

KOMPARASI DAN ANALISIS AI BASE NOISE SUPPRESSION: STUDI KASUS RTX VOICE

COMPARISON AND ANALYSIS ON AI BASE NOISE SUPPRESSION: STUDY CASE OF RTX VOICE

Yefta Christian¹, Herman², Chris Tan³

^{1,2,3}Universitas Internasional Batam
yefta@uib.ac.id

ABSTRACT

Since the COVID-19 pandemic, many people have realized that many of the activities we do daily can be done through online meetings. However, when we conduct online meetings, there is often a problem where the speaker's voice is not clear due to external interference such as keyboard sounds, other people's voices and several other factors. To overcome this problem, AI noise suppression can be used, which can help eliminate distracting sounds to clarify the speaker's voice. Currently there are two types of AI noise suppression, namely using hardware such as RTX Voice which uses AI capabilities on graphics cards and software that uses models that are trained on applications or through the cloud. The research method used in this research is experimentation. The results of this study prove that by using AI noise suppression the speaker's voice can become clear by removing other sounds that can interfere with the clarity of the speaker's voice.

keywords: Noise Suppression, Artificial Intelligence, RTX Voice, Krisp, Comparison

ABSTRAK

Sejak pandemi COVID-19 banyak orang menyadari banyak kegiatan yang kita lakukan sehari-hari dapat dilakukan melalui *online meeting*. Namun ketika kita melakukan *online meeting* sering terjadi masalah dimana suara pembicara tidak jelas karena gangguan luar seperti suara *keyboard*, suara orang lain dan beberapa faktor lainnya. Untuk mengatasi masalah tersebut dapat digunakan AI *noise suppression*, yang dapat membantu menghilangkan suara mengganggu untuk memperjelas suara pembicara. Saat ini ada dua jenis AI *noise suppression* yaitu menggunakan *hardware* seperti RTX Voice yang menggunakan kemampuan AI pada kartu grafis dan *software* yang menggunakan model yang di latih pada aplikasi atau melalui *cloud*. Metode penelitian yang digunakan pada penelitian ini ada eksperimen. Hasil dari penelitian ini membuktikan bahwa dengan menggunakan AI *noise suppression* suara pembicara dapat menjadi jelas dengan menghapus suara lainnya yang dapat mengganggu kejelasan suara pembicara.

Kata Kunci: Noise Suppression, Artificial Intelligence, RTX Voice, Krisp, Comparison

PENDAHULUAN

Sekarang kita dapat melakukan kelas, kerja, bertemu dengan teman dan banyak lagi secara *online*. Sejak masa pandemi COVID-19 orang sadar bahwa banyak kegiatan yang kita lakukan sehari-hari dapat dilakukan melalui *online meeting*. Untuk melakukan *online meeting* dibutuhkan komponen yang mendukung seperti perangkat keras dan perangkat lunak. Salah satu hal penting ketika melakukan *online meeting* adalah suara berbicara yang jelas agar tidak terjadi salah paham. Alasan suara berbicara tidak jelas terdapat beberapa faktor, seperti suara *keyboard*, orang lain yang berbicara di

belakang, suara jalan dan lain-lainnya (Fricke et al., 2022).

Ketika suara kita tidak jelas pada *online meeting* atau rekaman ada banyak hal yang dapat kita lakukan untuk memperbaiki masalah tersebut, seperti menggunakan mikrofon kondensor yang di mana mikrofon tersebut hanya mengambil suara dari satu arah atau menggunakan mikrofon yang memiliki pengaturan *pickup pattern* yang di mana pengguna dapat mengatur dari arah mana mikrofon tersebut akan mengambil suara. Namun tidak semua orang dapat mengganti mikrofon jika ingin karena sesuai dengan model dan kualitas yang dipilih dana yang diperlukan untuk

juga relatif tinggi. Selain mengganti mikrofon terdapat cara lain untuk mengatasi masalah tersebut.

Berdasarkan penelitian yang dilakukan (Chu et al., 2022) yang melakukan eksperimen tentang dampak suara angin dengan performa ANC (*Active Noise Control*) pada sebuah *headphone* di terowongan angin. Pada penelitian ini ditemukan kemampuan *noise reduction* berkurang dengan kecepatan angin. Dan dengan bantuan algoritma *noise suppression* yang dikembangkan untuk mengurangi suara angin saat berada di terowongan ini mendapatkan hasil yang lebih rendah lagi dibandingkan dengan algoritma konvensional.

Salah satu cara yang dapat dilakukan untuk memperjelas suara berbicara adalah dengan menggunakan *noise suppression*, yang di mana suara yang diterima ke komputer akan di filter untuk menghasilkan suara berbicara yang lebih jelas dengan menghilangkan suara yang mengganggu. Dengan perkembangan zaman sekarang ini *noise suppression* sudah didukung dengan bantuan AI untuk menganalisis suara berbicara untuk menghasilkan suara yang lebih jelas. Sekarang ini sudah banyak perangkat lunak yang mengandalkan AI untuk membantu mengatasi masalah berhubungan dengan kejelasan suara. Salah satu perangkat lunak yang menggunakan kemampuan *Graphic Processing Unit* (GPU) dan AI untuk menghapus suara yang mengganggu kejelasan suara berbicara adalah *Nvidia RTX Voice*. Seperti penelitian yang dilakukan oleh (Stumpf & Parker, 2021) yang telah melakukan eksperimen dengan melakukan siaran langsung di dalam laboratorium teknik mesin dengan menggunakan *RTX Voice* sebagai *noise suppression* untuk menghilangkan suara berisik pada ruangan tersebut. Perangkat lunak ini hanya bekerja pada komputer yang memiliki kartu grafis buatan *Nvidia*. Cara kerja *RTX Voice* adalah dengan dibuatnya mikrofon virtual pada komputer, yang di mana suara akan diambil dari mikrofon dan akan di filter melalui kartu

grafis untuk mendeteksi suara-suara yang mengganggu kejelasan pembicara. Semua ini terjadi pada jangka waktu yang sangat singkat bahkan instan.

