

PENERAPAN *RETRIEVAL-BASED VOICE CONVERSION* UNTUK
MEMBANGUN ASET SUARA KARAKTER MONSTER PADA
GAME

Diajukan Sebagai Syarat Untuk Menyelesaikan
Pendidikan Program Strata-1 Pada
Jurusan Teknik Informatika



Oleh:

Hafiz Muhammad Kurniawan
NIM : 09021182126003

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2024

LEMBAR PENGESAHAN SKRIPSI

**PENERAPAN *RETRIEVAL-BASED VOICE CONVERSION* UNTUK
MEMBANGUN ASET SUARA KARAKTER MONSTER PADA GAME**


Oleh:

Hafiz Muhammad Kurniawan
NIM : 09021182126003

Indralaya, 1 Desember 2024



Pembimbing,



Anggina Primanita, M.IT., Ph.D.
NIP. 198908062015042002

TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI

Pada hari senin tanggal 30 Desember 2024 telah dilaksanakan ujian komprehensif skripsi oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya

Nama : Hafiz Muhammad Kurniawan

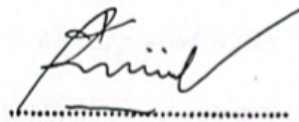
NIM : 09021182126003

Judul : Penerapan *Retrieval-Based Voice Conversion* untuk Membangun Aset Suara Monster pada *Game*

dan dinyatakan LULUS.

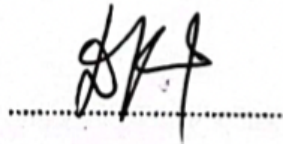
1. Ketua Penguji

Mastura Diana Marieska, M.T
NIP 198603212018032001



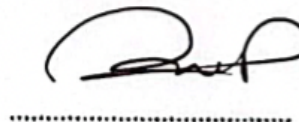
2. Penguji

Dian Palupi Rini, M.Kom., Ph.D
NIP 197802232006042002



3. Pembimbing

Anggina Primanita, M.IT., Ph.D.
NIM. 198908062015042002



Mengetahui,

Ketua Jurusan Teknik Informatika



Hadipurnawan Satria, Ph.D.
NIP. 198004182020121001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Hafiz Muhammad Kurniawan

NIM : 09021182126003

Program Studi : Teknik Informatika

Judul : Penerapan Retrieval-Based Voice Conversion untuk Membangun Aset Suara Karakter Monster pada Game

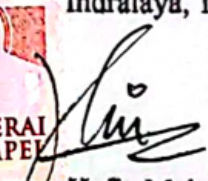
Hasil Pengecekan Software Turnitin : 6%

Menyatakan bahwa Skripsi saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam Skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari siapapun.



Indralaya, 1 Desember 2024


Hafiz Muhammad Kurniawan
NIM : 09021182126003

HALAMAN MOTTO DAN PERSEMBAHAN

*“Sungguh, Kami telah memberikan kepadamu **kemenangan yang nyata**. Agar Allah memberikan ampunan kepadamu (Muhammad) atas dosamu yang lalu dan yang akan datang serta menyempurnakan nikmat-Nya atasmu dan menunjukimu ke jalan yang lurus, dan agar Allah menolongmu dengan **pertolongan yang kuat (banyak)**”. (QS. Al-Fath: 1- 3)*

Kupersembahkan karya tulis ini kepada :

- Allah SWT
- Orang tuaku
- Saudara-saudaraku
- Keluarga besarku
- Sahabat dan teman seperjuangan
- Dosen pembimbing
- Almamater

ABSTRACT

Audio in video games plays a crucial role in creating an immersive and engaging gaming experience. However, indie game developers often face resource constraints, particularly in producing high-quality character voices. Voice cloning technology based on Retrieval-based Voice Conversion (RVC) offers an innovative solution by enabling accurate voice replication and transformation using minimal data. This technology provides flexibility for developers to create unique character voices without requiring direct voice actors, making it more cost- and time-efficient. This study aims to explore the application of RVC technology in generating monster character voices for video games, using a case study with voice clips from the anime *Goblin Slayer*. Through the voice conversion process, this study evaluates the performance of RVC models using objective standards such as Mel Cepstral Distortion (MCD) to measure voice transformation and Mel-Frequency Cepstral Coefficients (MFCC)-based analysis to assess the similarity between the original and converted voices. The results demonstrate that RVC technology is capable of producing voices significantly different from the original, enabling the creation of audio characters tailored to the narrative needs of the game. This technology offers an effective solution for indie game developers to create innovative and high-quality audio assets at low cost. This study highlights the importance of voice cloning technology in supporting competitive indie game development in the market.

