VIRTUAL ASSISTANT USED VOICE COMMAND

Wildan Azril Arvany[1], Anita Alfi Syahra[2], Roni Andarsyah[3]

[1,2,3] Universitas Logistik dan Bisnis Internasional
Sarijadi, Bandung, Jawa Barat, Indonesia
e-mail: [1]wildanazril27@gmail.com, [2]anitaalfi@gmail.com

Artikel Info : Diterima : 00-00-0000 | Direvisi : 00-00-0000 | Disetujui : 00-00-0000

Abstrak - Dengan keadaan perkembangan teknologi saat ini di bidang kecerdasan buatan dan komunikasi, banyak peneliti menulis tentang topik ini untuk mengembangkan teknik baru dan efektif. Asisten Suara adalah perangkat lunak yang dapat menafsirkan ucapan manusia dan merespons dengan ucapan yang disintesis. Apple Siri, Microsoft Cortana, dan Google Assistant adalah asisten suara yang paling umum dan disematkan pada teknologi. Dewasa ini, perkembangan teknologi semakin hari semakin meningkat. dengan bantuan sistem komputer, dimana hanya dapat melakukan beberapa tugas tetapi saat ini kecerdasan buatan, pembelajaran mesin dan beberapa teknologi lainnya memiliki komputer canggih sehingga dapat melakukan semua jenis tugas. Dalam artikel ini, "Virtual Assistant Used Voice Command" mencakup dua teknologi utama yaitu Synthesizer dan Pengakuan. Synthesizer ucapan mengambil input dan menghasilkan aliran audio sebagai output. Detektor suara di sisi berlawanan melakukan sebaliknya. Dibutuhkan aliran audio sebagai input dan dengan demikian mengubahnya menjadi transkripsi teks. Suara bisa menjadi sinyal informasi yang tak terbatas. Itu sebabnya membuat Asisten Pribadi Virtual sendiri untuk Windows yang hanya menggunakan Python 3.6 ke atas, yang dapat digunakan pada komputer, laptop atau PC. Aplikasi ini menggunakan Python sebagai bahasa pemrograman, sehingga memiliki library yang digunakan untuk menjalankan perintah.
Kata Kunci : Kecerdasan Buatan, Asisten Desktop, Python, Text-to-Speech, Asisten Virtual, Pengenalan Ucapan.

Abstracts *- With the current state of technological development in the field of artificial intelligence and communication, many researchers are writing on this topic to develop new and effective techniques. Voice Assistant is software that can interpret human speech and respond with synthesized speech. Apple Siri, Microsoft Cortana, and Google Assistant are the most common voice assistants and are embedded in the technology. Today, technological developments are increasing day by day. with the help of computer systems, which can only perform a few tasks but nowadays artificial intelligence, machine learning and some other technologies have advanced computers so that they can perform all types of tasks. In this article, "Virtual Assistant Used Voice Command" covers two main technologies namely Synthesizer and Recognition. The speech synthesizer takes input and produces an audio stream as output. The sound detector on the opposite side does the opposite. It takes an audio stream as input and thus converts it into a text transcription. Sound can be an infinite information signal. That's why we created our own Virtual Personal Assistant for Windows that only uses Python 3.6 and above, which can be used in computer, laptop, or PC. This application uses Python as a programming language, so it has a library used to execute commands.*
Keywords : Artificial Intelligence, Desktop Assistant, Python, Text-to-Speech, Virtual Assistant, Speech Recognition

1. PENDAHULUAN

Komunikasi merupakan salah satu ranah dalam ilmu sosial yang memiliki banyak implementasi didalam kehidupan bermasyarakat. Selain sebagai kebutuhan seorang individu untuk bisa tetap hidup, komunikasi juga menjadi salah satu dasar dari perkembangan teknologi yang semakin marak dilakukan pada hari ini. Perkembangan teknologi dalam ranah komunikasi sejatinya dilakukan untuk semakin mempermudah individu-individu dalam masyarakat untuk berinteraksi satu sama lain. Munculnya teknologi fisik seperti telepon genggam dan komputer serta teknologi perangkat lunak seperti sistem operasi dan situs menjadi sedikit dari banyaknya contoh teknologi untuk mempermudah proses komunikasi antar individu. Bentuk komunikasi yang

sebelumnya biasa dilakukan secara langsung tatap muka kini banyak berubah ke bentuk komunikasi tidak langsung (Kamil, 2016).

