

Konversi Suara Ucapan Bahasa Indonesia ke Sistem Bahasa Isyarat Indonesia (SIBI)

A. Nasri*¹

¹Program Studi Teknik Informatika STMIK Handayani Makassar
e-mail: andinasrimansur@gmail.com*

Abstract

With the development of speech recognition technology, a variety of software that aims to facilitate deaf people to communicate with each other has been developed. The system translates the sound of speech into sign language, or otherwise translate sign language into speech sounds. The system has been developed in a variety of languages such as English, Arabic, Spanish, Mexico, Indonesian and others. Especially for Indonesian has also been trying to do some research to make the system as such. But the system created is still limited to the Automatic Speech Recognition (ASR) that is used in which possessed a limited vocabulary. This research aims to develop a system to translate Indonesian speech to Indonesian sign language System (SIBI) with a larger corpus of data and use continue speech recognition to improve the accuracy of speech recognition system. The result test show the system obtained an average 90.50% accuracy and 9.50% WER. The accuracy was higher than the second study 48.75% and 66.67% for first study. The system can recognize spoken words continuously or sentence pronunciation. The results of performance testing system reached 0.83 seconds for speech to text and 8.25 seconds for speech to sign.

Keyword: Automatic Speech Recognition; Sign Language; Accuracy; WER

Abstrak

Dengan semakin berkembangnya teknologi *speech recognition*, berbagai *software* yang bertujuan untuk memudahkan orang tunarungu dalam berkomunikasi dengan yang lainnya telah dikembangkan. Sistem tersebut menterjemahkan suara ucapan menjadi bahasa isyarat atau sebaliknya bahasa isyarat diterjemahkan ke suara ucapan. Sistem tersebut sudah dikembangkan dalam berbagai bahasa seperti bahasa Inggris, Arab, Spanyol, Meksiko, Indonesia dan lain-lain. Khusus untuk bahasa Indonesia mulai juga sudah yang mencoba melakukan penelitian untuk membuat sistem seperti tersebut. Namun sistem yang dibuat masih terbatas pada *Automatic Speech Recognition* (ASR) yang digunakan dimana mempunyai kosa-kata yang terbatas. Dalam penelitian ini bertujuan untuk mengembangkan sistem penterjemah suara ucapan bahasa Indonesia ke Sistem Bahasa Isyarat Indonesia (SIBI) dengan data korpus yang lebih besar dan menggunakan *continue speech recognition* untuk meningkatkan akurasi sistem. Dari hasil pengujian sistem menunjukkan diperoleh hasil akurasi sebesar rata-rata 90,50 % dan *Word Error Rate* (WER) 9,50%. Hasil akurasi lebih tinggi dibandingkan penelitian kedua 48,75% dan penelitian pertama 66,67%. Disamping itu sistem juga dapat mengenali kata yang diucapkan secara kontinyu atau pengucapan kalimat. Kemudian hasil pengujian kinerja sistem mencapai 0,83 detik untuk *speech to text* dan 8,25 detik untuk *speech to sign*.

Kata kunci: Automatic Speech Recognition; Bahasa Isyarat; Akurasi; WER

1. Pendahuluan

Anak tunarungu adalah anak yang mengalami gangguan pada organ pendengarannya sehingga mengakibatkan ketidakmampuan mendengar, mulai dari tingkatan yang ringan sampai yang berat sekali yang diklasifikasikan kedalam tuli (*deaf*) dan kurang dengar (*hard of hearing*) [1]. Ketunarunguan bukan hanya mengakibatkan tidak berkembangnya kemampuan berbicara, lebih dari itu dampak paling besar adalah terbatasnya kemampuan berbahasa. Hal ini disebabkan karena mereka mengalami kesulitan untuk menguasai bahasa, miskin kosakata, sulit mengartikan kosakata, sulit dalam memahami arti sebuah konsep abstrak secara utuh dan akurat seperti konsep-konsep indah, bahagia, jujur, adil, dan sebagainya. Terlebih dalam mengartikan

kata-kata yang mengandung arti kiasan seperti "keras kepala". Sehingga anak tunarungu sering mengalami kesulitan dalam memahami sebuah kalimat, baik dalam bentuk suara ucapan ataupun tulisan. Sebaliknya mereka akan lebih mengerti dengan kata-kata dasar dan kalimat yang lebih sederhana dan disajikan secara pelan serta disertai dengan gambar sebagai representasi dari arti dan makna kata tersebut [2].

