Jurnal Linguistik Komputasional
Indonesian Association for Computational Linguistics

JLK
Vol 2, No 2
Sept 2019

Kalibrasi Rasio kemungkinan pada Sistem Rekognisi Pengucap Otomatis untuk Aplikasi Forensik di Indonesia

Miranti Indar Mandasari[1], Angga Dwi Firmanto[2], Fadjar Fathurrahman[3]
[1,2,3]*Program Studi Teknik Fisika, Fakultas Teknologi Industri, Institut Teknologi Bandung*
[1,2,3]*Jl. Ganesha no. 10, Bandung, 40132, Indonesia*
[1]mandasari@tf.itb.ac.id|miranti.indar.mandasari@gmail.com
[2]anggadwifirmanto@gmail.com, [3]fadjar@tf.itb.ac.id

Abstrak— **Kalibrasi LR merupakan tahapan yang sangat penting saat akan mengaplikasikan sistem rekognisi pengucap otomatis pada bidang forensik. Artikel ini memuat tahapan dan evaluasi terhadap sistem rekognisi pengucap yang dibangun menggunakan basis data suara ucap berbahasa Indonesia. Sistem dikembangkan menggunakan fitur MFCC, pemodelan GMM-UBM, dan normalisasi Z. Sistem dievaluasi kinerjanya berdasarkan gender laki-laki dan perempuan, serta dua skenario, yakni percakapan natural dan wawancara. Evaluasi sistem dilakukan menggunakan indikator performa dalam hal kemampuan diskriminasi dan kalibrasi sistem. Hasil evaluasi dengan berbagai indikator menunjukkan bahwa sistem rekognisi pengucap otomatis yang dibangun telah menunjukkan hasil yang sangat baik. Hal ini ditunjukkan dengan nilai EER terbaik sebesar 4.66%, dan nilai Cmc sebesar 0.04. Dengan begitu, sistem yang dikembangkan telah siap untuk dipakai sebagai alat analisis rekognisi pengucap otomatis untuk aplikasi forensik di Indonesia.**

Abstract— ***Likelihood ratio calibration is a very important step when applying an automatic speaker recognition system to forensic applications. This paper presents the process and evaluation of a speaker recognition system developed with spoken Indonesian database. The system is developed using MFCC feature, GMM-UBM modeling, and Z normalization. System performances were evaluated based on male and female genders, and two scenarios i.e., natural conversation and interview. System evaluations were done using performance measures for both discrimination and calibration abilities. Results show that based on various indicators, the system behaves very well. This is shown by the best achievable EER values of 4.66%, and Cmc values of 0.04. Therefore, the developed automatic speaker recognition system is now ready to be used for forensic applications in Indonesia.***

Kata kunci— **Kalibrasi, *likelihood ratio*, rekognisi pengucap, forensik, Indonesia.**
Keywords— ***Calibration, likelihood ratio, speaker recognition, forensic, Indonesia.***

I. PENDAHULUAN

Rekognisi pengucap, atau dalam istilah Bahasa Inggrisnya disebut sebagai *speaker recognition*, adalah suatu sistem yang berfungsi untuk mengenali identitas pengucap dari sebuah sampel rekaman suara ucap. Sistem rekognisi pengucap dapat dibangun dengan berbagai metode, diantaranya melalui pendekatan otomatis, semi-otomatis, dan manual [1]. Fitur yang digunakan pun bisa beragam, mulai dari fitur fonetik, akustik, sampai fitur statistik. Sistem rekognisi pengucap dapat digunakan dalam berbagai aplikasi. Sistem ini dapat digunakan sebagai alat autentikasi dalam skala komersil, misalkan untuk mengakses ponsel pintar. Selain aplikasi komersial, sistem rekognisi pengucap juga dapat digunakan untuk aplikasi forensik [2]. Pada aplikasi ini, sistem rekognisi pengucap dapat menghadapi berbagai tantangan, misalkan jumlah rekaman yang sedikit, perbedaan media perekaman, maupun tersangka yang tidak kooperatif dalam proses penyidikan [3,4].

Rekognisi pengucap merupakan salah satu sistem biometrik, yakni sistem yang mengenali identitas seseorang dari ciri-ciri biologisnya. Selain rekognisi pengucap, sistem pengenalan sidik jari, rekognisi wajah, dan pengenalan *Deoxyribonucleic acid* (DNA) merupakan contoh lain dari sistem biometrik [5]. Seperti penggunaan sistem biometrik lainnya pada aplikasi forensik, sangat penting untuk sistem rekognisi pengucap agar dapat memberikan keluaran berupa skor dalam bentuk rasio kemiripan, atau *likelihood ratio* (LR). Sehingga, sistem rekognisi pengucap perlu dilengkapi dengan kalibrasi agar dapat menghasilkan nilai skor LR yang dapat diandalkan dan terpercaya [6, 7].

Di Indonesia, rekognisi pengucap telah digunakan untuk aplikasi forensik sejak sekitar tahun 2007. Pada saat itu, telah disahkan undang-undang (UU) baru mengenai penggunakan rekaman suara ucap sebagai bukti yang sah di pengadilan. Selama ini, kebanyakan sistem rekognisi pengucap forensik yang dibangun dan digunakan di

Indonesia berbasiskan metode semi-otomatis dengan fitur fonetik-akustik [8]. Paradigma ini telah berubah dengan pengembangan sistem rekognisi pengucap otomatis dengan berbagai metode yang ada [9,10], mulai dari teknik *Gaussian Mixture Model – Universal Background Model* (GMM-UBM) [11], klasifikasi dengan *Support Vector Machine* (SVM) [12], pemodelan *i-vector* [13, 14], teknik perhitungan skor dengan *Probabilistic Linear Discriminant Analysis* (PLDA) [15, 16], sampai dengan teknik *deep-learning* yang berbasis jaringan syaraf tiruan [17, 18].