Sedangkan perangkat lunak yang menggunakan AI *software-base* untuk meningkatkan kejelasan suara tanpa mengandalkan GPU seperti *Krisp*, yang dimana *Krisp* dapat digunakan pada komputer apa pun dengan *headset* dan mikrofon apa pun. Perbedaan antara *Krisp* dan *RTX Voice* adalah *Krisp* tidak perlu menggunakan kartu grafis buatan *Nvidia*, *Krisp* adalah aplikasi berbasis *subscription* namun dapat dicoba menggunakan secara gratis melalui *free trial* dan *Krisp* memerlukan koneksi internet selama penggunaan. Koneksi internet diperlukan karena cara kerja *Krisp* adalah dengan menyesuaikan situasi suara dengan model yang lebih cocok untuk memaksimalkan kinerja filter yang diambil melalui *cloud*. Seperti penelitian yang dilakukan oleh (Mukherjee & Jain, 2021) di mana mereka menemukan dengan menggunakan AI yang dapat mendeteksi geolokasi dan kondisi suara sekitar untuk memilih model *noise cancelling* seperti *Krisp* dapat memberikan hasil yang lebih baik dibandingkan dengan menggunakan model yang konvensional.

Belakangan ini sudah banyak aplikasi *meeting* yang memiliki fitur AI *noise suppression* bawaan yang dapat digunakan oleh semua peserta *meeting*. Namun kualitas filter tersebut tidak sebagus filter yang menggunakan kartu grafis atau *software-base* terpisah seperti *Krisp*. Hal ini karena model AI yang diajarkan hanya cukup untuk menutupi sampel yang kecil sehingga jika ada suara di luar yang diajarkan kepada model, suara tersebut tidak akan ditutupi. Selain itu metode penutupan suara tidak dilakukan secara efisien sehingga mengganggu suara pembicaraan yang dapat berakhir suara menjadi lebih tidak jelas.

Penelitian mengenai perbandingan AI *noise suppression* ini sangat jarang dilakukan karena hal ini masih tergolong baru. Dalam penelitian ini analisis terhadap

kejelasan suara berbicara dilakukan dengan membandingkan kualitas suara ketika menggunakan *RTX Voice* dengan suara ketika menggunakan *Krisp* dengan suara tanpa *suppression*. Suara yang dihasilkan akan diubah menjadi sinyal digital yang kemudian dapat dilihat perbedaannya. Oleh karena itu penelitian ini diberi judul “Komparasi dan Analisis AI Base Noise Suppression: Studi kasus RTX Voice”.

Selama penulisan penelitian menggunakan landasan teori berikut:

1. *Noise Suppression*

Menurut (Braun & Tashev, 2020), *Noise Suppression* adalah *speech enhancement* yang bekerja untuk mengurangi *noise* ataupun suara latar belakang seperti suara orang musik di tempat umum, suara kerumunan, dan suara bisik lainnya.

Noise Suppression adalah sebuah proses yang dilakukan pada sebuah perangkat atau aplikasi untuk menghapus *noise* dan suara latar belakang untuk memperjelas suara pembicara.

2. *Artificial Intelligence*

Menurut (Siahaan et al., 2020), *Artificial Intelligence* adalah kecerdasan yang ditambahkan kepada sistem yang dapat diatur dalam konteks ilmiah. Kecerdasan tersebut dibuat agar dapat melakukan pekerjaan seperti yang dilakukan manusia. Dapat didefinisikan sebagai AI jika sistem mampu menafsirkan data external dengan benar, dapat dipelajari dan menggunakan pembelajaran untuk beradaptasi dan mencapai tujuan.

3. *RTX Voice*

RTX Voice adalah *plugin* yang dibuat oleh *Nvidia* menggunakan kemampuan kartu grafis *Nvidia RTX* dan kemampuan AI kartu grafis untuk menghilangkan suara latar belakang yang mengganggu kejelasan suara pada siaran langsung, rekaman, meeting dan lain-lain yang berhubungan dengan penggunaan mikrofon.

4. *Krisp*

Krisp adalah aplikasi yang dapat digunakan pada komputer maupun

smartphone yang berguna untuk memperjelas suara orang yang berbicara dengan menggunakan AI. Aplikasi ini akan membuat mikrofon virtual yang akan menerima suara dan memproseskan suara tersebut dengan model AI yang cocok pada situasi suara yang diterima. Aplikasi ini dapat digunakan pada perangkat apa pun selama perangkat tersebut dapat menerima inputan suara. Namun aplikasi ini perlu tersambung pada koneksi internet dan untuk menggunakan filternya secara tidak terbatas pengguna perlu membayar *subscription*, atau dengan menggunakan *free trial* yang di mana pengguna diberikan batas 1 jam penggunaan per hari.

5. *Komparasi*

Komparasi menurut (Hanik & Harsono, 2020) adalah metode yang digunakan untuk membandingkan data-data yang ditarik ke dalam konklusi baru. Dengan melakukan perbandingan, kita dapat mengetahui apakah dari dua data tersebut ada perbedaan yang signifikan yang dapat kita pelajari dan analisa. Dengan mengetahui perbedaan tersebut kita dapat mempelajari dan memperoleh informasi yang dapat digunakan untuk menjawab pertanyaan yang relevan.

6. *Analisis*

Analisis menurut (Nurhasanah et al., 2021) adalah proses identifikasi sesuatu dengan tujuan memecahkan masalah atau menemukan kekurangan dari keadaan sebenarnya sehingga dapat menemukan solusi yang diharapkan. Dengan melakukan analisis kita dapat memahami dan menjelaskan apa yang terjadi pada proses yang telah dilaksanakan dan menarik konklusi dari apa yang kita dapat pada analisis.