Kemajuan teknologi baru-baru ini dalam pengenalan ucapan telah membawa kita lebih dekat ke sistem kecerdasan buatan yang lengkap yang dapat berinteraksi dengan manusia dengan kecepatan komunikasi manusia-ke-manusia. Telah terbukti bahwa "*Voice Assistant*" (VA) yang dimanusiakan telah menjadi populer di beberapa sistem seperti Apple Siri dan Google Assistant di smartphone yang memungkinkan pengguna melakukan panggilan suara, Alexa di Amazon Echoes yang memungkinkan pengguna melakukan pemesanan belanja, asisten suara bertenaga AI di Mercedes yang memungkinkan pengemudi untuk mengubah pengaturan hands-free mobil. Dengan munculnya asisten suara ini, penting untuk memahami bagaimana perilaku asisten suara jika terjadi serangan yang disengaja. Banyak masalah keamanan asisten suara berasal dari perbedaan antara suara manusia dan mesin. Perangkat keras mikrofon asisten suara bertindak sebagai "telinga" yang mengubah gelombang akustik menjadi sinyal listrik, dan perangkat lunak pengenalan suara bertindak sebagai "otak" yang mengubah sinyal menjadi informasi semantik. Terlepas dari fungsinya yang tepat, sifat perangkat keras dan perangkat lunak yang tidak sempurna dapat menciptakan peluang bagi sinyal yang tidak biasa dalam komunikasi antarpribadi untuk diterima dan ditafsirkan dengan benar oleh asisten suara(N. Carlini, 2016).

2. LANDASAN TEORI

A. Voice Assistant

Sistem asisten suara terdiri dari tiga subsistem utama yaitu Perekaman audio, pengenalan ucapan, dan eksekusi perintah. Subsistem penangkap audio menangkap audio sekitar, yang diperkuat, disaring, dan di digitalkan sebelum di kirim ke subsistem deteksi audio. Kemudian sinyal digital mentah yang ditangkap diproses terlebih dahulu untuk menghilangkan frekuensi di luar rentang suara yang dapat didengar dan untuk menghilangkan segmen sinyal yang berisi suara yang terlalu lemah untuk dideteksi. Kemudian, sinyal yang diproses masuk ke sistem pengenalan suara. Biasanya, sistem pengenalan suara (SR) bekerja dalam dua fase diantaranya aktivasi dan pengenalan. Selama fase aktivasi, sistem tidak menerima input suara acak, tetapi menunggu aktivasi. Untuk mengaktifkan sistem, pengguna harus mengucapkan suatu kata kunci yang telah ditentukan sebelumnya atau menekan tombol khusus. Misalnya, Amazon's Echo menggunakan "Alexa" sebagai kata kuncinya. Apple Siri dapat diaktifkan dengan menekan dan menahan tombol Home selama sekitar satu detik, atau dengan menekan "Hey Siri" saat fitur "Allow Hey Siri" diaktifkan. Untuk mengenali kata kuncinya, mikrofon terus menangkap suara di lingkungan hingga suara tersebut ditangkap. Sistem kemudian menggunakan algoritma pengenalan suara tergantung speaker atau speaker independen untuk mengenali suara. Amazon Echo, misalnya, menggunakan algoritma yang tidak bergantung pada speaker dan akan menerima siapa pun yang mengatakan "Alexa" selama suaranya jelas dan keras. Sebagai perbandingan, Apple menggunakan Siri pada speaker. Siri harus dilatih oleh pengguna dan hanya akan menerima "Hey Siri" dari orang yang sama. Saat sistem Speech Recognition diaktifkan, ia masuk ke fase pengenalan dan biasanya menggunakan algoritme yang tidak bergantung pada speaker untuk mengubah ucapan menjadi teks, dalam hal ini perintah (C. Kasmi and J. L. Esteves, 2015).

Perhatikan bahwa SR yang bergantung pada speaker biasanya dilakukan secara lokal dan SR yang tidak bergantung pada speaker dilakukan melalui layanan cloud. Untuk menggunakan layanan cloud, sinyal yang diproses dikirim ke server, yang akan mengekstraksi fitur dan mengenali perintah melalui algoritme pembelajaran mesin (C. Ittichaichareon, 2018).

Diberi perintah yang dikenali, sistem eksekusi perintah akan meluncurkan aplikasi yang sesuai atau menjalankan operasi. Perintah yang dapat diterima dan tindakan yang sesuai bergantung pada sistem dan ditentukan sebelumnya. Asisten suara populer telah dibangun di ponsel cerdas, perangkat yang dapat dikenakan, perangkat rumah pintar, dan mobil. Smartphone memungkinkan pengguna untuk melakukan berbagai operasi melalui perintah suara, seperti menghubungi nomor telepon, mengirim pesan singkat, membuka halaman web, mengatur telepon ke mode pesawat, dll. Mobil modern menerima serangkaian perintah suara yang rumit untuk mengaktifkan dan mengontrol beberapa fitur dalam mobil, seperti navigasi, sistem hiburan, kontrol lingkungan, dan ponsel.