Artikel ini menampilkan tentang proses kalibrasi yang dilkukan pada pembuatan skor rasio kemungkinan di sistem rekognisi pengucap otomatis. Sistem rekognisi pengucap dibangun dengan menggunakan teknik GMM-UBM dan fitur suara berbasis *Mel-frequency cepstral coefficients* (MFCC). Kinerja dari sistem akan dievaluasi berdasarkan dua macam kinerja, yakni kinerja diskriminasi dan kinerja kalibrasi.

II. KALIBRASI *LIKELIHOOD RATIO*

Dalam penyajian bukti ahli, sistem rekognisi pengucap otomatis sebaiknya memberikan keluaran berupa skor rasio kemungkinan, atau *likelihood ratio* (LR). Sistem yang ideal harus dapat memberikan nilai LR terkalibrasi yang dapat diandalkan [19]. Nilai LR dapat dirumuskan sebagai:

$$LR = \frac{P(E|H_p, I)}{P(E|H_d, I)} \tag{1}$$

dimana E adalah bukti rekaman suara dari kasus forensik, H_p sebagai *prosecution hypothesis* atau hipotesa tuntutan, H_d sebagai *defence hypothesis* atau hipotesia pertahanan, dan I merepresentasikan informasi prior terkait dengan data forensik yang disajikan. Dalam kasus rekognisi pengucap, hipotesa tuntutan dirumuskan sebagai hipotesa dimana kedua sampel suara ucap yang dianalisis berasal dari satu orang pengucap yang sama. Sedangkan untuk hipotesa pertahanan, ia dirumuskan sebagai hipotesa dimana kedua sampel berasal dari pengucap atau orang yang berbeda.

Setelah seorang saksi ahli menghasilkan nilai LR, misalkan dari keluaran sistem rekognisi pengucap otomatis, nilai ini kemudian digunakan untuk mencari *posterior odds*. Dalam perhitungan *posterior odds* ini, pengadilan menambahkan juga nilai *prior odds*, atau kemungkinan dari informasi sebelumnya, dengan nilai LR. Proses perhitungannya mengikuti kaidah Bayes, yakni:

$$posterior\ odds = LR \times prior\ odds$$
$$\frac{P(H_p|E, I)}{P(H_d|E, I)} = \frac{P(E|H_p, I)}{P(E|H_d, I)} \times \frac{P(H_p|I)}{P(H_d|I)} \tag{2}$$

Berdasarkan persamaan di atas, nilai LR dapat dilihat sebagai ukuran nilai dari suatu bukti yang disajikan di pengadilan. Sangat penting untuk digarisbawahi bahwa, seorang saksi ahli hanya dapat mempresentasikan nilai LR di pengadilan. Proses pengambilan keputusan, disini perhitungan posterior odds, hanya dapat dilakukan oleh pengadilan atau hakim, dan bukan oleh saksi ahli [19, 20].

Nilai LR harus memiliki arti probabilitas. Cara untuk mengetahui apakah LR benar-benar memiliki arti probabilitas yang dapat diandalkan adalah dengan melalui proses kalibrasi. Kalibrasi, dalam hal ini kalibrasi LR, adalah proses untuk mentransformasi skor s dari keluaran sistem menjadi nilai LR yang terkalibrasi. Transformasi ini dapat dilakukan melalui

$$LR = w_0 + w_1 s \tag{3}$$

dimana w_0 dan w_1 merupakan parameter kalibrasi yang dioptimasi dengan teknik regresi logistik menggunakan sebuah set data latih. Parameter kalibrasi w_0 disebut sebagai parameter *offset*, dan w_1 sebagai parameter scaling atau *skala*.

Konsep kalibrasi pada bidang rekognisi pengucap dikhususkan pada konsep *proper scoring rule* atau aturan skor yang tepat [19, 21]. Sebuah sistem rekognisi pengucap otomatis harus dapat menghasilkan nilai LR yang dapat diandalkan agar dapat digunakan di pengadilan sebagai sebuah bukti ahli. Skor dari sistem rekognisi pengucap harus dikalibrasi agar dapat menghasilkan nilai LR yang handal ini.

III. EKSPERIMEN

Sebelum dilakukan proses kalibrasi, terlebih dahulu harus dibuat sebuah sistem rekognisi pengucap otomatis yang dibangun berdasarkan basis data yang cukup. Pada bagian ini akan dibahas tiga hal terkait eksperimen yang dilakukan, yakni basis data yang digunakan, konfigurasi sistem rekognisi pengucap, dan indikator kinerja untuk evaluasi performa sistem, baik secara diskriminatif maupun secara kalibratif.