7. *Audacity*

Audacity adalah aplikasi gratis dan *open source* yang digunakan untuk *multi-track* audio editor dan perekam untuk sistem operasi Windows, macOS, Linux dan sistem operasi lainnya. Aplikasi ini dapat digunakan untuk merekam suara dari mikrofon, merekam suara dari komputer, mengkonversi media musik fisik menjadi

media musik digital, mengedit suara, memberi efek pada suara, dan menganalisis suara.

8. Spectrogram

Spectrogram adalah representasi visual frekuensi sebuah sinyal yang berbeda seiring waktu. *Spectrogram* biasa digunakan pada bidang musik, ilmu bahasa, sonar, radar, dan pengolahan wicara. Visual yang ditunjukkan pada *spectrogram* adalah kekuatan atau kebisingan sinyal, di mana warna hitam berarti tidak ada sinyal pada frekuensi tersebut dan warna biru hingga oren menunjukkan adanya sinyal pada frekuensi tersebut, di mana warna biru menunjukkan sinyal yang lemah dan oren menunjukkan sinyal yang kuat.

9. Frekuensi Suara

Frekuensi Suara merujuk pada berapa banyak gelombang suara pada satu detik siklus suara, yang di mana semakin banyak gelombang suara pada satu detik tersebut maka suara yang diterima akan terdengar lebih tinggi nadanya. Telinga manusia dapat menerima suara pada jarak frekuensi sekitar 20 Hz hingga 20000 Hz dan jarak ini juga digunakan untuk merepresentasikan frekuensi pada suara. Pada frekuensi suara biasa dibagi menjadi 7 bagian yaitu: *Sub-Bass* (16 - 60 Hz), *Bass* (60 - 250 Hz), *Lower Midrange* (250 - 500 Hz), *Midrange* (500 - 2000 Hz), *Higher Midrange* (2 - 4 kHz), *Presence* (4 - 6 kHz), *Brilliance* (6 - 20 kHz).

9.1 *Sub-Bass* (16 - 60 Hz)

Pada frekuensi ini dibutuhkan *headphone* yang berkualitas atau memiliki frekuensi yang penuh atau memiliki *subwoofer* karena pada frekuensi ini kita tidak mendengarkan tapi lebih ke merasakan karena kita sangat kesulitan mendengarnya. Contohnya seperti instrumen *kick drums*, *bass synthesizers*, dan instrumen orkestra seperti harpa, dan bassoon.

9.2 *Bass* (60 - 250 Hz)

Pada frekuensi ini manusia lebih mudah mendengarkan suaranya karena berada pada frekuensi kemampuan manusia, namun untuk memproduksi suara

pada frekuensi ini membutuhkan lebih banyak tenaga dibanding frekuensi yang di atas. Pada frekuensi ini kita dapat mendengarkan apakah lagu ini terdengar lebih ringan atau lebih tebal.

9.3 *Lower Midrange* (250 - 500 Hz)

Frekuensi ini penting karena mengandung *fundamental frequency* di mana suara pada frekuensi ini dapat menentukan *pitch* suara dan frekuensi ini selalu bagian suara dengan volume paling keras.

9.4 *Midrange* (500 - 2000 Hz)

Frekuensi ini penting karena pada frekuensi ini mengandung dialog, vokal, dan suara efek pada film dan suara TV. Ketika produser ingin membuat lagi menjadi lebih “pop” di radio mereka akan fokus pada frekuensi ini. Namun jika frekuensi 1 kHz dinaikan terlalu tinggi maka musik dapat menyebabkan *Listener fatigue* atau pendengar menjadi capek, dan tidak nyaman.

9.5 *Higher Midrange* (2 - 4 kHz)

Frekuensi ini adalah frekuensi yang paling sensitif, di mana suara konsonan seperti huruf k, p, s, t dapat ditemukan pada frekuensi ini. Oleh karena itu jika lagu kekurangan frekuensi ini maka kita akan langsung menyadarinya, namun jika terlalu di empasiskan maka lagu akan menjadi sulit di dengar.

9.6 *Presence* (4 - 6 kHz)

Frekuensi ini dapat membuat lagu menjadi lebih terasa hangat dan kemanisan pada alat musik orkestra, *synths*, dan *high strings*. Namun jika berlebihan maka akan menjadi kekerasan dan bunyi berdesis. Jika dilakukan dengan benar maka dapat terasa seperti sedang berada di depan musik yang sedang dimainkan.

9.7 *Brilliance* (6 - 20 kHz)

Pada *range* ini sesuai dengan namanya akan membuat lagu menjadi lebih cemerlang, namun jika diberikan berlebihan maka akan menghasilkan bunyi berdesis. Di atas 10 kHz sering disebut sebagai “air” band.

10. OBS

Menurut (Qorib & Zaniyati, 2021), OBS atau Open Broadcast Software adalah aplikasi dapat digunakan untuk merekam video dan melakukan siaran langsung yang mudah diakses dan gratis. Aplikasi OBS dirilis pada 1 september 2012 yang dibangun menggunakan bahasa pemrograman C/C++ dan dibangun menggunakan Qt. Aplikasi ini memiliki banyak fitur yang dapat mempermudah pengguna seperti melakukan live editing saat merekam atau melakukan siaran langsung, mengatur besar window yang ditampilkan, mengatur sumber suara dan volume, dan banyak lagi.

11. Virtual Audio Cable

Virtual Audio Cable adalah aplikasi yang membuat perangkat suara virtual, dimana input suara akan diteruskan ke output suara dan sebaliknya, suara output diteruskan ke input suara. Dengan aplikasi ini pengguna dapat membuat suara yang dimainkan di desktop menjadi suara yang keluar dari mikrofon kita dan sebaliknya, suara dari mikrofon kita menjadi suara yang keluar dari desktop kita.