Keywords : Game Audio, Voice Cloning, Retrieval-based Voice Conversion, Indie Game Developers, Mel Cepstral Distortion, Mel-Frequency Cepstral Coefficients.

ABSTRAKSI

Audio dalam gim video memegang peranan penting dalam menciptakan pengalaman bermain yang imersif dan mendalam. Namun, pengembang gim indie sering kali menghadapi keterbatasan sumber daya, terutama dalam menghasilkan suara karakter yang berkualitas. Teknologi kloning suara berbasis *Retrieval-based Voice Conversion* (RVC) menawarkan solusi inovatif dengan memungkinkan replikasi dan transformasi suara secara akurat menggunakan data yang minimal. Teknologi ini memberikan fleksibilitas bagi pengembang untuk menciptakan suara karakter unik tanpa memerlukan aktor suara langsung, sehingga lebih efisien dari segi biaya dan waktu. Penelitian ini bertujuan mengeksplorasi penerapan teknologi RVC dalam menghasilkan suara karakter monster untuk gim video, dengan studi kasus menggunakan cuplikan suara dari anime *Goblin Slayer*. Melalui proses konversi suara, penelitian ini mengevaluasi kinerja model RVC menggunakan standar objektif seperti *Mel Cepstral Distortion* (MCD) untuk mengukur tingkat perubahan suara, serta analisis berbasis *Mel-Frequency Cepstral Coefficients* (MFCC) untuk menentukan tingkat kesamaan antara suara asli dan suara hasil konversi. Hasil penelitian menunjukkan bahwa teknologi RVC mampu menghasilkan suara yang signifikan berbeda dari suara asli, memungkinkan terciptanya karakter audio yang sesuai dengan kebutuhan narasi gim. Teknologi ini memberikan solusi efektif bagi pengembang gim indie untuk menciptakan aset audio yang inovatif dan berkualitas tinggi dengan biaya rendah. Penelitian ini menegaskan pentingnya teknologi kloning suara dalam mendukung pengembangan gim indie yang kompetitif di pasar.

Kata Kunci : Audio Gim, Kloning Suara, *Retrieval-based Voice Conversion*, Pengembang Gim Indie, *Mel Cepstral Distortion*, *Mel-Frequency Cepstral Coefficients*.

KATA PENGANTAR

Puji dan syukur atas berkat dan rahmat Allah SWT yang telah memberikan hidayah, rahmat, dan petunjuk-Nya sehingga penulis dapat menyelesaikan tugas akhir ini, disusun guna memenuhi salah satu syarat penyelesaian pendidikan Strata-1 di Fakultas Ilmu Komputer Program Studi Teknik Informatika, Universitas Sriwijaya. Dalam penelitian dan penulisan skripsi ini, penulis tidak terlepas dari berbagai hambatan dan kesulitan yang pada dasarnya memberikan hikmah tersendiri bagi penulis.

Penulis menyadari bahwa penulisan ini tidak dapat terselesaikan tanpa dukungan dari berbagai pihak baik moril maupun material. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada:

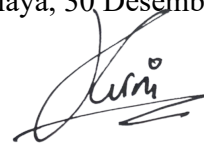
1. Allah Subhanahu wa Ta'ala yang telah memberikan rezeki berupa kesehatan, kecerdasan, kelancaran sehingga bisa menyelesaikan skripsi ini.
2. Nabi Muhammad Shallallahu 'Alaihi wa sallam yang telah memberikan jalan yang terang dan ilmu yang bermanfaat pada umatnya.
3. Orang tuaku, sebagai pengingat dan penyemangat dirumah sehingga penulis tetap semangat dalam menyelesaikan skripsi ini.
4. Saudara dan keluarga besar tercinta yang sentiasa ada memberikan dukungan penuh kepada penulis.
5. Dekan Fakultas Ilmu Komputer Universitas Sriwijaya Bapak Prof. DR. Erwin, S.Si., M.Si.

6. Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Hadipurnawan Satria, Ph.D.
7. Dosen Pembimbing Akademik, Bapak Julian Supardi, S.Pd., M.T., Ph.D.
8. Dosen Pembimbing Skripsi, Ibu Anggina Primanita, S.Kom., M.IT., Ph.D. yang telah menyediakan waktu, senantiasa memberikan bimbingan, masukan, perhatian, dan motivasi untuk menyelesaikan tugas akhir ini.
9. Seluruh Dosen Jurusan Teknik Informatika dan Dosen Fakultas Ilmu Komputer yang telah banyak memberikan ilmu selama masa perkuliahan.
10. Sahabat bimbingan skripsi yang senantiasa memberikan dukungan, inspirasi, dan menjadi tempat berdiskusi yang berarti bagi penulis.
11. Teman-teman seperjuanganku khususnya anggota grup Budak Corporate, Epan, Iman, Anhar, Dzaki, Apandi, Bima, Wahyu, Zidane dan Adit yang selalu memberi dukungan dan masukan selama penulisan skripsi ini.
12. Teman-teman penghuni ruang labotarium pengolahan citra kampus Palembang yang selalu memberikan dukungan, masukan dan gagasan kepada penulis.
13. Semua pihak yang terlibat dalam penyusunan skripsi yang tidak bisa disebutkan satu per satu namun telah berkontribusi dalam penyusunan skripsi.

Penulis sadar bahwa masih banyak kekurangan selama penyusunan skripsi dikarenakan kurangnya manajemen waktu yang baik serta pengalaman dan pengetahuan yang terbatas. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Semoga semua pihak yang telah membantu dalam penulisan skripsi ini mendapatkan pahala di sisi Allah Swt. dan penulis berharap semoga skripsi ini bermanfaat, khususnya bagi para akademisi. *Aamiin*.

Indralaya, 30 Desember 2024

A handwritten signature in black ink, appearing to read 'Kurniawan', with a stylized flourish underneath.

Hafiz Muhammad Kurniawan

DAFTAR ISI

| | Halaman |
|---|---------|
| HALAMAN JUDUL | i |
| HALAMAN PENGESAHAN SKRIPSI | ii |
| TANDA LULUS UJIAN KOMPREHENSIF SKRIPSI..... | iii |
| HALAMAN PERNYATAAN | iv |
| HALAMAN MOTTO DAN PERSEMBAHAN | v |
| ABSTRACT | vi |
| ABSTRAKSI..... | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI..... | xi |
| DAFTAR TABEL | xiv |
| DAFTAR GAMBAR..... | xv |
| DAFTAR LAMPIRAN..... | xvi |

BAB I PENDAHULUAN

| | | |
|-----|------------------------------|-----|
| 1.1 | Pendahuluan..... | I-1 |
| 1.2 | Latar Belakang Masalah | I-1 |
| 1.3 | Rumusan Masalah..... | I-3 |
| 1.4 | Tujuan Penelitian | I-4 |
| 1.5 | Manfaat Penelitian | I-4 |
| 1.6 | Batasan Masalah | I-5 |
| 1.7 | Sistematika Penulisan | I-5 |
| 1.8 | Kesimpulan | I-6 |

BAB II TINJAUAN PUSTAKA

| | | |
|-------|--|------|
| 2.1 | Pendahuluan..... | II-1 |
| 2.2 | Landasan Teori..... | II-1 |
| 2.2.1 | Audio pada permainan Gim | II-1 |
| 2.2.2 | <i>Retrieval-based Voice Conversion (RVC)</i> | II-3 |
| 2.2.3 | <i>Mel Frequency Cepstral Coefficient (MFCC)</i> | II-4 |
| 2.2.4 | <i>Mel-Cepstral Distance (MCD)</i> | II-5 |

| | | |
|-----|---|-------|
| 2.3 | <i>Multimedia Development Life Cycle (MDLC)</i> | II-7 |
| 2.4 | Penelitian Lain yang Relevan | II-9 |
| 2.5 | Kesimpulan | II-10 |