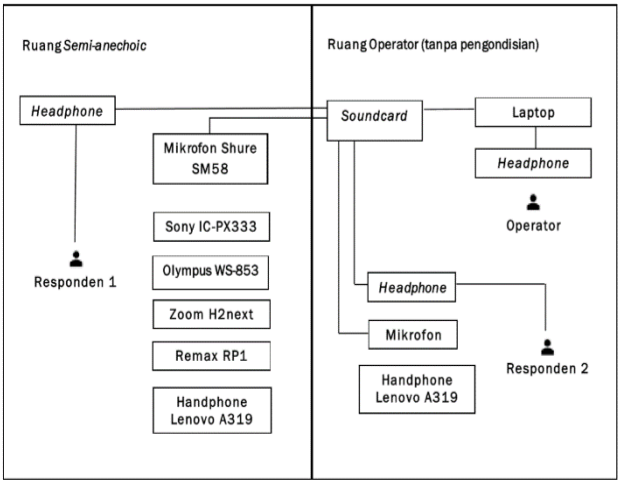
A. Basis data suara ucap

Dalam pengembangan sebuah sistem rekognisi pengucap otomatis, diperlukan suatu basis data yang dapat digunakan sebagai bahan pelatihan dan pengujian sistem tersebut. Dalam eksperimen ini digunakan sebuah basis data yang berisikan sejumlah rekaman suara ucap yang berbahasa Indonesia [22, 23]. Basis data ini dibangun dengan dua macam skenario, yakni percakapan natural dan wawancara. Skenario wawancara dipakai untuk melakukan mimik terhadap kasus forensik suara ucap yang banyak terjadi, dimana tersangka biasanya direkam suara ucapnya ketika proses investigasi kasus berlangsung. Percakapan natural banyak ditemukan pada kasus forensik di Indonesia, dimana para pelaku kejahatan berkomunikasi via telepon dengan rekan pelaku kejahatan. Percakapan telepon ini kemudian dipakai sebagai barang bukti di pengadilan.

Gambar 1 menunjukkan konfigurasi alat rekaman suara ucap yang dilakukan pada pembangunan basis data ini.

Perekaman dilakukan di ruang *semi-anechoic* agar dapat meminimalisir bising latar pada hasil rekaman. Media perekaman yang dipakai terdiri dari beberapa jenis mikrofon dan juga telepon seluler [24]. Subjek yang akan direkam suaranya berada di dalam ruang semi-anechoic, sedangkan operator berada di luar. Ketika melakukan rekaman dalam skenario percakapan natural, subjek akan melakukan percakapan dengan orang yang sudah dikenalnya. Sedangkan dalam skenario wawancara, subjek harus menjawab pertanyaan dari operator. Untuk eksperimen yang telah dilakukan,

Pada eksperimen yang dilakukan, diambil 90 orang subjek pengucap yang setengahnya bergender laki-laki, dan setengahnya lagi bergender perempuan. Skenario yang dipakai adalah kedua skenario, yakni percakapan natural dan wawancara. Adapun media perekaman yang dipakai adalah mikrofon Shure. Untuk training *universal background model* (UBM), digunakan teknik *leave one out*. Misalkan akan dilakukan pemodelan UBM untuk subjek pertama, maka ke empat puluh empat subjek lainnya digunakan untuk membangun model UBM untuk subjek pertama. Untuk proses kalibrasi, dilakukan metode self-calibration. Pada metode ini, parameter kalibrasi dilatih menggunakan data yang juga digunakan sebagai data uji. Pendekatan ini dilakukan untuk mengakomodir jumlah subjek yang minim pada basis data rekaman suara yang digunakan.

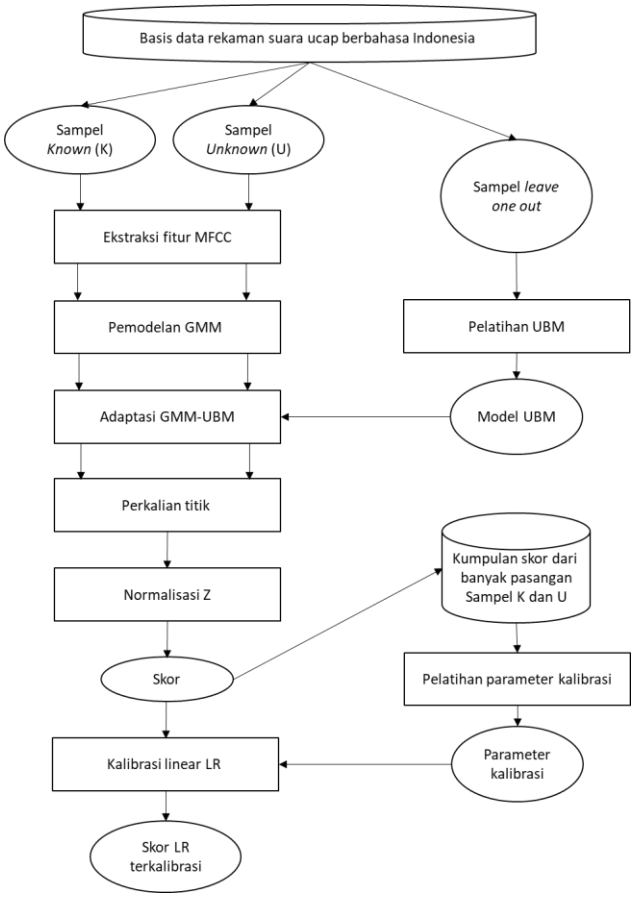


Gambar. 1 Konfigurasi perekaman data suara ucap.

B. Konfigurasi sistem rekognisi pengucap

Setelah dilakukan pengumpulan basis data rekaman suara ucap berbahasa Indonesia, basis data ini kemudian digunakan untuk melatih dan menguji sistem rekognisi pengucap. Gambar 2 menunjukkan diagram alir dari sistem rekognisi pengucap otomatis. Terdapat tiga tahapan penting pada sistem tersebut, yakni tahapan ekstraksi fitur MFCC [25], pemodelan GMM-UBM [11], dan kalibrasi linier LR [19, 21].

Dari basis data yang digunakan, dipilih sepasang sampel suara *known* (K) dan *unknown* (U) [8, 22]. Sampel suara K biasanya adalah suara yang diketahui identitas pengucapnya, sedangkan sampel suara U adalah suara yang tidak diketahui identitasnya. Pada kasus forensik secara umum, suara K didapat pada saat tersangka diwawancara oleh penyidik, sedangkan sampel U didapat dari suara rekaman hasil sadapan ketika kejahatan berlangsung. Sepasang sampel K dan U akan menghasilkan satu buah skor dan nilai LR. Sampel K dan U dari pengucap yang sama pada basis data akan membentuk kumpulan skor target s_t, sedangkan jika sampel K dan U dari pengucap yang berbeda, maka mereka akan membentuk kumpulan skor non-target s_n.