12. Metode Eksperimen

Menurut (Rohmi Khalida & Astawan, 2021), metode eksperimen adalah metode pembelajaran yang mengarahkan seseorang untuk melaksanakan sebuah percobaan secara mandiri sehingga dapat mengalami dan membuktikan sendiri hal yang dipelajari. Metode ini memberikan kesempatan untuk mengalami suatu percobaan secara mandiri dan mengikuti semua prosesnya hingga dapat menarik kesimpulan sendiri. Dengan metode eksperimen pelaksana dapat mengkonstruksi pengetahuan secara mandiri dan mengembangkan kemampuan berpikir ataupun kreativitas

METODE

Pada penelitian ini penulis menggunakan jenis penelitian eksperimen. Penelitian ini bertujuan untuk mencari tahu pengaruh AI *noise suppression* pada kualitas suara berbicara. Sebelum memulai penelitian ini, penulis akan mencari

referensi mengenai hal yang berhubungan dengan *noise suppression*, cara kerja AI *noise suppression*, dan penjelasan mengenai kejelasan suara. Setelah penulis mendapatkan referensi yang cukup dari penelitian yang sudah pernah dilakukan sebelumnya, penulis akan merancang proses yang akan dilakukan selama penelitian dari referensi yang telah didapatkan.

Penelitian akan dilakukan dengan eksperimen oleh karena itu pengumpulan data akan dilakukan dengan melakukan perekaman suara pada kondisi yang menantang sebagai data awal yang akan digunakan untuk diterapkannya berbagai jenis AI *noise suppression* dan juga akan digunakan sebagai garis dasar perbandingan. Tahap selanjutnya adalah mengolah data yang baru saja direkam dan proses menggunakan filter AI *noise suppression*. Pada tahap ini file suara akan dibuat menjadi data yang dapat dilihat yaitu membuat gambar *spectrogram* dari masing-masing file suara. Selain itu file suara juga akan proses untuk menghasilkan data frekuensi suara berupa angka frekuensi (Hz) dan level (dB).

Setelah data tersebut diubah, penulis dapat melakukan analisis masing-masing data yang telah dikumpulkan dan dilakukan perbandingan untuk melihat perbedaan sebelum dan setelah. Dari hasil analisis penulis akan membahas apa saja yang berubah sebelum dan sesudah filter AI *noise suppression* diterapkan pada file suara, dan dari hasil yang didapatkan dapat memahami hal-hal yang membuat hasil kedua AI berbeda satu sama lain dari referensi yang telah dicari. Penulis akan menjelaskan mengapa hasil dari kedua AI berbeda dan apa yang dapat membuat mereka berbeda. Terakhir penulis akan menarik kesimpulan dari keseluruhan penelitian yang telah dilakukan.

1. Penentuan Konsep

Pada tahap ini penulis akan menentukan apa saja yang akan menjadi faktor yang diperhatikan ketika menganalisis hasil dari data yang telah

dikumpulkan. Setelah melihat beberapa referensi mengenai *noise suppression*, penulis akan memilih hal-hal yang akan terjadi selama perekaman untuk memastikan suara rekaman tersebut dapat mencoba potensial AI *noise suppression* semaksimal mungkin dan mendapatkan hasil yang dapat terlihat jelas. Dengan adanya situasi yang ditentukan penulis dapat menganalisis apakah suara hal-hal yang terjadi dapat mempengaruhi kejelasan suara pembicara setelah diterapkan filter sehingga membuat suara menjadi lebih jelas atau menjadi tidak jelas.

2. Pengumpulan Data

Pada tahap ini, akan dilakukan perekaman suara yang kemudian digunakan sebagai garis dasar dan sebagai suara yang akan diterapkan filter AI *noise suppression* untuk dilakukan perbandingan. Selama perekaman hal-hal yang sudah ditentukan akan terjadi selama perekaman suara untuk menantang kemampuan AI *noise suppression* akan dilaksanakan sehingga mendapatkan hasil yang memenuhi kriteria yang sudah ditentukan. Untuk mendapatkan hasil suara setelah diterapkan filter yang konsisten, penulis akan menggunakan *virtual audio cable* untuk memanipulasi sumber suara desktop menjadi suara yang keluar dari mikrofon. Dengan ini penulis dapat menggunakan file suara yang sama namun dengan filter yang berbeda untuk mendapatkan hasil akhir yang konsisten. Jika dilakukan 3 kali perekaman pada waktu berbeda untuk mencoba 2 jenis AI *noise suppression* maka hasil yang didapatkan tidak akan konsisten karena perbedaan durasi, perbedaan suara yang dihasilkan pembaca dan perbedaan kondisi suara latar belakang yang dapat memungkinkan terjadinya hasil yang tidak konsisten.

3. Pengolahan dan Analisis Data

Pada tahap pengolahan dan analisis data penulis akan mengumpulkan jurnal yang akan digunakan sebagai referensi dalam proses pengolahan dan analisis data untuk memahami apa yang dilakukan AI *noise suppression* dan memahami hasil dari

penerapan filter tersebut. Setelah mengumpulkan referensi, penulis akan menentukan hal-hal yang akan dilakukan sebagai proses pengolahan data. Data rekaman tersebut akan diolah menjadi gambar *spectrogram* yang akan menunjukkan sinyal suara dan kekuatan sinyal suara pada waktu tersebut. Berdasarkan penelitian (Circic et al., 2021) dengan mengubah sinyal suara menjadi gambar *spectrogram* tidak hanya membuat data tersebut menjadi lebih mudah dianalisis tetapi juga menunjukkan hal yang unik tentang suara tersebut. Dengan *spectrogram* kita dapat melihat suara apa saja yang berhasil dihapus atau di minimalis untuk mendapatkan hasil suara yang lebih jelas. Selain membuat *spectrogram* penulis juga akan membuat grafik yang menunjukkan frekuensi suara yang dihasilkan dari rekaman tersebut. Frekuensi suara akan menunjukkan level suara dalam hitungan desibel pada frekuensi tertentu, yang akan menunjukkan seberapa keras suara pada frekuensi tertentu. Pada grafik ini dapat dilihat suara apa saja yang berhasil di redupkan untuk memperjelas suara.