Pada umumnya tingkat intelegensi anak tunarungu secara potensial sama dengan anak normal lainnya, akan tetapi secara fungsional perkembangannya dipengaruhi oleh tingkat kemampuan berbahasa. Dengan keterbatasan informasi dan kurangnya daya abstraksi anak mengakibatkan proses pencapaian pengetahuan yang lebih luas menjadi terhambat. Padahal anak tunarungu juga berhak mendapatkan pendidikan yang layak seperti anak lainnya. Sebagaimana disebutkan dalam UU No. 20 Tahun 2003 tentang Sistem Pendidikan Nasional Pasal 32 bahwa: "Pendidikan Khusus (Pendidikan Luar Biasa) merupakan pendidikan bagi peserta didik yang memiliki tingkat kesulitan dalam mengikuti proses pembelajaran karena kelainan fisik, emosional, mental, sosial" [3]. Sedini mungkin seharusnya mereka memperoleh pendidikan dan pengajaran yang layak serta informasi yang lebih luas. Pihak yang diharapkan berperan utama dalam hal ini adalah dari keluarga, sekolah, maupun dari orang-orang disekitarnya [4]. Akan tetapi yang menjadi kendala adalah sebagian besar dari mereka tidak tahu bahasa isyarat untuk berkomunikasi.

Bahasa isyarat yang ada di Indonesia ada dua yaitu Sistem Isyarat Bahasa Indonesia (SIBI) dan Bahasa Isyarat Indonesia (BISINDO). Pada tahun 1993, Departemen Pendidikan dan Kebudayaan Dasar hal ini Direktorat Dasar, Direktorat Jenderal Pendidikan Dasar dan Menengah, mengambil kebijakan untuk dibakukan Kamus Sistem Isyarat Bahasa Indonesia sebagai isyarat nasional [5].

Dengan semakin berkembangnya teknologi *speech recognition*, berbagai *software* yang bertujuan untuk memudahkan orang tunarungu dalam berkomunikasi dengan yang lainnya telah dikembangkan. Sistem tersebut menterjemahkan suara ucapan menjadi bahasa isyarat atau sebaliknya bahasa isyarat diterjemahkan ke suara ucapan. Sistem tersebut sudah dikembangkan dalam berbagai bahasa seperti bahasa Inggris, Arab, Spanyol, Meksiko, Indonesia dan lain-lain. Khusus untuk bahasa Indonesia mulai diteliti oleh Septiarti dan Haryanto dengan judul "Konverter Suara dengan Input Bahasa Indonesia ke Video, Gerakan Bahasa Isyarat Dengan Metode *Speech Recognition (Hidden Markov Model)* Untuk Penderita Tunarungu" [6]. Penelitian ini menggunakan SIBI sebagai isyarat *output* dengan hasil penelitian menunjukkan akurasi *speech to text* 66.67 %. Kemudian pada penelitian dilakukan oleh Noora dkk, dengan judul "*Speech to Video (S2V) Sistem Komunikasi Portabel untuk Penderita Tunarungu*" [7]. Penelitian ini menggunakan metode yang sama yaitu HMM dengan data *training* yang lebih baik dari sebelumnya. Hasil pengujian menunjukkan bahwa sistem dapat mengenali kata yaitu sekitar 89.84%, apabila data uji yang digunakan sama dengan data referensi. Sistem ini bisa mengenali kata sekitar 48.75% apabila data uji yang digunakan berbeda dengan data referensi.

Berdasarkan uraian di atas maka tujuan dari penelitian ini adalah untuk mengembangkan sistem penterjemah suara ucapan bahasa Indonesia ke bahasa isyarat Indonesia dengan menggunakan Sistem Isyarat Bahasa Indonesia (SIBI). Kemudian untuk memperjelas arti dan makna kata atau kalimat akan ditampilkan gambar yang berkaitan dengan arti kata tersebut [8].