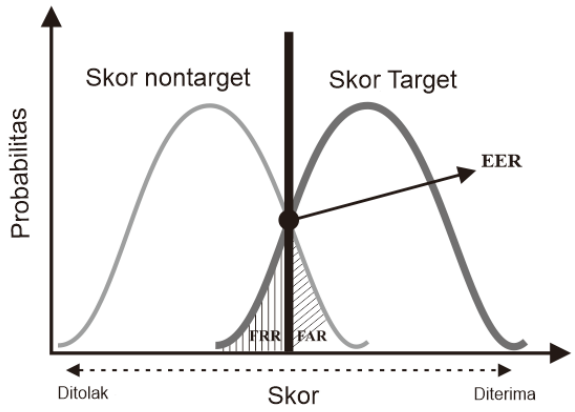
Gambar. 2 Diagram alir sistem rekognisi pengucap otomatis yang berbasiskan fitur MFCC, pemodelan GMM-UBM, dan kalibrasi LR.

Sampel suara diekstrak fiturnya dengan menggunakan filter Mel, sehingga fitur yang dihasilkan bernama fitur *Mel Frequency Cepstral Coefficient* (MFCC). Bersamaan dengan proses ekstraksi fitur, dilakukan juga tahap *Voice Activity Detection* (VAD), yakni tahapan pencarian bagian rekaman suara yang memiliki *active speech* atau suara ucap manusia aktif. Pada eksperimen, fitur yang dihasilkan adalah fitur 60 dimensi, dikalikan dengan jumlah *time frame* berdasarkan kerangka windows selebar 20 ms yang

saling bertumpuk setiap 10 ms. Keenampuluh dimensi ini terdiri dari 19 frekuensi Mel, ditambah 1 intensitas energi, kemudian 20 nilai *delta* dan 20 nilai *double delta*.

Setelah didapatkan fitur MFCC, selanjutnya fitur dimasukkan ke proses pemodelan *Gaussian Mixture Model* (GMM) dan adaptasi dengan *Universal Background Model* (UBM) menjadi model GMM-UBM. Jumlah Gaussian yang digunakan untuk sampel suara pria adalah 256, sedangkan untuk wanita adalah 128. Angka ini dipilih berdasarkan referensi yang dihasilkan dari eksperimen sebelumnya. Adaptasi model GMM dengan UBM dilakukan dengan metode adaptasi Maximum A-Priori (MAP). Model UBM dibangun dengan teknik *leave one out* karena keterbatasan jumlah data suara yang berada pada sistem basis data.

Untuk menghasilkan skor, dilakukan perkalian titik terhadap hasil vektor pemodelan GMM-UBM dari sampel suara K dan U. Penggunaan teknik perkalian titik ini disebut juga sebagai teknik *dot-scoring* di bidang rekognisi pengucap. setelah melewati tahap perkalian titik, skor-skor tersebut kemudian dinormalisasi dengan teknik normalisasi Z, atau dikenal sebagai teknik *zero* atau *Z-normalization*. Skor-skor ternormalisasi Z kemudian dikumpulkan di dalam kumpulan skor mentah dan dikelompokkan menjadi skor target dan non-target. Kumpulan skor ini kemudian digunakan untuk melakukan pelatihan terhadap parameter-parameter kalibrasi linier LR, w_0 dan w_1, dengan menggunakan teknik regresi logistik atau *logistic regression*. Setelah melalui proses kalibrasi linier, skor mentah akan bertransformasi menjadi skor LR terkalibrasi dengan arti LR yang baik [19].



Gambar. 3 Bentuk distribusi skor *target* dan *non-target* hasil pengolahan suara K dan U dari sistem rekognisi pengucap otomatis.

C. Indikator Kinerja Sistem Rekognisi

Indikator kinerja sistem rekognisi pengucap dapat dilihat dari bentuk distribusi skor target saat pengucap K dan U sama, dan distribusi skor non-target saat pengucap K dan U berbeda. Gambar 3 menunjukkan distribusi tersebut. Terdapat satu ukuran performa sistem rekognisi pengucap yang banyak digunakan, yakni nilai *Equal Error Rate* (EER). Sistem rekognisi pengucap merupakan sistem *binary-classifier* atau pemisah kelas biner, dimana hanya terdapat dua macam keluaran atau kelas. Kelas tersebut adalah kelas skor target dan non-target. Seperti pada sistem pemisah kelas biner lainnya, terdapat dua macam kesalahan yang mungkin, yakni kesalahan *False Rejection Rate* (FRR) dan *False Alarm Rate* (FAR). FRR menunjukkan probabilitas kesalahan klasifikasi skor target menjadi non-target, sedangkan FAR menunjukkan probabilitas kesalahan klasifikasi skor non-target menjadi target. Nilai EER didapatkan ketika nilai FRR dan FAR sama.