4. Pembahasan Hasil Penelitian

Pada tahap pembahasan hasil penelitian, penulis akan membahas hasil yang telah ditemukan dari data yang telah diolah dan dianalisis serta menjelaskan korelasinya dengan dua jenis AI *noise suppression* yang digunakan pada perekaman suara tersebut. Dari hasil gambar *spectrogram* penulis akan menunjukkan perbedaan dari tiga suara tersebut serta menjelaskan hasil yang ditampilkan untuk lebih lanjut memperjelas dampak AI *noise suppression* terhadap file suara tersebut. Setelah itu penulis juga akan menunjukkan grafik yang dibuat dari data frekuensi suara di mana *axis-x* akan menunjukkan frekuensi (Hz) dan *axis-y* akan menunjukkan level (dB). Dari grafik ini penulis akan menjelaskan arti dari frekuensi dan jenis suara yang berada pada kelompok frekuensi tersebut, serta menjelaskan level suara.

5. Menarik Kesimpulan

Langkah terakhir pada penelitian ini adalah penulis akan menarik kesimpulan dari hal-hal yang telah ditemukan oleh penulis selama penelitian ini dan kesimpulan penelitian yang telah dilakukan.

Perangkat Keras atau Hardware adalah perangkat fisik yang digunakan untuk melakukan sebuah proses. Pada penelitian ini perangkat keras akan digunakan untuk melakukan perekaman, penerapan filter AI *noise suppression*, dan melakukan analisis. Perangkat keras yang digunakan selama proses penelitian ini adalah:

Tabel 1. Spesifikasi Hardware

Nama	Keterangan
Komputer	Lenovo Legion 5 15ACH6H
Prosesor	AMD Ryzen 7 5800H
Memori	16GB DDR4 3200 MHz
Kartu Grafis	NVIDIA® GeForce RTX™ 3060 Laptop GPU
<i>Earphone</i>	KZ ZSN PRO X
Sistem Operasi	Windows 10 Home 64-bit

Selain *hardware* juga digunakan perangkat lunak atau *software*. Perangkat lunak atau *software* adalah sekumpulan instruksi atau perintah yang memberi tahu komputer apa yang perlu dilakukan. Pada penelitian ini perangkat lunak akan digunakan untuk mengumpulkan data, dan menganalisis data. Perangkat lunak yang digunakan selama penelitian ini adalah:

1. Audacity

Perangkat lunak ini digunakan sebagai sarana untuk menganalisis data, yang di mana aplikasi ini akan menghasilkan data frekuensi suara dan menghasilkan gambar *spectrogram*.

2. Virtual Audio Cable

Perangkat lunak ini digunakan untuk memanipulasi sumber suara, sehingga suara *output* akan menjadi suara *input*.

3. OBS

Perangkat lunak ini digunakan untuk merekam suara sebelum diterapkan

filter, dan menerapkan filter AI *noise suppression* pada suara.

4. RTX Voice

RTX Voice digunakan sebagai AI *noise suppression* yang berbasis *hardware*.

5. Krisp

Krisp digunakan sebagai AI *noise suppression* yang berbasis *software*.

6. Microsoft Excel

Perangkat lunak ini digunakan untuk mengubah data frekuensi suara menjadi bentuk grafik.

Metode yang digunakan untuk mengumpulkan data adalah dengan melakukan perekaman suara dan memanipulasi sumber suara untuk menerapkan filter pada suara yang telah direkam. Data suara filter dikumpulkan dengan cara tersebut dengan tujuan mendapatkan hasil yang lebih konsisten karena menggunakan suara yang sama dengan suara garis dasar. Selama perekaman akan ada pembicara yang membacakan naskah yang akan ditentukan, sebagai suara yang ditargetkan oleh filter untuk diperjelas. Setelah perekaman, suara yang telah direkam tersebut akan diproses menggunakan filter *RTX Voice*, dan *Krisp* untuk mendapatkan hasil yang konsisten karena menggunakan file suara yang sama.

Namun, karena *RTX Voice* dan *Krisp* adalah aplikasi yang menerapkan AI *noise suppression* secara langsung, penulis perlu melakukan manipulasi sumber suara agar dapat menggunakan filter tersebut pada file suara yang sudah direkam. Untuk memanipulasi sumber suara diperlukan aplikasi *virtual audio cable* yang di mana aplikasi ini akan membuat perangkat virtual baru pada komputer dengan nama "*Cable Output*" pada bagian mikrofon. Dengan ini kita dapat memanipulasi sumber suara yang berasal dari desktop komputer menjadi suara yang muncul dari mikrofon atau *audio input*. Setelah itu penulis akan menggunakan OBS untuk merekam hasil suara setelah diterapkan filter. Untuk filter diterapkan dengan benar, pada aplikasi OBS perlu mengubah pengaturan perangkat mikrofon menjadi *virtual audio cable* agar

mengubah suara yang diterima menjadi suara yang keluar dari desktop dan menerapkan filter pada suara desktop itu. Setelah mendapatkan tiga file suara yaitu suara sebelum diterapkan filter, suara setelah diterapkan filter *RTX Voice*, dan suara setelah diterapkan filter *Krisp*, suara tersebut akan dianalisis dengan melakukan konversi sinyal suara menjadi gambar *spectrogram*. Untuk mendapatkan gambar tersebut tiga file suara pertama akan dimasukkan ke dalam aplikasi *Audacity* kemudian kita dapat mengubah jenis representasi sinyal audio dari *waveform* menjadi *spectrogram*.