2. Metode Penelitian

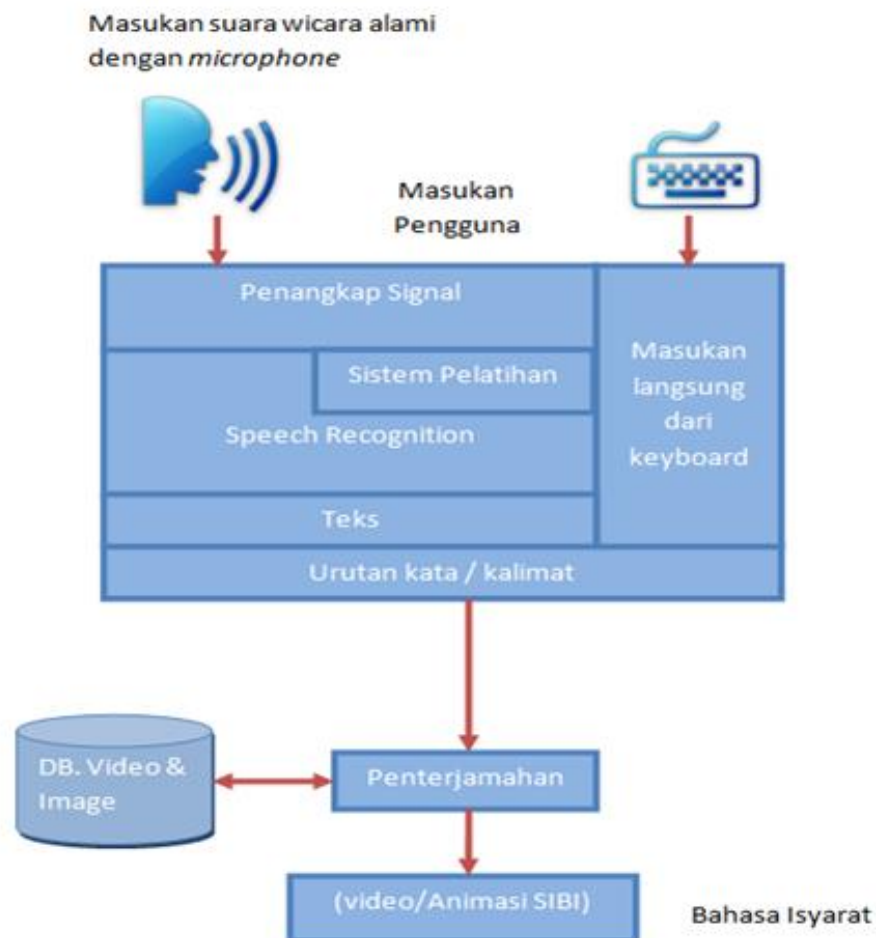
Pengambilan data berupa suara yang direkam dari beberapa responden laki-laki dan wanita dewasa dan anak-anak untuk dijadikan *sample* data suara latih dan data suara uji dengan mengucapkan kata dalam Bahasa Indonesia yang umum digunakan sehari-hari dan disimpan dalam bentuk file format wav. Proses perekaman dilakukan pada kondisi tenang untuk menghindari *noise* yang masuk. Kata bahasa Indonesia yang digunakan adalah suku kata bahasa lisan, bukan suku kata bahasa tulis. Hal ini perlu diperhatikan karena suku kata bahasa lisan dan suku kata bahasa tulis sangat berbeda. Sebagai contoh kata "belajar" memiliki suku kata bahasa lisan (be-la-jar), sedangkan suku kata bahasa tulisan-nya (bel-a-jar). Penggunaan suku kata bahasa lisan dapat membantu pembentukan data *Acoustic Model* dan *Language Model* serta mengurangi jumlah data yang harus dikenali tanpa mengurangi jumlah kata Bahasa Indonesia yang ada.