Nilai EER merupakan indikator kinerja sistem rekognisi pengucap otomatis yang umum digunakan. Nilai EER menunjukkan performa diskriminasi antara skor target dan non-target. Nilai EER yang semakin kecil menunjukkan semakin baiknya performa diskriminasi suatu sistem. Selain dari segi diskriminasi kedua jenis skor, performa sistem juga dapat dianalisis melalui seberapa baik kemampuan kalibrasi sistemnya. Pada artikel ini, akan digunakan indikator C_{llr}, yakni *cost of log likelihood ratio calibration*. Nilai C_{llr} ini dapat dirumuskan dengan

$$C_{llr} = \frac{1}{N_{H_p}} \sum_{i=1}^{N_{H_p}} \log_2 \left(1 + \frac{1}{LR_i}\right) + \frac{1}{N_{H_d}} \sum_{j=1}^{N_{H_d}} \log_2 (1 + LR_j) \qquad (4)$$

dimana N_{H_p} adalah jumlah skor target yang memenuhi hipotesa tuntutan, dan N_{H_d} adalah jumlah skor non-target yang memenuhi hipotesa pertahanan. Nilai LR pada persamaan tersebut adalah nilai skor rasio kemungkinan yang telah dikalibrasi. Nilai C_{llr} menunjukan performa sistem terhadap semua kemungkinan informasi *prior* yang mungkin terjadi. [6, 19, 21]

Dengan menggunakan algoritma *Pool Adjacent Violators* (PAV), nilai optimal C_{llr} dapat dicari. Nilai ini disebut dengan nilai C_{llr}^{min} yakni nilai minimum C_{llr} yang dapat diraih oleh sistem. Sebuah indikator kinerja kalibrasi sistem *mis-calibration cost*, atau C_{mc}, adalah

$$C_{mc} = C_{llr} - C_{llr}^{min} \qquad (5)$$

Nilai C_{mc} mengindikasikan calibration loss, atau kerugian kalibrasi dari sistem rekognisi pengucap otomatis.

Nilai C_{llr}^{min} merupakan indikator kinerja sistem yang menunjukkan performa diskriminasi sistem dalam semua kemungkinan prior. Dua nilai lainnya, C_{llr} dan C_{mc} merupakan ukuran performa kalibrasi suatu sistem rekognisi pengucap. nilai C_{llr} dapat dilihat sebagai rangkuman skalar dari seberapa baiknya nilai log LR yang dihasikan oleh sistem rekognisi pengucap otomatis. [19]

IV. HASIL DAN DISKUSI

Berdasarkan hasil eksperimen, kinerja dari sistem rekognisi pengucap otomatis yang telah dibangun dapat dianalisis. Tabel I menunjukkan performa diskriminatif sistem rekognisi pengucap menggunakan ukuran nilai EER dalam satuan persen. Secara umum, nilai EER untuk subjek laki-laki lebih baik daripada perempuan. Ini merupakan hasil klasik yang bisa dihasilkan oleh suatu sistem rekognisi pengucap, dimana suara perempuan lebih sulit untuk didiskriminasi dibandingkan suara laki-laki. Nilai EER terbaik adalah untuk subjek laki-laki dengan skenario wawancara, dengan nilai EER 4.66%.

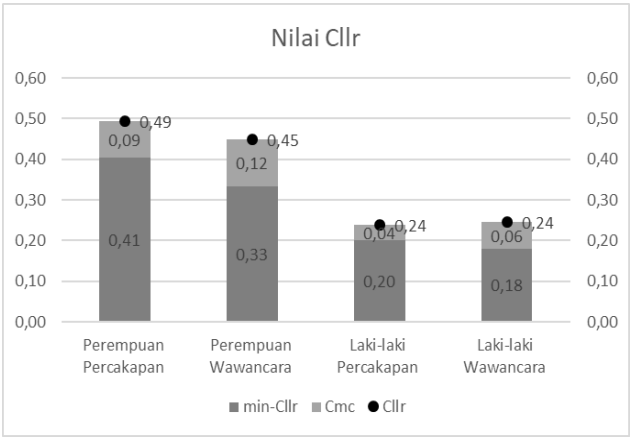
Terdapat dua skenario wicara yang digunakan pada eksperimen, yakni skenario percakapan natural dan wawancara. Berdasarkan Tabel I, nilai EER pada skenario wawancara lebih kecil daripada nilai pada skenario percakapan. Ini berarti bahwa rekognisi pengucap lebih mudah dilakukan pada rekaman suara ucap dengan skenario wawancara. Hal ini dikarenakan, pada skenario percakapan banyak terdapat variabilitas pada suara pengucap. Pada umumnya, ketika seseorang bercakap dengan skenario percakapan natural, ia akan lebih ekspresif dalam berucap yang menyebabkan munculnya banyak variabilitas seperti nada suara, intonasi, dan pengucapan kata-kata yang non-formal. Hal ini akan menyebabkan munculnya variabilitas yang lebih besar pada fitur yang diekstrak.

Tabel II menunjukkan nilai C_{llr}^{min}, C_{mc}, dan C_{llr} dari sistem rekognisi pengucap otomatis yang dievaluasi. Nilai-nilai pada tabel kemudian digambarkan pada grafik di Gambar 3. Pada gambar tersebut dapat dilihat bahwa nilai C_{llr} merupakan penjumlahan dari nilai C_{llr}^{min} dan nilai C_{mc}. Nilai C_{llr}^{min} merupakan indikator performa diskriminasi sistem, sedangkan nilai C_{mc} dan C_{llr}merupakan indikator performa kalibrasi sistem.

Tren yang dilihat pada nilai C_{llr}^{min} sama dengan tren yang dilihat pada nilai EER di Tabel I. Disini, nilai C_{llr}^{min} didapatkan untuk subjek laki-laki dengan skenario wawancara, dengan nilai 0.18. Perlu digaris bawahi bahwa nilai C_{llr}^{min}, C_{mc}, dan C_{llr} harus berada pada angka di bawah 1.00. Jika nilai-nilai tersebut lebih besar dari 1.00, maka dapat dikatakan bahwa sistem menghasilkan skor LR yang tidak memiliki arti bermakna.