BAB III METODOLOGI PENELITIAN

| | | |
|-------|--|--------|
| 3.1 | Pendahuluan..... | III-1 |
| 3.2 | Pengumpulan Data..... | III-1 |
| 3.2.1 | Jenis dan Sumber Data | III-1 |
| 3.2.2 | Metode Pengumpulan Data | III-1 |
| 3.3 | Tahapan Penelitian..... | III-2 |
| 3.3.1 | Mengumpulkan Data | III-3 |
| 3.3.2 | Menentukan Kerangka Kerja Penelitian..... | III-3 |
| 3.3.3 | Menentukan Kriteria Pengujian..... | III-4 |
| 3.3.4 | Menentukan Format Data Pengujian | III-4 |
| 3.3.5 | Menentukan Alat Bantu Penelitian..... | III-5 |
| 3.3.6 | Menentukan Pengujian Penelitian | III-5 |
| 3.3.7 | Analisis Hasil Pengujian dan Kesimpulan Penelitian | III-6 |
| 3.3.8 | Metode Pengembangan Perangkat Lunak | III-7 |
| 3.4 | Metode Pengembangan Perangkat Lunak..... | III-7 |
| 3.4.1 | <i>Concept</i> | III-7 |
| 3.4.2 | <i>Design</i> | III-8 |
| 3.4.3 | <i>Material Collecting</i> | III-8 |
| 3.4.4 | <i>Assembly</i> | III-8 |
| 3.4.5 | <i>Testing</i> | III-8 |
| 3.4.6 | <i>Distribution</i> | III-9 |
| 3.5 | Manajemen Proyek Penelitian | III-9 |
| 3.6 | Kesimpulan | III-10 |

BAB IV PENGEMBANGAN PERANGKAT LUNAK

| | | |
|-------|----------------------------------|------|
| 4.1 | Pendahuluan..... | IV-1 |
| 4.2 | Fase-Fase Pengembangan | IV-1 |
| 4.2.1 | <i>Concept</i> (Konsep)..... | IV-1 |
| 4.2.2 | <i>Design</i> (Perancangan)..... | IV-2 |

| | | |
|---------|---|-------|
| 4.2.2.1 | <i>Use Case Diagram</i> | IV-2 |
| 4.2.2.2 | <i>Activity Diagram</i> | IV-3 |
| 4.2.2.3 | <i>Class Diagram</i> | IV-6 |
| 4.2.2.4 | <i>Sequence Diagram</i> | IV-6 |
| 4.2.2.5 | Sketsa Antarmuka Perangkat Lunak..... | IV-9 |
| 4.2.3 | <i>Material Collecting</i> (Pengumpulan Materi) | IV-10 |
| 4.2.4 | <i>Assembly</i> | IV-11 |
| 4.2.4.1 | Antarmuka Aplikasi..... | IV-11 |
| 4.3 | Kesimpulan | IV-13 |

BAB V HASIL DAN ANALISIS PENELITIAN

| | | |
|---------|--|------|
| 5.1 | Pendahuluan..... | V-1 |
| 5.2 | Hasil Penelitian | V-1 |
| 5.2.1 | Konfigurasi Pengujian | V-1 |
| 5.2.2 | Evaluasi Pengujian (Testing)..... | V-2 |
| 5.2.2.1 | <i>Conversational Speech</i> | V-2 |
| 5.2.2.2 | <i>Emotional Expressions</i> | V-3 |
| 5.2.2.3 | <i>Phonetically-Rich Sentences</i> | V-3 |
| 5.2.2.4 | <i>Monster</i> | V-3 |
| 5.2.3 | Data Hasil Penelitian | V-4 |
| 5.3 | Analisis Hasil Pengujian Model | V-6 |
| 5.4 | Kesimpulan | V-16 |

BAB VI KESIMPULAN DAN SARAN

| | | |
|-----|-------------------|------|
| 6.1 | Pendahuluan | VI-1 |
| 6.2 | Kesimpulan | VI-1 |
| 6.3 | Saran..... | VI-2 |

| | |
|---------------------|-----|
| DAFTAR PUSTAKA..... | xv |
| LAMPIRAN | xix |

DAFTAR TABEL

| | Halaman |
|---|---------|
| Tabel III-1. Rancangan Perbandingan Suara Asli dan Suara Hasil RVC..... | III-4 |
| Tabel III-2. Rancangan Analisis Pengujian Model Terbaik | III-6 |
| Tabel III- 3. Format Nilai Score Model | III-7 |
| Tabel III- 4. Managemen waktu proyek penelitian..... | III-9 |
| Tabel IV- 1. Aktor Use Case Diagram | IV-3 |
| Tabel IV- 2 Definisi Use Case | IV-3 |
| Tabel V- 1. Hasil Perbandingan pada Sound-Similar | V-4 |
| Tabel V- 2. Hasil Perbandingan pada FastDTW | V-5 |
| Tabel V- 3. Hasil Perbandingan pada MCD..... | V-5 |
| Tabel V- 4. Hasil Pengujian Model Terbaik..... | V-14 |
| Tabel V- 5. Nilai Score Model | V-15 |