Dalam prosesnya, penelitian kami telah mengalami beberapa tahap. Peneliti telah melakukan tinjauan literatur yang berhubungan dengan objek yang diteliti. Beberapa karya ilmiah telah dipelajari dan dianalisis apakah dapat digunakan secara langsung atau tidak langsung untuk

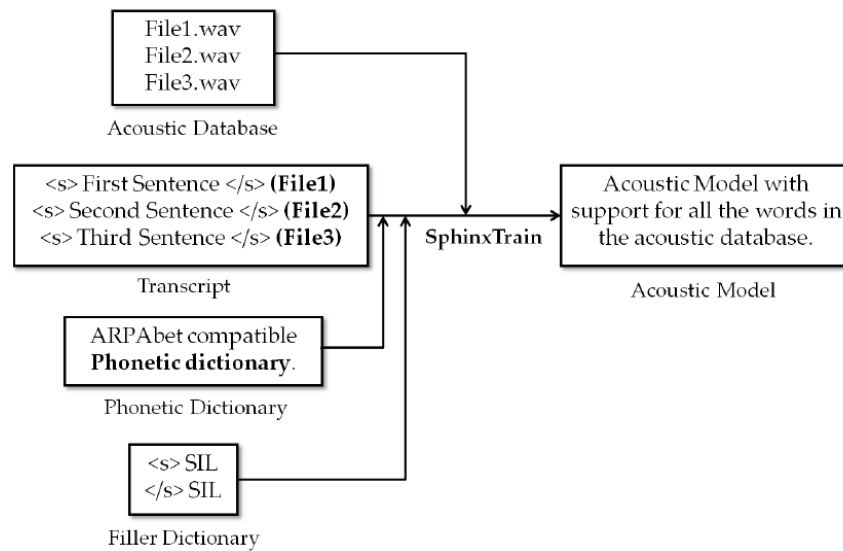
memberikan solusi dari masalah yang ada dengan mengekstrak masalah yang relevan dan hasil yang didapatkan dari penelitian yang sebelumnya. Kemudian literatur tentang perkembangan sistem penterjemah bahasa dan pengolahan bahasa alami yang dibaca dan dianalisis secara rinci untuk digunakan dalam mencari solusi dari masalah yang ada.

Studi empiris dilakukan pada lembaga tunarungu di Makassar. Kami mengunjungi lembaga untuk mengumpulkan informasi tentang bahasa isyarat yang digunakan saat ini. Wawancara dan pertemuan dengan instruktur telah dilakukan untuk mendapatkan pengetahuan yang lebih baik tentang Bahasa Isyarat Indonesia. Penelitian dilakukan untuk mengetahui bagaimana Bahasa Isyarat diajarkan dan digunakan saat ini di Indonesia dan mengamati langsung kehidupan sehari-hari dari penyandang tunarungu sehingga dapat lebih memahami tantangan yang dihadapi oleh mereka.

Kemudian tahap selanjutnya melakukan perancangan sistem. Sistem yang dirancang akan membentuk suatu sistem penterjemah suara ucapan Indonesia ke bahasa isyarat Indonesia. Sistem penterjemah dilakukan dengan mengubah suara menjadi teks dan kemudian dari teks menjadi bahasa isyarat. *Speech Recognition Engine* digunakan untuk melakukan konversi suara ke teks dan selanjutnya hasil konversi yang berupa teks digunakan untuk pencarian bahasa isyarat yang sesuai. Sistem yang akan dirancang dapat dilihat seperti yang ditunjukkan pada gambar 1. Setelah itu dilakukan penulisan *skrip* program atau *coding* dengan menggunakan bahasa pemrograman java dan Sql untuk database MySQL. Pengujian akurasi sistem dilakukan dengan menggunakan data uji langsung dari responden yang dipilih.



Gambar 1. Rancangan sistem



Gambar 2. Proses pelatihan data akustik

Data yang didapatkan berupa data korpus suara dan korpus teks. Data suara diperoleh dengan melakukan perekaman terhadap responden yang sudah dipilih. Data korpus teks dalam penelitian ini dibangun dengan mengambil data dari beberapa sumber seperti laman berita *online* Indonesia, laman *Sketchengine.co.uk*, dan laman kumpulan cerita rakyat Indonesia. Data tersebut kemudian diolah untuk dapat dijadikan data pelatihan pada sistem pengenalan suara. Data suara dirubah ke dalam format suara WAV dengan frekuensi 16 Khz 16 Bit *mono*.

