kapak dan Atap.

4.1 Ucapan Terima Kasih

Ucapan terima kasih disampaikan kepada guru-guru saya, Bp. Samiadji dan Bp. Risanuri. Merupakan suatu kehormatan yang sangat besar mendapatkan bimbingan beliau. Terima kasih kepada Ibu Siti Agrippina Alodia Yusuf dan Ibu Rita serta para terapis wicara. dan teman-teman CLP serta sesama partisipan dalam penelitian ini.

Daftar Pustaka

- [1] Charisma, A. (2013). Sistem Verifikasi Penutur Menggunakan Metoda Mel Frequency Cepstral Coefficients-Vector Quantisation (Mfcc-Vq) Serta Sum Square Error (SSE) Dan Pengenalan Kata Menggunakan Metoda Logika Fuzzy. *Jurnal Teknik Elektro ITP*, 2(2), 71.
- [2] Budi, D., & Suthami, A. (2018). Sistem Pengenalan Dan Verifikasi Pembicara HMM. *Prosiding CITEE 2018*, 68-73.
- [3] Darwito, Haryadi Amran. (2012). *Identifikasi Level Keparahan Cleft Lip and Palate*, 82.
- [4] Dixit, A., Vidwans, A., & Sharma, P. (2016, March). Improved MFCC and LPC algorithm for bundelkhandi isolated digit speech recognition. In *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)* (pp. 3755-3759). IEEE.
- [5] Effendi, M. (2019). *Pengenalan Pengaruh Suara Konsonan Terhadap Vokal Menggunakan MFCC Dan SVM*. 03(02), 34–39.
- [6] Ferdiansyah, V., & Purwarianti, A. (2011, July). Indonesian automatic speech recognition system using English-based acoustic model. In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics* (pp. 1-4). IEEE.
- [7] He, L., Zhang, J., Liu, Q., Zhang, J., Yin, H., & Lech, M. (2018). Automatic detection of glottal stop in cleft palate speech. *Biomedical Signal Processing and Control*, 39, 230-236.
- [8] 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).
- [9] 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.
- [10] Larangeira, F. R., Dutka, J. D. C. R., Whitaker, M. E., de Souza, O. M. V., Lauris, J. R. P., da Silva, M. J. F., & Pegoraro-Krook, M. I. (2016). Speech nasality and nasometry in cleft lip and palate. *Brazilian journal of otorhinolaryngology*, 82(3), 326-333.
- [11] 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.
- [12] Rullo, R., Di Maggio, D., Festa, V. M., & Mazzarella, N. (2009). Speech assessment in cleft palate patients: a descriptive study. *International journal of pediatric otorhinolaryngology*, 73(5), 641-644.
- [13] Sardjono. (2005). *Terapi Wicara*. Jakarta. Departemen Pendidikan Nasional.
- [14] Supandi, A. (2014). Angka Kejadian Sumbing Bibir di RSUP Prof. Dr. R.D. Kandou Manado Periode 2011-2013. *e-CliniC*, 2(2).
- [15] Vijayalakshmi, P., Reddy, M. R., & O'Shaughnessy, D. (2007). Acoustic analysis and detection of hypernasality using a group delay function. *IEEE Transactions on biomedical engineering*, 54(4), 621-629.

- [16] Yessivirna, R., & Marji, D. E. R. (2013). Klasifikasi Suara Berdasarkan Gender (Jenis Kelamin) dengan Metode K-Nearest Neighbor (KNN). *Jurnal Doro*, 2(3).
- [17] Yusuf, S. A. A., & Hidayat, R. (2019, September). MFCC Feature Extraction and KNN Classification in ECG Signals. In *2019 6th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)* (pp. 1-5). IEEE.
- [18] Yuwono, W. (1987). *Klasifikasi Tingkat Keparahan Bibir Sumbing Pada Citra 2D*. 1–3.