DAFTAR GAMBAR

| | Halaman |
|---|---------|
| Gambar II- 2. Multimedia Development Life Cycle | II-7 |
| Gambar III- 1. Tahapan Penelitian..... | III-2 |
| Gambar III- 2. Kerangka Kerja Penelitian..... | III-3 |
| Gambar IV- 1 Use Case Diagram | IV-2 |
| Gambar IV- 2. <i>Activity</i> Diagram Mulai Konversi Suara..... | IV-4 |
| Gambar IV- 3. <i>Activity</i> Diagram Mulai <i>Tranning RVC Model</i> | IV-4 |
| Gambar IV- 4. <i>Activity</i> Diagram Mulai Mengubah <i>Format Audio File</i> | IV-5 |
| Gambar IV- 5. <i>Activity</i> Diagram Tampilan <i>User Guide</i> | IV-5 |
| Gambar IV- 6. Rancangan <i>Class</i> Diagram..... | IV-6 |
| Gambar IV- 7. <i>Sequence</i> Diagram Konversi Suara | IV-7 |
| Gambar IV- 8. <i>Sequence</i> Diagram Training Model | IV-7 |
| Gambar IV- 9. <i>Sequence</i> Diagram <i>Conversion Audio File</i> | IV-8 |
| Gambar IV- 10. <i>Sequence</i> Diagram <i>User Guide</i> | IV-8 |
| Gambar IV- 11. <i>Storyboard Model Conversion Display</i> | IV-9 |
| Gambar IV- 12. <i>Storyboard Training Model Display</i> | IV-9 |
| Gambar IV- 13. <i>Storyboard Conversion Audio File Display</i> | IV-10 |
| Gambar IV- 14. <i>Storyboard User Guide Display</i> | IV-10 |
| Gambar IV- 15. Antarmuka <i>Model Conversion</i> | IV-11 |
| Gambar IV- 16. Antarmuka <i>Training Model</i> | IV-12 |
| Gambar IV- 17. Antarmuka <i>Conversion Audio File</i> | IV-12 |
| Gambar IV- 18. Antarmuka <i>User Guide</i> | IV-13 |
| Gambar V - 1. Grafik Batang Perbandingan pada Sound-Similar | V-6 |
| Gambar V - 2. Grafik Batang Perbandingan pada FastDTW | V-7 |
| Gambar V - 3. Grafik Batang Perbandingan pada MCD | V-8 |
| Gambar V - 4. Grafik Evaluasi pada Jenis Suara <i>Conversional Speech</i> | V-10 |
| Gambar V - 5. Grafik Evaluasi pada Jenis Suara <i>Emotional Expressions</i> | V-11 |
| Gambar V - 6. Grafik Evaluasi pada Suara <i>Phonetically-Rich Sentences</i> | V-12 |
| Gambar V - 7. Grafik Evaluasi pada Jenis Suara Monster | V-13 |

DAFTAR LAMPIRAN

| | | |
|-------------------|--------------------------------|-------|
| LAMPIRAN 1 | Dataset Suara | xviii |
| LAMPIRAN 2 | Kode Program | xviii |
| LAMPIRAN 3 | Visual Bentuk Suara | xviii |
| LAMPIRAN 4 | Evaluasi Kemiripan Suara | xviii |

BAB I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan judul yang diangkat secara rinci. Bab ini menjelaskan tentang penjelasan secara umum terhadap penelitian yang dikerjakan.

1.2 Latar Belakang Masalah

Pada saat ini *Video Game* atau gim video telah menjadi salah satu bentuk hiburan yang paling mendominasi di era digital saat ini, menawarkan pengalaman yang mendalam dan interaktif yang tidak ditemukan dalam bentuk hiburan lainnya (Sinclair, 2020). Kenaikan popularitas gim video telah menggeser preferensi masyarakat, di mana banyak individu, baik anak-anak maupun orang dewasa, kini lebih memilih bermain gim daripada menonton film (Bandral dan Kaur, 2018). Gim video memungkinkan pemain untuk terlibat secara langsung dalam dunia virtual, mengendalikan narasi dan aksi melalui berbagai input seperti tombol atau gerakan (Bandral dan Kaur, 2018). Hal ini menciptakan pengalaman yang imersif, hasil dari kombinasi harmonis antara elemen visual dan audio yang dirancang untuk membawa pemain ke dalam dunia yang interaktif dan responsif.