B. Microphone

Subsistem penangkap audio terutama perekam suara yang terdengar melalui mikrofon, yang merupakan transduser yang mengubah gelombang suara di udara (yaitu suara) menjadi sinyal listrik. Sebagian besar mikrofon adalah mikrofon kondensor, dan ada dua jenis mikrofon kondensor yang digunakan pada perangkat yang mengaktifkan suara yaitu Mikrofon kondensor elektret (ECM) dan mikrofon sistem mikroelektromekanis (MEMS). Dengan ukuran yang kecil, konsumsi daya yang lebih rendah, dan karakteristik suhu yang sangat baik, mikrofon MEMS mendominasi perangkat seluler, termasuk smartphone dan laptop. Namun, mikrofon ECM dan MEMS bekerja dengan cara yang sama. Mikrofon kondensor adalah kondensor celah udara yang berisi diafragma bergerak dan elektroda tetap (N. Roy, 2018).

Di hadapan gelombang suara, tekanan udara yang disebabkan oleh gelombang suara mencapai membran, yang tertekuk sebagai respons terhadap perubahan tekanan udara, sementara elektroda lainnya tetap diam. Pergerakan membran menyebabkan perubahan kapasitansi antara membran dan elektroda tetap. Karena muatan kapasitor tetap hampir konstan, perubahan kapasitansi menghasilkan sinyal AC. Dengan cara ini, tekanan udara diubah menjadi sinyal listrik.

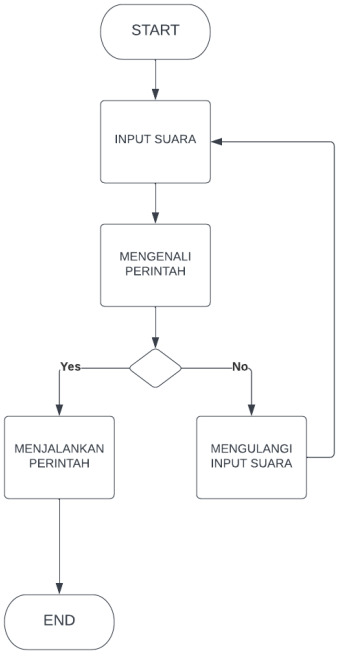
C. Speech Recognition

Adapun asisten virtual Google Now, Cortana dan Siri ketiganya menggunakan metode yang sama untuk mengeksekusi perintah pengguna melalui suara pengguna. Pada dasarnya, perintah suara menawarkan kemudahan dibandingkan metode lain seperti menggunakan keyboard. Karena suara merupakan alat komunikasi yang sederhana dan alami yang memudahkan orang untuk memberi perintah.Suara dapat memiliki karakteristik yang berbeda dan spesifik, karena setiap orang memiliki spektrum suara, frekuensi dan perbedaan yang luas.

Secara teknis, *Speech* atau *Automatic Speech Recognition (ASR)* adalah sebuah teknologi dan sistem yang memungkinkan sebuah komputer untuk menerima masukan ucapan berupa kata-kata yang diucapkan atau diucapkan, walaupun keluarannya saat ini masih terbatas pada kosa kata tertentu, namun tetap menjanjikan tahap perkembangan untuk seluruh dunia. Teknologi ini memungkinkan untuk mengenali dan memahami kata yang diucapkan dengan mendigitalkan kata-kata tersebut, setelah itu mesin mencocokkan sinyal digital dengan pola suara tertentu yang tersimpan di perangkat mesin. Mesin mengubah kata yang diucapkan menjadi sinyal digital dengan mengubah sekelompok gelombang suara. Perhitungan dan mencocokkan kode tertentu untuk mengenali kata yang diucapkan. Keluaran dari tag kata yang diucapkan dapat dilihat dalam bentuk tertulis atau dibaca sebagai pekerjaan dalam bentuk perintah yang dapat dibaca mesin, seperti menekan tombol pada ponsel dengan asisten digital bawaan (I.Satish, 2018).

