

Sintesis Suara Bernyanyi Dengan Teknologi *Text-To-Speech* untuk Notasi Musik Angka dan Lirik Lagu Berbahasa Indonesia

Jonathan*¹, Yohanes Suyanto²

¹Prodi Elektronika dan Instrumentasi FMIPA, UGM Yogyakarta, Indonesia

²Departemen Ilmu Komputer dan Elektronika FMIPA, UGM Yogyakarta, Indonesia

e-mail: *jonathan11@mail.ugm.ac.id, yanto@ugm.ac.id

Abstrak

Bernyanyi merupakan aktivitas seni yang secara umum tidak dapat dilepaskan dari kehidupan manusia. Hal ini menjadikan penelitian dengan melibatkan teknologi untuk mengembangkan seni bernyanyi dapat bermanfaat bagi lingkup yang luas. Penelitian ini bermaksud untuk melakukan sintesis suara bernyanyi dengan memanfaatkan teknologi TTS (*text-to-speech*) yang memiliki kapabilitas untuk menghasilkan suara dengan pelafalan kata tertentu pada frekuensi tertentu. Pada penelitian ini, masukan yang digunakan untuk sistem berupa teks lagu dengan format TXT yang mewakili notasi musik angka dan lirik lagu berbahasa Indonesia. Masukan akan melalui proses pemrosesan teks yang mengubah teks lagu menjadi transkripsi fonem, dan proses sintesis suara yang melakukan sintesis suara berdasarkan transkripsi fonem tersebut. Secara umum sistem berhasil melakukan sintesis suara bernyanyi dengan sejumlah fitur berdasarkan aturan notasi musik angka. Berdasarkan pengujian berupa aktivitas mendengarkan hasil suara bernyanyi sintesis dari 5 buah teks lagu oleh 30 orang responden, ditemui bahwa hasil sintesis suara bernyanyi memiliki akurasi sebesar 81.71% dengan standar deviasi 6.24%. Aturan penulisan teks lagu kemudian bisa dikatakan *user-friendly* dengan jumlah kompilasi ulang paling banyak untuk menuliskan sintaks tanpa error adalah 3 kali pada pengujian penulisan 8 bar teks lagu uji oleh 30 orang responden.

Kata kunci—sintesis suara bernyanyi, *text-to-speech*, notasi musik angka, bahasa Indonesia

Abstract

Singing is a work of art that can not be separated from human life. It then makes a research about develop the art of singing by technology will brings a useful impact for such a wide aspect of human life. This research is trying to synthesize singing voice with TTS (*text-to-speech*) technology, as it capability to produce sound with certain pronunciation at certain frequency of sound. Inputs that used in the system are texts of song in TXT format that contain the information of numbered musical notation and lyrics in Indonesian. These inputs will converted to a phonetic transcription, for then synthesize of song voice can done based on the transcription. In general, the system made successfully synthesize song voices with some feature that based on the convention of numbered musical notation. Based on 30 people of respondents, the song voice synthesized has 81.71% of accuracy with 6.24% of deviation standard. The syntax of song text also reputed as a *user-friendly* convention with only up to 3 times re-compilation done to synthesize 8 bar of song text by each of respondents without any error.

Keywords—singing voice synthesize, *text-to-speech*, numbered musical notation, Indonesian

1. PENDAHULUAN

Bernyanyi merupakan aktivitas seni yang dapat berwujud sebagai sarana untuk kegiatan belajar-mengajar, hiburan, dan ritual pada agama tertentu. Mengembangkan teknologi yang dapat menunjang seni seperti pada penelitian [1], [2], [3] dan [4] kemudian bisa dikatakan sebagai hal yang dapat berguna pada lingkup yang luas dari aspek kehidupan manusia. Sebagaimana dilakukan pada penelitian [4], teknologi TTS (*text-to-speech*) merupakan salah satu contoh teknologi yang berpotensi untuk menunjang seni bernyanyi.