Dengan *spectrogram* kita dapat melihat suara pada frekuensi mana yang berhasil dihapus setelah diterapkan dan seberapa bersih suara yang dihasilkan. Selain itu juga akan dibuat grafik frekuensi suara yang dapat digunakan untuk menunjukkan perbedaan suara sebelum dan sesudah penerapan filter per kelompok frekuensi. Untuk membuat grafik frekuensi suara pertama perlu dimasukkan suara ke dalam aplikasi *Audacity*. Setelah memilih *track* suara pada menu *analyze* terdapat pilihan plot *spectrum*, yang akan memberikan pilihan algoritma, fungsi *size*, *axis* serta gambar grafik frekuensi suara. Namun dengan grafik yang sudah diberikan kita tidak bisa melihat perbandingan ketiga suara secara langsung, oleh karena itu penulis akan *export* hasil frekuensi tersebut ke dalam bentuk teks yang kemudian dimasukkan ke dalam Microsoft Excel. Pada *Microsoft Excel* angka yang dihasilkan dari *export* akan ditampilkan dalam bentuk grafik sehingga menampilkan tiga garis pada posisi berbeda yang dapat digunakan untuk membandingkan ketiga hasil tersebut

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pada tahap perekaman suara penulis membuat skenario yang menantang agar dapat melihat potensial maksimal dari filter *AI noise suppression* tersebut. Skenario yang terjadi selama perekaman adalah:

1. Suara Statis

Suara statis sering ditemukan ketika tingkat sensitivitas mikrofon terlalu tinggi, namun hal ini juga sering terjadi pada mikrofon memiliki kualitas rendah. Pada skenario ini mikrofon yang digunakan adalah mikrofon dari *earphone* dengan level volume 100 dan *boost* pada +30 dB di pengaturan audio pada Windows 10 agar suara pembicara dapat terdengar jelas dan semua suara yang berada pada lingkungan sekitar dapat ditangkap. Dengan suara *statik* kita dapat mengevaluasi apakah *AI noise suppression* dapat menghilangkan suara statis tanpa mengganggu kejelasan suara berbicara karena suara statis selalu ada pada rekaman.

2. Musik *Background*

Saat perekaman suara, sedang ada lagu yang dimainkan melalui *speaker* yang tidak terlalu jauh dan cukup keras untuk tertangkap oleh mikrofon saat perekaman. Dengan ini kita akan mengetahui apakah *AI noise suppression* mampu membedakan orang yang sedang berbicara dengan suara orang sedang menyanyi pada lagu yang dimainkan.

3. Orang Lain Berbicara

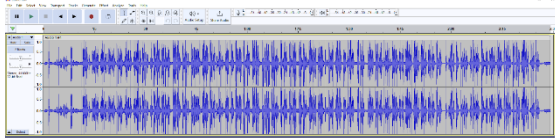
Selain orang yang sedang merekam suara, ada juga orang lain yang berbicara di ruangan yang sama. Karena *AI noise suppression* yang digunakan dilatih dengan tujuan memperjelas suara orang yang sedang berbicara, apakah *AI* tersebut dapat membedakan orang yang sedang merekam dengan orang lain yang sedang berbicara dan menghapus suara yang tidak berhubungan.

4. Suara "*popping*"

Suara "*popping*" muncul terutama pada pengucapan konsonan letup atau *plosive* seperti huruf 'p' pertama dalam kata bahasa Inggris "*popping*". Suara *popping* ini sulit dihilangkan karena suara ini dapat menutupi suara lain, oleh karena itu jika *AI noise suppression* tidak dapat mengatasi dengan baik hasil dari filter adalah suara yang menghilang karena suara *popping* dianggap sebagai suara yang mengganggu kejelasan pembicaraan.

Naskah yang akan dibaca pada rekaman ini diambil dari *Harvard Sentences* yaitu naskah yang di standarisasi untuk pengujian *Voice over IP (VoIP)*, seluler, dan sistem telekomunikasi lainnya. Kalimat pada *Harvard Sentences* adalah kalimat yang seimbang secara fonetis sehingga cocok digunakan untuk kualitas pembicaraan.

Perekaman suara awal sebelum diterapkan filter akan dilakukan menggunakan aplikasi *audacity*. Setelah suara direkam melalui *audacity*, hasil rekaman tersebut akan direkam ke dalam file format MP3, dan juga akan disimpan ke dalam bentuk *.aup3* yaitu format file *audacity 3 Project File* sebagai file audio sebelum dikonversi menjadi file MP3.



Gambar 1. Hasil Rekaman Suara Pertama

B. Pengolahan Data

Pengolahan data dimulai dengan menerapkan filter pada hasil suara yang telah direkam. Pertama file suara tersebut akan di rekam dengan menerapkan filter *RTX Voice* dan *Krisp*.

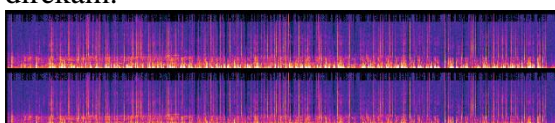


Gambar 2. Hasil Rekaman RTX Voice

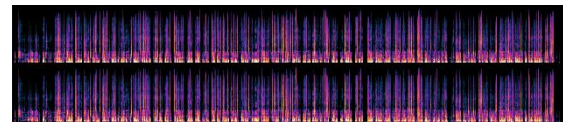


Gambar 3. Hasil Rekaman Krisp

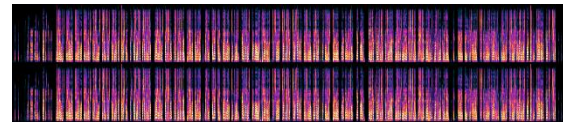
Setelah merekam hasil suara setelah diterapkan filter *RTX Voice* dan *Krisp*, tahap selanjutnya adalah untuk membuat gambar *spectrogram* dari hasil suara yang direkam.