TABEL I
PERFORMA DISKRIMINASI SISTEM BERDASARKAN NILAI EER

Gender	EER (%)		Jumlah Gaussian
	Percakapan	Wawancara	
Perempuan	10.58	9.59	128
Laki-laki	6.08	4.66	256



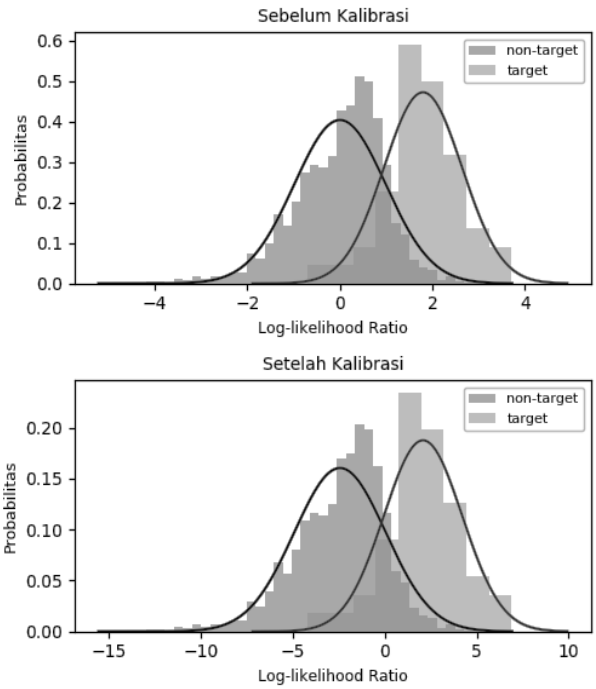
Gambar. 3 Kinerja sistem rekognisi pengucap dari ukuran nilai C_{llr}^{min} sebagai indikator performa diskriminasi, serta nilai C_{llr} dan C_{mc} sebagai indikator performa kalibrasi.

Performa kalibrasi dari sistem rekognisi pengucap sudah berada pada level yang baik. Hal ini terlihat dari nilai-nilai C_{mc} dan C_{llr} yang rendah dan lebih kecil dari 1.00. Nilai C_{mc} adalah ukuran performa kalibrasi sistem yang menunjukkan tingkat kesalahan kalibrasi nilai LR. Baik untuk subjek laki-laki maupun perempuan, nilai C_{mc} sangat rendah, yakni berada di rentang nilai 0.04 sampai dengan 0.12. Nilai-nilai C_{mc} dan C_{llr} yang rendah pada sistem rekognisi pengucap, menunjukkan kemampuan sistem dalam menghasilkan skor LR yang terkalibrasi dengan baik. Sistem yang seperti ini, dianggap siap untuk digunakan dalam aplikasi forensik [19].

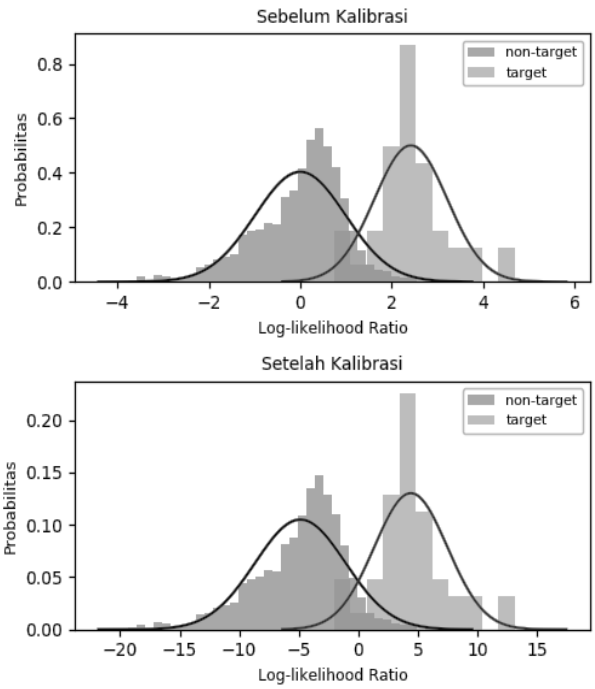
Selain dengan menggunakan ukuran kinerja sistem dengan nilai EER, C_{llr}^{min}, C_{mc}, dan C_{llr}, perform sistem rekognisi pengucap otomatis juga dapat dilihat dari bentuk distribusi skor nya. Gambar 4 menunjukkan distribusi skor target dan non-target hasil keluaran sistem rekognisi pengucap. Gambar distribusi skor ditampilkan untuk (a) subjek perempuan pada skenario percakapan, (b) subjek perempuan pada skenario wawancara, (c) subjek laki-laki pada skenario percakapan, dan (d) subjek laki-laki pada skenario wawancara. Distribusi skor ditunjukkan untuk kondisi skor pada saat sebelum dan sesudah kalibrasi linier LR dilakukan terhadap skor mentah yang dihasilkan oleh sistem rekognisi pengucap.