Dalam konteks pembuatan gim, audio merupakan elemen yang sangat penting. Audio dalam gim video tidak hanya berfungsi sebagai pelengkap visual, tetapi juga berperan dalam membangun atmosfer, mendukung narasi, dan

meningkatkan keterlibatan pemain (Sinclair, 2020). Elemen audio dalam gim dapat dibagi menjadi tiga kategori utama antara lain musik latar, suara suasana dan efek suara (Horowitz dan Looney, 2014). Setiap elemen ini memiliki peran yang signifikan dalam menciptakan pengalaman bermain yang imersif dan memuaskan. Efek suara, khususnya dianggap sebagai salah satu elemen audio yang paling berpengaruh dalam meningkatkan kualitas pengalaman bermain gim (Andersen, et.al 2021). Seorang pengembang gim yang bekerja dengan sumber daya terbatas, perlu memahami dan memanfaatkan elemen audio ini untuk menciptakan gim yang berkualitas tinggi.

Seiring berkembangnya teknologi informasi, salah satu inovasi yang dapat digunakan dalam industri gim adalah *voice cloning* atau kloning suara. Teknologi *voice cloning* ini memungkinkan replikasi suara seseorang secara digital, memberikan fleksibilitas bagi pengembang dalam menghasilkan suara karakter yang unik dan autentik tanpa harus melibatkan aktor suara secara langsung (Jung dan Kim, 2020). Penelitian awal mengenai *voice cloning* dimulai dengan pendekatan *text-to-speech* yang dikembangkan dengan metode Tacotron untuk menghasilkan suara sintesis (Wang et al., 2017). Dengan kemajuan dalam kecerdasan buatan, seperti pada teknologi *Retrieval based Voice Conversion (RVC)*, pengembang kini dapat melakukan konversi suara dengan akurasi tinggi dan sumber daya minimal (Ren, 2024). RVC memungkinkan perubahan timbre suara dengan menggunakan data yang relatif kecil, menjadikannya solusi ideal bagi pengembang gim individu (*indie*) yang mungkin memiliki keterbatasan dalam hal anggaran dan waktu.

Dengan memanfaatkan teknologi *voice cloning*, pengembang gim *indie* dapat menciptakan suara karakter yang lebih kaya dan beragam, meningkatkan kualitas narasi dan keterlibatan pemain dalam gim. *Voice cloning* tidak hanya mempermudah proses pembuatan audio gim, tetapi juga membuka peluang bagi pengembang untuk meningkatkan pengalaman audio secara keseluruhan dan melibatkan *audiens* dengan cara yang baru dan lebih imersif (Matthew et al., 2024). Oleh karena itu, teknologi ini memiliki potensi besar untuk mendukung pengembangan gim *indie* yang lebih inovatif dan kompetitif di pasar yang semakin berkembang.

Mengacu pada permasalahan dan argumen yang telah dipaparkan, pemanfaatan teknik RVC sebagai alat bantu bagi pengembang gim *indie* dalam pembuatan aset suara karakter dapat meningkatkan pengalaman pembuatan gim. Oleh karena itu, penulis akan melakukan penelitian tentang Penerapan *Retrieval-based Voice Conversion* untuk Membangun Aset Suara Karakter Monster pada *Game*.

1.3 Rumusan Masalah

Beberapa rumusan masalah berdasarkan latar belakang yang telah disebutkan yang akan dibahas dalam penelitian ini adalah:

1. Bagaimana penerapan teknologi *Retrieval-based Voice Conversion* (RVC) dalam pembuatan aset suara karakter monster pada gim?
2. Seberapa konsisten RVC dalam menciptakan variasi suara karakter monster?

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian yang dilakukan adalah sebagai berikut:

1. Menerapkan teknologi *Retrieval-based Voice Conversion* (RVC) untuk membuat aset suara karakter monster dalam gim.
2. Mengevaluasi konsistensi RVC dalam menciptakan variasi suara karakter monster.