3. METODE PENELITIAN

Saat pengguna meminta asisten pribadi untuk melakukan tugas, *natural language audio signal* direkonstruksi menjadi perintah yang mungkin dianalisis atau data digital yang dapat dianalisis oleh perangkat lunak, dan kemudian informasi ini dibandingkan dengan informasi perangkat lunak. Lalu, mencari respon yang benar dari asisten virtual adalah menggerakkan mesin sesuai dengan perintahnya sendiri.



Gambar 3.1 *Flowchart* Aplikasi

Flowchart adalah representasi simbolik dari suatu algoritma atau prosedur untuk memecahkan masalah. Alur aplikasi ini adalah:

a) Saat suara direkam oleh mikrofon, sinyal analog dari suara manusia diubah menjadi sinyal digital. Setelah didigitalkan, banyak model dapat digunakan untuk menyalin audio ke teks.

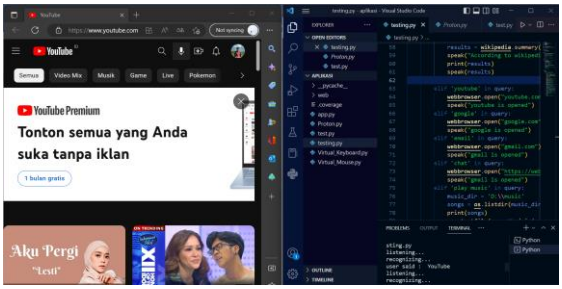
b) Sistem mengenali sinyal digital dan kemudian mengeksekusinya sesuai dengan program yang ditulis pada backend python. Backend python bekerja untuk mendapatkan output sebagai ganti input suara yang disediakan oleh pengguna melalui modul *Speech Recognition*.

c) Apabila suara dapat dikenali, maka aplikasi akan langsung mengeksekusi perintah. Apabila tidak, aplikasi akan merespon untuk mengulang kembali perintah suara.

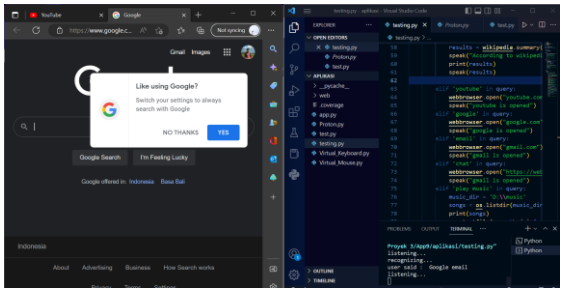
4. HASIL DAN PEMBAHASAN

A. Implementasi Aplikasi

Asisten virtual adalah perangkat lunak yang memahami perintah dan melakukan tugas yang ditentukan oleh pelanggan. Asisten virtual menggunakan NLP untuk mencocokkan suara atau teks pengguna dengan perintah yang akan dieksekusi. Asisten virtual memungkinkan untuk menggunakan komputer, laptop atau PC dengan perintah sendiri. Ini adalah proses cepat yang menghemat waktu. Hal ini dapat memanfaatkan dengan Python dan kecerdasan buatan. Tujuan utama adalah mendukung pengguna dalam tugas mereka dengan perintah suara. Ini dapat dilakukan dalam dua langkah. Pertama, dibutuhkan input suara pengguna dan mengubahnya menjadi kalimat bahasa Inggris menggunakan pengenalan suara Kedua, aplikasi akan menjalankan perintah suara dari pengguna sesuai dengan suara yang dikenali oleh sistem.



Gambar 4.1 Tampilan perintah membuka aplikasi YouTube



Gambar 4.2 Tampilan perintah membuka Google

B. Pengujian

Dalam pengujian aplikasi menggunakan metode pengujian *code coverage* untuk mengetahui apakah fungsi dan komponen alat dapat bekerja dengan baik. Pengujian yang dilakukan meliputi pengujian *source code* dan aplikasi saat dijalankan. Setiap fitur dan fungsi diuji dari sudut pandang pengguna untuk mengetahui hasil yang dicapai. Berikut adalah hasil pengujian :

```
PS D:\Proyek 3\App9\aplikasi> coverage run testing.py
listening...
recognizing...
user said :  time is it
listening...
recognizing...

listening...
recognizing...
user said :  Oita how are you
listening...
recognizing...
user said :  hello hello
listening...
recognizing...
user said :  open YouTube
listening...
recognizing...
user said :  open Google
listening...
recognizing...
```

Gambar 4.3 Aplikasi dijalankan

```
PS D:\Proyek 3\App9\aplikasi> coverage report
Name         Stmts   Miss  Cover
------------------------------
testing.py      77     21    73%
------------------------------
TOTAL           77     21    73%
PS D:\Proyek 3\App9\aplikasi>
```