TABEL II
PERFORMA DISKRIMINASI SISTEM BERDASARKAN NILAI C_{llr}^{min}, SERTA PERFORMA KALIBRASI SISTEM BERDASARKAN NILAI C_{mc} DAN C_{llr}

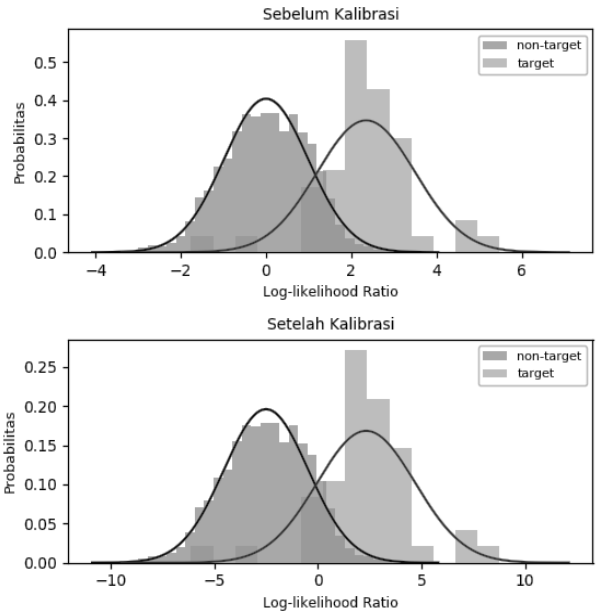
Gender	Skenario		Ukuran Kinerja
	Percakapan	Wawancara	
Perempuan	0.41	0.33	C_{llr}^{min}
Laki-laki	0.20	0.18	
Perempuan	0.09	0.12	C_{mc}
Laki-laki	0.04	0.06	
Perempuan	0.49	0.45	C_{llr}
Laki-laki	0.24	0.24	



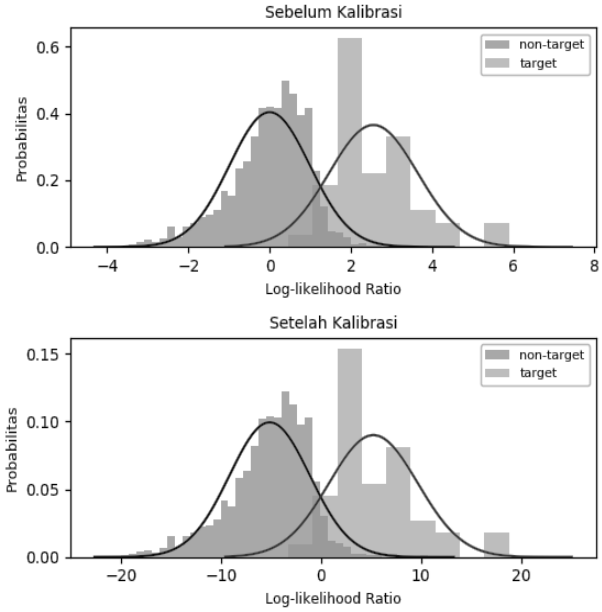
(a) Perempuan - Percakapan (EER = 10.58%)



(c) Laki-laki - Percakapan (EER = 6.08%)



(b) Perempuan - Wawancara (EER = 9.59%)



(d) Laki-laki - Wawancara (EER = 4.66%)

Gambar. 4 Distribusi skor target dan non-target saat sebelum dan setelah dilakukan proses kalibrasi LR pada subjek pengucap perempuan dan laki-laki, serta pada skenario percakapan dan wawancara.

Nilai EER merupakan ukuran performa diskriminasi sistem rekognisi pengucap. Dari grafik-grafik distribusi skor pada Gambar 4, dapat dilihat bahwa kondisi eksperimen dengan EER rendah menunjukkan semakin jauhnya perpisahan antara distribusi skor target dan non-target. Namun, perlu juga untuk diperhatikan bahwa bentuk distribusi skor tidak berubah saat sebelum dan setelah kalibrasi dilakukan. Ini dikarenakan kalibasi linear tidak merubah performa diskriminasi sistem, melainkan ia merubah performa kalibrasi dari sistem.

Setelah kalibrasi dilakukan, dapat dilihat bahwa skor akan bergeser dengan batas tengahnya semakin mendekati nilai nol. Nilai *log-likelihood ratio* (log LR) yang ditunjukkan pada Gambar akan bernilai nol jika probabilitas bukti rekaman suara E saat mendukung hipotesa tuntutan H_p adalah sama dengan saat ia mendukung hipotesa pertahanan H_d, seperti diperlihatkan pada persamaan (1). Sebelum kalibrasi, nilai nol biasanya tidak berada pada tempatnya. Setelah kalibrasi dilakukan, yakni proses merenggangkan dan menggeser skor, nilai log LR, atau LR yang dihasilkan akan lebih handal dan terpercaya. Artinya, sistem menghasilkan nilai LR yang dapat diterjemahkan menjadi pengertian awalnya seperti ditampilkan pada persamaan (1), yakni perbandingan antara probabilitas bukti E dalam konteks hipotesa tuntutan H_p dan probabilitas bukti E dalam konteks hipotesa hipotesa pertahanan H_d. Kesimpulannya, proses kalibrasi LR dapat dikatakan telah berhasil dilakukan untuk memperbaiki performa sistem rekognisi pengucap otomatis, sehingga ia dapat digunakan sebagai alat untuk mempersiapkan bukti di pengadilan dalam kasus-kasus forensik.

V. KESIMPULAN

Artikel ini memuat hasil eksperimen kalibrasi LR pada sistem rekognisi pengucap otomatis yang berbasiskan fitur MFCC, pemodelan GMM-UBM, dan normalisasi Z. Kalibrasi yang dilakukan menggunakan metode kalibrasi linier. Meskipun dengan keterbatasan jumlah data latih yang tersedia pada basis data, sistem telah berhasil mencapai nilai EER 4.66% untuk subjek laki-laki pada skenario wawancara. Ini menunjukkan performa diskriminasi skor target dan non-target sistem yang sangat baik. Selain itu, performa kalibrasi sistem juga sangat baik karena nilai C_{llr} yang berhasil mencapai nilai di bawah 1.00 untuk seluruh gender dan skenario. Setelah kalibrasi dilakukan, sistem rekognisi pengucap dapat menghasilkan skor LR yang terkalibrasi. Basis data yang digunakan dalam pembangunan sistem rekognisi pengucap adalah basis data suara ucap berbahasa Indonesia. Sehingga, skor LR terkalibrasi hasil keluaran sistem rekognisi pengucap otomatis ini kemudian dapat digunakan sebagai bukti pada kasus forensik, khususnya di Indonesia.