Gambar 4. Spectrogram Rekaman Suara Awal

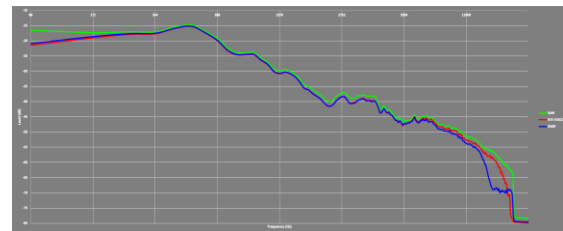


Gambar 5. Spectrogram Hasil Rekaman RTX Voice



Gambar 6. Spectrogram Hasil Rekaman Krisp

Tahap terakhir dalam proses pengolahan data ini adalah membuat grafik frekuensi suara yang dihasilkan oleh hasil rekaman suara tersebut.



Gambar 7. Hasil Grafik Frekuensi Suara

Warna hijau menunjukkan garis frekuensi suara awal sebelum *noise suppression*, garis berwarna merah menunjukkan frekuensi suara setelah *RTX Voice* dan garis berwarna biru menunjukkan frekuensi suara setelah *Krisp*.

C. Hasil Analisis

Dari data yang sudah diproses dapat di lihat perubahan yang signifikan dari suara sebelum penerapan AI *noise suppression*, setelah diterapkan *RTX Voice* dan setelah diterapkan *Krisp*.

Dimulai dengan menganalisa hasil rekaman suara pertama sebelum diterapkan *noise suppression* dan sebelum mengolah data apa pun. Dari hasil rekaman di *audacity* sudah dapat melihat hasil sinyal setelah suara selesai direkam sangat berisik. Dari awal rekaman sebelum pembicara mulai berbicara sudah dapat dilihat bahwa semua suara yang berada disekitar tertangkap dengan jelas yang dapat dilihat pada 2 garis vertikal yang muncul pada awal. Selain itu juga dapat dilihat bahwa suara *background* tertangkap jelas pada mikrofon karena sebelum pembicara mulai berbicara dan setelah selesai berbicara

terdapat garis horizontal yang cukup tebal karena suara yang ada disekitar.

Dibandingkan dengan hasil rekaman *RTX Voice* sinyal suara yang dihasilkan lebih bersih dibandingkan rekaman awal dimana pada awal dapat dilihat 2 garis vertikal yang dapat terlihat sebelum sekarang menjadi lebih kecil dan dihapuskan. Selain itu sinyal yang dihasilkan juga menjadi jauh lebih tenang dimana setiap 1 kalimat dapat terlihat jedah dimana tidak ada suara yang terekam dan sinyal yang dihasilkan pembicara menjadi lebih jelas. Dapat di lihat juga pada setiap jedah garis horizontal yang dihasilkan lebih tipis yang menandakan tidak ada suara yang tertangkap.

Pada hasil rekaman *Krisp* sinyal suara yang dihasilkan jika dibandingkan dengan hasil rekaman *RTX Voice* terlihat mirip, di mana *Krisp* berhasil memberikan hasil lebih bersih dan hasil yang lebih jelas. Dapat dilihat juga bahwa hasil dari rekaman *Krisp* lebih bagus karena suara pada awal rekaman berhasil dihapus.

Selanjutnya dapat dianalisis hasil *spectrogram* yang dihasilkan dari rekaman suara. Pada *spectrogram* terdapat warna putih, oren, ungu, biru tua, dan hitam, dimana warna putih menunjukkan sinyal yang sangat kuat dan warna hitam menunjukkan tidak adanya sinyal.

Dari *spectrogram* rekaman suara awal dapat di lihat banyak warna ungu yang muncul karena adanya sinyal suara *background* dan suara statis yang dihasilkan mikrofon sehingga tidak ada sinyal berwarna hitam karena selalu ada suara yang tertangkap selama perekaman. Oleh karena itu sinyal suara berwarna putih dan oren tidak terlalu jelas yang berarti suara pembicaraan tidak terlalu jelas karena terganggu oleh suara lainnya.

Dibandingkan dengan *spectrogram* hasil *RTX Voice*, sinyal warna ungu sudah sangat banyak dihapus sehingga hasilnya sinyal berwarna hitam menjadi lebih banyak yang menunjukkan AI *noise suppression* dari *RTX Voice* berhasil menghapus suara *background*. Namun

setelah dihapusnya suara *background*, dapat dilihat banyak sinyal berwarna biru tua di sekitar sinyal warna oren. Hal ini menunjukkan bahwa suara pembicara tidak berhasil diperjelas melainkan hanya menghapus suara *background* dan meninggalkan hasil suara pembicara apa adanya. Oleh karena itu dari hasil suara dari rekaman dengan *RTX Voice* terdengar patah-patah dan menghilang-hilang karena suara tersebut tercampur dengan suara *background* dan gagal menghasilkan suara baru untuk memperjelas suara pembicara.

Dari hasil rekaman *Krisp*, sama seperti *RTX Voice* berhasil menghapus suara *background* yang dapat dilihat dari *spectrogram* di mana sinyal berwarna ungu tidak dapat ditemukan melainkan adanya sinyal berwarna hitam. Namun hasil dari rekaman dengan *Krisp* menghasilkan suara pembicara yang lebih jelas dibandingkan dengan *RTX Voice*. Dapat di lihat bahwa sinyal yang dihasilkan rekaman *Krisp* lebih konsisten dan lebih jelas, yang berarti suara pembicara lebih jelas dan tidak terpotong-potong, karena AI tersebut berhasil menghapus suara *background* tanpa mengganggu suara pembicara. Selain itu kita dapat melihat bahwa sinyal suara tersebut memiliki sinyal horisontal yang lebih tinggi dan konsisten. Hal ini karena selain memperjelas suara, *Krisp* juga mengubah *level* suara sehingga suara pembicara menjadi lebih keras dan memiliki volume yang konsisten.