1.5 Manfaat Penelitian

Diharapkan perpada penelitian yang dijalankan dapat memberi manfaat sebagai berikut:

1. Hasil penelitian ini diharapkan dapat membantu *developer game indie* dalam menciptakan aset suara karakter monster yang autentik dan imersif dengan menggunakan teknologi *Retrieval-based Voice Conversion* (RVC).
2. Menambah literatur dan kontribusi dalam bidang pengembangan audio gim dan *voice cloning*, khususnya dalam penerapan teknologi RVC pada pembuatan aset suara untuk gim.
3. Hasil penelitian ini dapat digunakan sebagai referensi untuk pengembangan lebih lanjut dalam integrasi teknologi suara di industri gim dan sebagai landasan untuk penelitian selanjutnya.

1.6 Batasan Masalah

Agar ruang lingkup masalah pada penelitian ini jelas dan terhindar dari penyimpangan, batasan masalah dalam penelitian ini ditetapkan pada suara yang dihasilkan oleh teknologi *Retrieval-based Voice Conversion* RVC. Suara yang dihasilkan akan lebih optimal dengan durasi 2 detik untuk setiap karakter monster dan suara yang dihasilkan adalah satu jenis suara karakter monster yaitu suara goblin. Pembatasan ini diterapkan agar proses konversi suara lebih efisien dan relevan dengan kebutuhan pembuatan aset suara dalam pengembangan gim.

1.7 Sistematika Penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut:

BAB I. PENDAHULUAN

Pada bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, sistematika penulisan dan kesimpulan dari judul yang diangkat.

BAB II. KAJIAN LITERATUR

Pada bab ini membahas tentang dasar-dasar teori yang digunakan dalam penelitian yaitu teori *Voice Conversion*, algoritma *Retrieval-based*, komponen suara serta penelitian-penelitian lain yang relevan.

BAB III. METODOLOGI PENELITIAN

Bab ini akan membahas tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci

dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

BAB IV. PENGEMBANGAN PERANGKAT LUNAK

Bab ini akan membahas proses pengembangan perangkat lunak *Voice Conversion* menggunakan algoritma *Retreival-base* dan implementasi dari tiap fase berdasarkan metode *Multimedia Develompent Life Cycle*.

BAB V. HASIL DAN ANALISIS PENELITIAN

Pada bab ini akan dijelaskan mengenai proses pengujian berdasarkan format yang sudahdirancangkan dari perangkat lunak serta analisis terhadap hasil penelitian.

BAB VI. KESIMPULAN

Pada bab ini akan dijelaskan mengenai kesimpulan dari penelitian dan pengujian yang dilakukan beserta saran yang diharapkan untuk digunakan sebagai referensi dalam pengembangan penelitian selanjutnya.

1.8 Kesimpulan

Berdasarkan rumusan masalah yang diteliti, tujuan penelitian, dan batasan masalah yang diteliti dapat ditetapkan berdasarkan latar belakang masalah. Judul penelitian yang dilakukan adalah “Penerapan *Retrieval-based Voice Conversion* untuk Membangun Aset Suara Karakter Monster pada *Game*”.

DAFTAR PUSTAKA

- Abdul, Z. K., & Al-Talabani, A. K. (2022). Mel Frequency Cepstral Coefficient and its Applications: A Review. *IEEE Access*, *10*(November), 122136–122158. <https://doi.org/10.1109/ACCESS.2022.3223444>
- Alhumud, A. M., Al-qurishi, M., Alomar, Y. O., Alzahrani, A., & Soussi, R. (2024). Improving Automated Speech Recognition Using Retrieval-Based Voice Conversion. *Tiny Papers @ ICLR, 2022*, 1–7. <https://doi.org/271461600>
- Andersen, F., Danny, King, C. L., & Gunawan, A. A. S. (2021a). Audio Influence on Game Atmosphere during Various Game Events. *Procedia Computer Science*, *179*, 222–231. <https://doi.org/10.1016/j.procs.2021.01.001>
- Andersen, F., Danny, King, C. L., & Gunawan, A. A. S. (2021b). Audio Influence on Game Atmosphere during Various Game Events. *Procedia Computer Science*, *179*(2019), 222–231. <https://doi.org/10.1016/j.procs.2021.01.001>
- Bandral, N., & Kaur, R. (2018). A Psychological Inquiry into the Role of Music in Video Games. *Language in India*, *18*(5), 297.
- Binanto, I. (2010). *Multimedia Digital - Dasar Teori dan Pengembangannya* (Issue 25).
- Farkaš, T. (2024). Understanding Auditory Space in Digital Games for Visually Impaired People. *Acta Ludologica*, *7*(1), 136–150. <https://doi.org/10.34135/actaludologica.2024-7-1.136-150>