REFERENSI

[1]. Kinnunen, T. and Li, H., 2010. An overview of text-independent speaker recognition: From features to supervectors. Speech communication, 52(1), pp.12-40.

[2]. Beigi, H., 2011. Speaker recognition. Springer US.

[3]. Campbell, J. P., Shen, W., Campbell, W. M., Schwartz, R., Bonastre, J. F. and Matrouf, D., 2009. Forensic speaker recognition. Institute of Electrical and Electronics Engineers.

[4]. Neustein, A. and Patil, H. A., 2010. Forensic speaker recognition, Vol. 1., Springer.

[5]. Meuwly, D. and Veldhuis, R., 2012. Forensic biometrics: From two communities to one discipline. 2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), IEEE.

[6]. van Leeuwen, D. A. and Brümmer, N., 2013. The distribution of calibrated likelihood-ratios in speaker recognition. arXiv preprint arXiv:1304.1199.

[7]. Mandasari, M. I., McLaren, M. L. and van Leeuwen, D. A., 2011. Evaluation of i-vector speaker recognition systems for forensic application. Interspeech conference, ISCA, Florence, Italy.

[8]. Sarwono, J., Mandasari, M. I., and Suprijanto, 2010. Forensic speaker identification: an experience in Indonesians court, Proceedings of 20th International Congress on Acoustics, Sydney, Australia.

[9]. Stefanus, I., Sarwono, R. J. and Mandasari, M. I., 2017. GMM based automatic speaker verification system development for forensics in Bahasa Indonesia. 2017 5th International Conference on Instrumentation, Control, and Automation (ICA), pp. 56-61, IEEE.

[10]. Firmanto, A. D., Mandasari, M. I., Suprijanto, and Fathurrahman, F., 2019. Applying GMM-UBM framework for Indonesian forensic speaker verification. AIP Conference Proceedings, Vol. 2088, No. 1, p. 050013, AIP Publishing.

[11]. Matějka, P., et al., 2011. Full-covariance UBM and heavy-tailed PLDA in i-vector speaker verification. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE.

[12]. Greenberg, C. S., Bansé, D., Doddington, G. R., Garcia-Romero, D., Godfrey, J. J., Kinnunen, T., & Reynolds, D. A., 2014, June. The NIST 2014 speaker recognition i-vector machine learning challenge. In Odyssey: The Speaker and Language Recognition Workshop, pp. 224-230.

[13]. Garcia-Romero, D. and McCree, A., 2014, May. Supervised domain adaptation for i-vector based speaker recognition. 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4047-4051, IEEE.

[14]. Dehak, N., Kenny, P. J., Dehak, R., Dumouchel, P., & Ouellet, P., 2010. Front-end factor analysis for speaker verification. IEEE Transactions on Audio, Speech, and Language Processing, 19(4), 788-798.

[15]. Matějka, P., Glembek, O., Castaldo, F., Alam, M. J., Plchot, O., Kenny, P., & Černocky, J., 2011. Full-covariance UBM and heavy-tailed PLDA in i-v,ector speaker verification. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4828-4831.

[16]. Kanagasundaram, A., Vogt, R. J., Dean, D. B., & Sridharan, S., 2012. PLDA based speaker recognition on short utterances. Speaker and Language Recognition Workshop (Odyssey 2012), ISCA.

[17]. Vasilakakis, V., Cumani, S., Laface, P., & Torino, P., 2013. Speaker recognition by means of deep belief networks. Proceedings of Biometric Technologies in Forensic Science, pp. 52-57.

[18]. Liu, Y., Qian, Y., Chen, N., Fu, T., Zhang, Y., & Yu, K., 2015. Deep feature for text-dependent speaker verification. Speech Communication, vol. 73, pp. 1-13.

[19]. Mandasari, M. I., 2018. Speaker Recognition System in Forensic Conditions: The Calibration and Evaluation of the Likelihood Ratio. Doctoral dissertation, Radboud University Nijmegen, the Netherlands.

[20]. Drygajlo, A., & Haraksim, R., 2017. Biometric Evidence in Forensic Automatic Speaker Recognition. Handbook of Biometrics for Forensic Science, pp. 221-239, Springer, Cham.
[21]. Mandasari, M. I., Saeidi, R., McLaren, M., and van Leeuwen, D. A., 2013. Quality measure functions for calibration of speaker recognition systems in various duration conditions. IEEE Transactions on Audio, Speech, and Language Processing, 21(11), pp.2425-2438.
[22]. Sarwono, R. S. J., Mandasari, M. I., and Stefanus, I., 2016. Bahasa Speech Database for Automatic Speaker Recognition System Development in Indonesia. Reports of research assisted by the Asahi Glass Foundation, pp.1-10.
[23]. Firmanto, A. D., Stefanus, I., Ikhwanudin, R., Mandasari, M. I., 2017. Desain Perekaman Basis Data Suara Ucap untuk Pengembangan Sistem Rekognisi Pengucap Otomatis Forensik Berbahasa Indonesia, Seminar Instrumentasi dan Kontrol, Yogyakarta.
[24]. Mandasari, M. I., Sudarsono, A. S., Sarwono, R. S. J., and Firmanto, A. D., 2018, July. The effect of recording devices towards MFCC based speech features in a typical forensic scenario found in Indonesia. Proceedings of 25th International Congress on Sound and Vibration (ICSV25), Hiroshima, Japan.
[25]. Tiwari, V., 2010. MFCC and its applications in speaker recognition. International journal on emerging technologies, 1(1), pp. 19-22.