Data dilatih menggunakan *cmusphinx* untuk mendapatkan *Acoustic Model* dan *Language Model*. Proses pelatihan dapat dilihat pada gambar 2. Sementara *Dictionary* dibuat secara manual berdasarkan *phonem* set Indonesia seperti pada table 1. Pelatihan dilakukan dengan melakukan eksperimen sebanyak 8 kali dengan nilai yang berbeda. Setiap eksperimen dilakukan perubahan konfigurasi *sphinxtrain* dengan merubah jumlah *senone*. Hal ini dilakukan untuk mendapatkan hasil pelatihan dengan nilai WER atau akurasi yang bagus untuk implementasi dan pengembangan ASR.

Tabel 1. Contoh keterangan tabel

Kategori fonetik	Fonem	Kata	Bentuk dasar urutan fonem	Urutan fonem setelah translasi
	/a/	saya	/s a y a/	/ s a y a/
	/e/	enak	/e n a k/	/ey n a k/
	/E/	kEmana	/k E m a n a/	/k eg m a n a/
	/i/	ingin	/i n g i n/	/i n g i n/
	/o/	orang	/o r a n g/	/o r a n g/
	/u/	untuk	/u n t u k/	/u n t u k/
	Diphthongs			
	/ai/	sungai	/s u n g a i/	/s u n g a i/
	/au/	danau	/d a n a u/	/d a n a u/
	/oi/	amboi	/a m b o i/	/a m b o i/
	Semi-vowels			
	/w/	wanita	/w a n i t a/	/w a n i t a/
	/y/	saya	/s a y a/	/s a y a/
	Consonants			
Plosives	/b/	berapa	/b e r a p a/	/b eg r a p a/
	/p/	petani	/p e t a n i/	/p eg t a n i/

	/d/	dia	/d i a/	/d i a/
	/t/	teman	/t e m a n/	/t e g m a n/
	/g/	giat	/g i a t/	/g i y a t/
	/k/	kamu	/k a m u/	/k a m u/
	/kh/	khairul	/k h a i r u l/	/k h a i r u l/
Africates	/j/	juga	/j u g a/	/j u g a/
	/c/	cinta	/c i n t a/	/c i n t a/
	/v/	video	/v i d e o/	/v i d e o/
Fricatives	/f/	maaf	/m a a f/	/m a a f/
	/z/	jenazah	/j e n a z a h/	/j e g n a z a h/
	/s/	saya	/s a y a/	/s a y a/
Liquids	/sy/	syahdu	/s y a h d u/	/s y a h d u/
	/h/	hujan	/h u j a n/	/h u j a n/
	/r/	ramai	/r a m a i/	/r a m a i/
	/l/	lambat	/l a m b a t/	/l a m b a t/
Nasals	/m/	mana	/m a n a/	/m a n a/
	/n/	mana	/m a n a/	/m a n a/
	/ny/	nyanyian	/n y a n y i a n/	/n y a n y i a n/
	/ng/	lambang	/l a m b a n g/	/l a m b a n g/

Dalam proses pelatihan model akustik, *sphinxtrain* melakukan penghitungan untuk mendapatkan hasil akurasi dari hasil pelatihan. *Sphinxtrain* melakukan *decoding* berdasarkan data uji yang disertakan dalam pelatihan. Dari 30 responden pembicara dipilih 4 laki-laki dan 4 perempuan untuk sampel data uji. Data uji yang disertakan berjumlah 1089 file *audio* yang berisikan suara ucapan. Hasil pelatihan model akustik yang dilakukan dapat dilihat pada tabel 2 dibawah.

Tabel 2. Hasil pelatihan model akustik

Senones	Jumlah Kata	Correct (%)	Error (%)	Insertion	Deletion	Substitution	Akurasi (%)	Wer (%)
150	11678	94,94	8,61	297	145	563	91,39	8,61
200	11678	94,48	7,76	261	128	517	92,24	7,76
250	11678	94,5	7,41	223	174	468	92,59	7,41
300	11678	94,47	7,7	253	146	500	92,3	7,7
350	11678	94,67	7,29	229	136	486	92,71	7,29
400	11678	94,4	7,75	251	141	513	92,25	7,75
450	11678	94,43	8,01	284	140	511	91,99	8,01
500	11678	94,85	7,61	287	92	510	92,39	7,61
Nilai rata-rata		94,6	7,62	246,22	135,56	507,89	92,38	7,62