- Flipsen, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics and Phonetics*, 20(4), 303–312.
<https://doi.org/10.1080/02699200400024863>
- Horowitz, S., & Looney, S. (2014). *The Essential Guide to Game Audio*. Routledge.
<https://doi.org/10.4324/9781315886794>
- Jung, S., & Kim, H. (2020). *Neural voice cloning with a few low-quality samples*. *NeurIPS*, 1–11. <http://arxiv.org/abs/2006.06940>
- Kamble, A., Tathe, A., Kumbharkar, S., Bhandare, A., & Mitra, A. C. (2023). *Custom Data Augmentation for low resource ASR using Bark and Retrieval-Based Voice Conversion*. <http://arxiv.org/abs/2311.14836>
- Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional Expression: Advances in Basic Emotion Theory. In *Journal of Nonverbal Behavior* (Vol. 43, Issue 2, pp. 133–160). Springer New York LLC.
<https://doi.org/10.1007/s10919-019-00293-3>
- Kominek, J., Schultz, T., & Black, A. W. (2008). Synthesizer Voice Quality of New Languages Calibrated With Mean Mel Cepstral Distortion. *SLTU 2008 - 1st International Workshop on Spoken Languages Technologies for Under-Resourced Languages*, 63–68.
- Kubichek, R. F. (1993). Mel-Cepstral distance measure for objective speech quality assessment. *IEEE Pac Rim Conf Commun Comput Signal Process*, 125–128.
<https://doi.org/10.1109/pacrim.1993.407206>

- Masriah, Atmojo, W. T., & Ayunda, A. T. (2023). Penerapan metode multimedia development life cycle dalam pembuatan aplikasi wisata berbasis mobile. *Prosiding Sains Nasional Dan Teknologi*, 13(1), 173. <https://doi.org/10.36499/psnst.v13i1.9133>
- Nercessian, S. (2020). Zero-shot singing voice conversion. *Proceedings of the 21st International Society for Music Information Retrieval Conference, ISMIR 2020*, 70–76.
- Onuh Matthew Ijiga, Idoko Peter Idoko, Lawrence Anebi Enyejo, Omachile Akoh, Solomon Ileanaju Ugbane, & Akan Ime Ibokette. (2024). Harmonizing the voices of AI: Exploring generative music models, voice cloning, and voice transfer for creative expression. *World Journal of Advanced Engineering Technology and Sciences*, 11(1), 372–394. <https://doi.org/10.30574/wjaets.2024.11.1.0072>
- Putrasyah, A., Magister, J., Komputer, I., Komputer, F. I., & Sriwijaya, U. (2023). *Classification of Recorded Voice Similarity Using Mel-Frequency*. 6, 959–969.
- Raza, A. A., Hussain, S., Sarfraz, H., Ullah, I., & Sarfraz, Z. (2009). Design and development of phonetically rich Urdu speech corpus. *2009 Oriental COCOSDA International Conference on Speech Database and Assessments*, 38–43. <https://doi.org/10.1109/ICSDA.2009.5278380>
- Ren, Z. (2024). *Selection of Optimal Solution for Example and Model of Retrieval Based Voice Conversion* (pp. 468–475). https://doi.org/10.2991/978-94-6463-370-2_48

- Rovithis, E., Moustakas, N., Floros, A., & Vogklis, K. (2019). Audio legends: Investigating sonic interaction in an augmented reality audio game. *Multimodal Technologies and Interaction*, 3(4), 1–18. <https://doi.org/10.3390/mti3040073>
- Sinclair, J. (2020). *Principles of Game Audio*. Routledge. <https://www.routledge.com/Principles-of-Game-Audio-and-Sound-Design-Sound-Design-and-Audio-Implementation-for-Interactive-and-Immersive-Media/Sinclair/p/book/9781138738973>
- Vasilijević, A., & Petrinović, D. (2011). Perceptual significance of cepstral distortion measures in digital speech processing. *Automatika*, 52(2), 132–146. <https://doi.org/10.1080/00051144.2011.11828412>
- Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., Yang, Z., Xiao, Y., Chen, Z., Bengio, S., Le, Q., Agiomyrgiannakis, Y., Clark, R., & Saurous, R. A. (2017). Tacotron: Towards End-to-End Speech Synthesis. *Interspeech 2017*, 4006–4010. <https://doi.org/10.21437/Interspeech.2017-1452>