Gambar 4.4 Hasil pengujian

5. KESIMPULAN DAN SARAN

Hasilnya menunjukkan bahwa teknik ini dapat digunakan secara efektif untuk pengenalan suara. Kami menambahkan fitur seperti B. yang hanya mendengarkan suara pengguna dan tidak diaktifkan oleh kebisingan sekitar. Modularitas proyek ini membuatnya mudah dipahami dan lebih fleksibel. Kami dapat menambahkan lebih banyak fitur ke aplikasi tanpa memengaruhi fungsinya. Semua paket yang diperlukan untuk bahasa pemrograman Python diinstal dan kode dijalankan menggunakan VS Code Integrated Development Environment (IDE). Versi Python yang digunakan dalam proyek ini adalah 3.x dan berbagai data audio dari lingkungan juga dicari.

Dalam komunikasi antara manusia dan mesin, pengaturan dilakukan oleh sinyal analog, yang diubah menjadi gelombang digital oleh sinyal audio. Teknologi ini tersebar luas, memiliki kegunaan yang tidak terbatas dan memungkinkan mesin merespons suara pengguna secara akurat dan menyediakan fungsionalitas yang berguna dan berharga. Sistem Pengenalan Ucapan (SRS) tersebar luas dan memiliki aplikasi tak terbatas. Analisis tersebut mengungkapkan fitur utama dari prosedur tersebut. Dalam beberapa hari mendatang, sistem yang direncanakan diharapkan dapat diimplementasikan sebagai aplikasi multibahasa, sehingga aplikasi tersebut dapat digunakan dalam bahasa Anda sendiri tanpa masalah. Selain itu, sistem yang direncanakan akan dibangun dengan IoT. Di masa mendatang, sistem yang kami rencanakan akan dapat menafsirkan deskripsi teks dengan lebih baik. Pengenalan gambar digunakan lebih detail dari gambar yang diambil.

6. DAFTAR PUSTAKA

[1] N. Carlini, P. Mishra, T. Vaidya, Y. Zhang, M. Sherr, C. Shields, D. Wagner, and W. Zhou (2016), "Hidden voice commands," in Proceedings of the USENIX Security Symposium.
[2] C. Kasmi and J. L. Esteves (2015), "IEMI threats for information security: Remote command injection on modern smartphones," IEEE Transactions on Electromagnetic Compatibility, vol. 57, no. 6, pp. 1752– 1755.
[3] C. Ittichaichareon, S. Suksri, and T. Yingthawornsuk (2018), "Speech recognition using MFCC," in Proceedings of the International Conference on Computer Graphics, Simulation and Modeling,
[4] N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury (2018), "Inaudible voice commands: The long-range attack and defense," in Proceeding of the 15th USENIX Symposium on Networked Systems Design and Implementation. USENIX Association.
[5] I. Satish And L. V. Kiran (2018), "Integrating Google Speech Recognition With Android Home Screen Application For Easy And Fast Multitasking," No. May, Pp. 0–3.
[6] K. V Kulhalli, K. Sirbi, M. A. J. Patankar, And Research (2017), "Personal Assistant With Voice Recognition Intelligence," Int. J. Eng. Res. Technol., Vol. 10, No. 1, Pp. 416–419.
[7] K. Khairunizam, D. Danuri, And J. Jaroji (2017), "Aplikasi Pemutar Musik Menggunakan Speech Recognition," Inovtek Polbeng - Seri Inform., Vol. 2, No. 2, P. 97, Doi: 10.35314/Isi.V2i2.196.
[8] I Komang Setia Buana, "Implementasi Aplikasi Speech to Text untuk Memudahkan Wartawan Mencatat Wawancara dengan Python," Jurnal Sistem dan Informatika (JSI), vol. 14, no. 2, hlm. 135–142, Agu 2020, doi: 10.30864/jsi.v14i2.293.
[9] N. F. I. Prayoga, "Analisis Speaker Recognition Menggunakan Metode Dynamic Time Warping (DTW) Berbasis Matlab," AVITEC, vol. 1, no. 1, Agu 2019, doi: 10.28989/avitec.v1i1.492.
[10] A. Fapal, T. Kanade, B. Janrao, M. Kamble, dan M. Raule, "Personal Virtual Assistant for Windows Using Python," www.irjmets.com @International Research Journal of Modernization in Engineering, vol. 485, no. 07, hlm. 485–491, 2021, [Daring]. Available: www.irjmets.com