Dari rangkaian 8 kali eksperimen diperoleh hasil rata-rata WER 7,62 % dengan akurasi rata-rata 92,38 %. Hasil akurasi tertinggi didapatkan pada saat konfigurasi *sphinxtrain* dengan jumlah *senone* 350. Hasil tersebut menjadi dasar untuk pengujian berikutnya. Perbedaan yang muncul tidak terlalu signifikan diantara semua eksperimen. Dengan demikian pada eksperimen ini dengan konfigurasi jumlah *senone* yang sedikit berbeda tidak terlalu mempengaruhi kualitas model akustik. Walaupun pada dasarnya semakin banyak *senone* yang dimiliki model maka semakin akurat mendiskriminasikan suara. Akan tetapi disini lain perlu diperhatikan jika terlalu banyak memberikan *senone* maka model tidak akan cukup generik untuk mengenali suara samar [9]. Ini artinya bahwa WER akan semakin tinggi terhadap data samar.

3. Hasil dan diskusi

Dari hasil pelatihan model akustik maka dilakukan pengujian secara *live* atau pengujian langsung terhadap sistem ASR dengan menggunakan *microphone*. Dalam pengujian ini dipilih 4 orang responden pembicara laki-laki dan 4 orang responden pembicara perempuan yang tidak terlibat dalam pengambilan data suara untuk pelatihan. Masing-masing responden membaca 100 kalimat, sehingga jumlah data uji keseluruhan yaitu 800 kalimat dengan jumlah kata 4232. Responden membaca dengan volume suara normal, kecepatan membaca sedang dan dalam ruangan yang sunyi.

Hasil ASR dibandingkan dengan transkrip kalimat teks pengujian. Kemudian dihitung jumlah kata yang benar dan yang salah. Kata yang salah terdiri dari kata yang tidak muncul (*deletion*), kata yang muncul tapi tidak ada dalam transkrip (*insertion*), dan kata yang diganti (*substitution*). Jumlah kata yang salah dibagi dengan jumlah total kata didapatkan tingkat kesalahan kata atau WER%. Sementara itu, akurasi ketepatan hampir sama dengan tingkat kesalahan kata, tetapi tidak menghitung sisipan.

Tabel 3. Hasil pengujian ASR secara *live*

Responden	Total kata	Benar (%)	Salah (%)	Sisipan	Penghapusan	Substitusi	Wer (%)	Akurasi (%)
F1	529	92,25	8,13	2	4	37	8,13	91,87
F2	529	91,12	10,4	8	3	44	10,4	89,6
F3	529	92,25	8,7	5	6	35	8,7	91,3
F4	529	92,25	8,51	4	10	31	8,51	91,49
M1	529	92,82	9,07	10	1	37	9,07	90,93
M2	529	92,25	10,96	17	4	37	10,96	89,04
M3	529	92,25	9,45	9	10	31	9,45	90,55
M4	529	90,55	10,78	7	6	44	10,78	89,22
						Rata-rata	9,5	90,5

Pengujian kinerja khususnya dalam hal kecepatan proses telah dilakukan. Pengujian kinerja dilakukan dengan menggunakan Komputer Sony Vaio dengan prosessor *Intel Core i5 2,30 Ghz*. Menggunakan RAM 8 Ghz dengan Sistem Operasi Windows 7 64 bit. Kemudian kami memilih tes dataset yang mewakili berbagai jenis masukan seperti yang ditunjukkan pada tabel 4.

Tabel 4. Hasil pengujian kinerja sistem

Masukan	Tipe	Speech to text	Text to sign	Speech to sign
Membaca	Kata	0,83	8,25	8,29
Saya membacakan buku kemarin	Kalimat Standar	0,97	20,02	20,06

Pengujian setiap masukan diulang sebanyak tiga (3) kali dengan rata-rata lamanya waktu menunggu mulai dari pengenalan perkataan hingga awal terjemahan video telah direkam. Dapat disimpulkan bahwa hasil rata-rata lamanya waktu menunggu berkisar dari 1,0 sampai 1,5 detik. Jika waktu rata-rata dihitung mulai dari pengenalan kata sampai akhir terjemahan video maka akan membutuhkan waktu yang lebih lama. Hal ini disebabkan oleh durasi video animasi rata-rata berkisar 1,72 detik. Sementara dalam satu (1) kata bisa saja terdiri dari satu (1) atau lebih video file animasi. Itu terjadi apabila kata tersebut mengandung imbuhan. Misalnya kata yang mengandung imbuhan "membacakannya" terdiri dari tiga (3) imbuhan dan kata dasar, jadi total video animasi dalam kata "membacakannya" terdiri dari empat (4) video animasi, yaitu animasi "me", "baca", "kan", dan animasi "nya".