Terakhir dapat dianalisis grafik frekuensi suara yang terdapat garis hijau yang menunjukkan frekuensi suara rekaman pertama, garis warna merah yang menunjukkan frekuensi suara rekaman setelah *RTX Voice* dan garis warna biru yang menunjukkan frekuensi suara rekaman setelah *Krisp*. Pada *axis-x* menunjukkan frekuensi suara antara 20-20000 Hz dan pada *axis-y* menunjukkan *level* (dB) suara. Pada grafik frekuensi suara dapat di lihat frekuensi sekitar 20-500 Hz terlihat garis frekuensi hijau cukup tinggi yang berarti suara pada frekuensi tersebut yaitu suara *sub-bass* dan *bass*

terdengar lebih jelas dibanding dengan garis merah dan biru yang memiliki level lebih rendah. Sedangkan pada frekuensi sekitar 500-6000 Hz dapat di lihat perubahan antara rekaman pertama dan setelah *RTX Voice* dan *Krisp* relatif kecil. Hal ini dikarenakan pada frekuensi ini suara yang muncul ada suara vokal atau suara berbicara seseorang dan suara konsonan. Karena tujuan dari *AI noise suppression* adalah untuk memperjelas suara pembicara, pada frekuensi ini tidak ada perubahan yang besar agar suara pembicara tidak berubah setelah di rubah. Pada Frekuensi 6000-20000 Hz terdapat perubahan paling besar dari semua. Karena suara manusia berbicara berada pada frekuensi 250-6000 Hz frekuensi di atas ini akan dianggap suara yang mengganggu kejelasan pembicaraan, oleh karena itu suara pada frekuensi ini akan diturunkan untuk mendapatkan kejelasan suara pembicara.

SIMPULAN

Berdasarkan hasil eksperimen perbandingan *AI noise suppression* yang telah dilakukan beberapa hal yang dihasilkan, antara lain:

1. Dengan menggunakan *AI noise suppression*, suara background dan suara berisik yang ditangkap mikrofon dapat dihapus sehingga menghasilkan suara pembicara menjadi lebih jelas.
2. Tidak semua *AI noise suppression* di buat sama. *RTX Voice* walau menggunakan kemampuan *AI* pada kartu grafis, hasil yang didapatkan tidak sebagus dengan *Krisp* menggunakan aplikasi dan tersambung pada *cloud* yang akan mencocokkan kondisi suara dengan model yang cocok untuk memaksimalkan kemampuan.
3. *Spectrogram* dapat menunjukkan karakteristik suara dan hal unik pada suara. Kita dapat melihat karakteristik suara rekaman awal yang sangat berisik yang menggambarkan suara pembicara tidak terlalu jelas karena gangguan suara lain. Suara setelah *RTX Voice* menjadi lebih tenang namun suara pembicara

terganggu karena *AI RTX Voice* tidak mampu memperjelas suara pembicara setelah menghapus suara lainnya. Dan suara setelah *Krisp* menjadi lebih tenang namun tidak seperti *RTX Voice* suara pembicara tetap jelas dan dioptimalkan sehingga suara menjadi lebih jelas dan keras.

4. Dengan frekuensi suara kita dapat menganalisis frekuensi suara yang ditargetkan oleh *AI noise suppression* untuk memperjelas suara pembicara.

DAFTAR PUSTAKA

- Braun, S., & Tashev, I. (2020). *Data augmentation and loss normalization for deep noise suppression*. <http://arxiv.org/abs/2008.06412>
- Chu, Y., Zhao, S., He, L., & Niu, F. (2022). Wind noise suppression in filtered-x least mean squares-based active noise control systems. *The Journal of the Acoustical Society of America*, 152(6), 3340–3345. <https://doi.org/10.1121/10.0016443>
- Ciric, D., Peric, Z., Nikolic, J., & Vucic, N. (2021, March 17). Audio Signal Mapping into Spectrogram-Based Images for Deep Learning Applications. 2021 20th International Symposium INFOTEH-JAHORINA, INFOTEH 2021 - Proceedings. <https://doi.org/10.1109/INFOTEH51037.2021.9400698>
- Fricke, L., Kuzmic, J., & Vatolkin, I. (2022). *Suppression of Background Noise in Speech Signals with Artificial Neural Networks, Exemplarily Applied to Keyboard Sounds*. 367–374. <https://doi.org/10.5220/0011537400003332>
- Hanik, N. R., & Harsono, S. (2020). Implementasi Model Pembelajaran Komparasi yang Diintegrasikan dengan Pendekatan Kolaboratif Ditinjau dari Kemampuan Analisis Mahasiswa. *Jurnal Komunikasi*

- Pendidikan*, 4(2).
<http://journal.univetbantara.ac.id/index.php/komdik>
- Mukherjee, A., & Jain, V. (2021). Geolocation Driven Reinforcement Learning-Powered In Geolocation Driven Reinforcement Learning-Powered In Transit Headset Noise Cancellation Mechanism Transit Headset Noise Cancellation Mechanism.
https://www.tdcommons.org/dpubs_series/4133
- Nurhasanah, A., Pribadi, R. A., & Nur, M. D. (2021). ANALISIS KURIKULUM 2013. *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, 7(02), 484–493.
<https://doi.org/10.36989/DIDAKTIK.V7I02.239>
- Qorib, A., & Zaniyati², H. S. (2021). Penggunaan Open Broadcast Software Studio Dalam Mendesain Video Pembelajaran Era Pandemi. *12*.
<https://doi.org/10.36835/syaikhuna.v12i1.4320>
- Rohmi Khalida, B., & Astawan, G. (2021). *Penerapan Metode Eksperimen untuk Meningkatkan Hasil Belajar IPA Siswa Kelas VI SD*. 4.
<https://doi.org/10.23887/jippg.v4i2>
- Siahaan, M., Harsana Jasa, C., Anderson, K., Rosiana, M. V., Lim, S., & Yudianto, W. (2020). Penerapan Artificial Intelligence (AI) Terhadap Seorang Penyandang Disabilitas Tunanetra. In *Journal of Information System and Technology* (Vol. 01).
- Stumpf, C., & Parker, A. (2021). Live-Streaming In Mechanical Engineering: An Innovative Approach To Remote Lab Delivery