4. Kesimpulan

Setelah melalui tahap perancangan dan pengujian aplikasi konversi suara ke bahasa isyarat Indonesia untuk anak tunarungu, maka dapat ditarik kesimpulan bahwa aplikasi ini dapat digunakan selain untuk belajar bahasa isyarat. Aplikasi SIBI dikhususkan untuk anak usia sekolah baik SD, SMP dan SMA. Aplikasi ini juga dapat digunakan oleh orang normal untuk memberikan

informasi sederhana atau perintah singkat kepada penyandang tunarungu. Hasil akurasi ASR yang diperoleh sebesar 90%. Hal ini masih perlu ditingkatkan terutama dalam kondisi lingkungan yang ribut dan berbagai variasi volume dan kecepatan berbicara dari para pengguna. Sehingga aplikasi ini dapat digunakan pada area yang vital seperti rumah sakit, bandara dan sebagainya. Adapun saran-saran terhadap pengembangan aplikasi ini untuk kedepan adalah perlu menambahkan data latih korpus yang lebih besar baik korpus suara maupun teks agar ASR dapat memberikan hasil yang lebih akurat. Juga selain menggunakan animasi avatar perlu juga menggunakan video manusia asli untuk representasi bahasa isyarat.

Referensi

- [1] D. P. Hallahan, J. M. Kauffman, and P. C. Pullen, "Exceptional Children: Introduction to Special Education. New Prentice Hall International." Inc, 1991.
- [2] UTONO and DANY ARIF BUDI, "PERANCANGAN BUKU VISUAL UNTUK ANAK TUNARUNGU USIA 10-12 TAHUN SEBAGAI MEDIA ALTERNATIF PEMBELAJARAN BAHASA | UTONO, DANY ARIF BUDI | Undergraduate Theses - ITS Institutional Repository," 2012. [Online]. Available: <http://digilib.its.ac.id/ITS-Undergraduate-3100012046277/18947>. [Accessed: 25-Sep-2020].
- [3] P. R. Indonesia, "Undang-undang Republik Indonesia nomor 20 tahun 2003 tentang sistem pendidikan nasional," *Jakarta Pemerintah Republik Indones.*, 2003.
- [4] D. R. Wasito, D. Sarwindah, and W. Sulistiani, "Penyesuaian Sosial Remaja Tuna Rungu yang Bersekolah di Sekolah Umum," *J. Insa. Univ. Hang Tuah Surabaya*, vol. 12, p. 3, 2010.
- [5] C. P. Wijayanto, "Membangun Aplikasi Pelatihan Bahasa Isyarat Berbasis Komputer Pada Orang Tunarungu," in *Sekolah Tinggi Manajemen Informatika Dan Komputer*, 2009.
- [6] R. Septiarti and H. Haryanto, "Konverter Suara dengan Input Bahasa Indonesia ke Video Gerakan Bahasa Isyarat dengan Metode Speech Recognition (Hidden Markov Model) untuk Penderita Tunarungu," *Semantik*, vol. 2, no. 1, 2012.
- [7] N. M. Rodliyah, I. K. E. Purnama, and A. Zaini, "Speech to Video (S2V), Sistem Komunikasi Portable untuk Penderita Tunarungu," *J. Tek. POMITS*, vol. 1, no. 1, pp. 1–8, 2013.
- [8] Mosyarofah, "Anak Tunarungu melalui Media Variasi Gambar pada Kelas I / B Semester II SDLBN Cilacap Tahun Pelajaran 2008 / 2009," UNIVERSITAS SEBELAS MARET, 2009.
- [9] J. Otander, "Basic concepts of speech recognition – CMUSphinx Open Source Speech Recognition." [Online]. Available: <https://cmusphinx.github.io/wiki/tutorialconcepts/>. [Accessed: 14-Nov-2019].