TTS adalah sistem yang mengkonversi teks ke dalam bentuk ucapan. Sebagaimana dilakukan pada penelitian [5], TTS dapat dipadukan dengan OCR (*optical character recognition*) atau komponen lainnya untuk menunjang suatu sistem yang menjalankan fungsi membaca untuk membantu pekerjaan manusia. Jika diperhatikan lebih jauh, dapat ditemukan bahwa teknologi TTS mampu memenuhi komponen-komponen dasar yang diperlukan untuk melakukan sintesis suara bernyanyi. Komponen-komponen dasar yang dimaksud adalah nada (atau frekuensi gelombang suara) dan interpretasi lirik (yang dibentuk dari pengolahan teks dan fonemnya). Penyesuaian fonem sesuai lirik lagu, tinggi-rendah frekuensi sesuai nada pada lagu, dan durasi setiap fonem sesuai ritmik (ketukan pada tempo tertentu) pada lagu, dapat digunakan sebagai dasar untuk melakukan sintesis suara bernyanyi. Pengolahan lirik dan nada (bagian *front-end* pada sistem TTS) dapat dilakukan dengan membuat sistem pengolahan teks yang mengolah lirik dan nada dengan teknik *regular expression* pada aplikasi Lex (*Lexical Analyzer Generator*) untuk proses tokenisasi, yang kemudian dilanjutkan dengan pemrosesan dengan bahasa tingkat tinggi seperti bahasa C. Penggunaan Lex sebagai aplikasi tingkat rendah memungkinkan untuk membuat sistem penerjemah teks dengan pelafalan bahasa Indonesia. Proses konversi dari fonem ke dalam bentuk suara (bagian *back-end* pada sistem TTS) kemudian bisa dilakukan dengan metode penyambungan *diphone*, sehingga dengan basis data yang relatif sedikit bisa dihasilkan suara ucapan yang beragam.

Penelitian [4] dan [6] merupakan contoh dari penelitian yang berupaya melakukan sintesis berdasarkan aturan dari bahasa tertentu. Penelitian [4] melakukan sintesis suara bernyanyi dalam bahasa Mandarin. Metode yang digunakan pada penelitian ini adalah *unit-selection synthesis* dengan mengambil ukuran sintesis per *demi-syllable* (setengah silabel) untuk menghindari jumlah kombinasi fonem yang terlalu besar, dan tidak menggunakan unit berukuran *diphone* untuk menghindari pencampuran dari unit-unit yang berdekatan (yang cukup banyak ditemui pada bahasa Mandarin). Pembuatan pustaka kemudian dilakukan dengan melakukan rekaman suara wanita yang menyanyikan 44 lagu berbahasa Mandarin, untuk mendapatkan 5882 silabel dengan $F_s = 22,050$ Hz. Penelitian [6] kemudian menunjukkan berbagai permasalahan yang berbeda untuk konteks pelafalan bahasa Melayu.

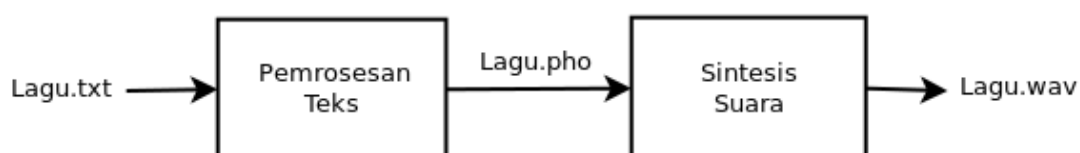
Penelitian [7] mengembangkan sistem TTS untuk bahasa Indonesia. Penelitian sintesis suara ucapan ini berfokus pada dinamika intonasi (*prosody*) pada nada bicara yang dapat merepresentasikan emosi pada kata-kata yang diucapkan. Masukan sistem yang menggunakan bahasa Indonesia menjadikan proses perubahan teks menjadi fonem memperhatikan aturan umum pelafalan dalam bahasa Indonesia. Aturan kemudian dikonversi ke struktur *context-free grammar* (CFG) dengan notasi *Backus-Naur Form* (BNF). Proses sintesis kemudian dilakukan dengan menggunakan aplikasi Praat untuk mendapatkan nilai F_0 dari suatu sampel suara, dan aplikasi MBROLA untuk melakukan konversi fonem dan *prosody*-nya (dalam *file* dengan ekstensi *.pho*) ke bentuk suara ucapan.

Sistem sintesis suara bernyanyi sederhana yang dibuat dalam penelitian ini diharapkan mampu berguna dalam perkembangan penelitian TTS secara mendasar. Penggunaan aplikasi-aplikasi tingkat rendah seperti Lex memungkinkan peneliti dan pengembang untuk mampu berkreasi lebih leluasa, misalnya mengembangkan sistem untuk konversi kata-kata berbahasa Indonesia ke dalam bentuk fonem sesuai aturan pelafalan pada [8]. Melihat potensi dan

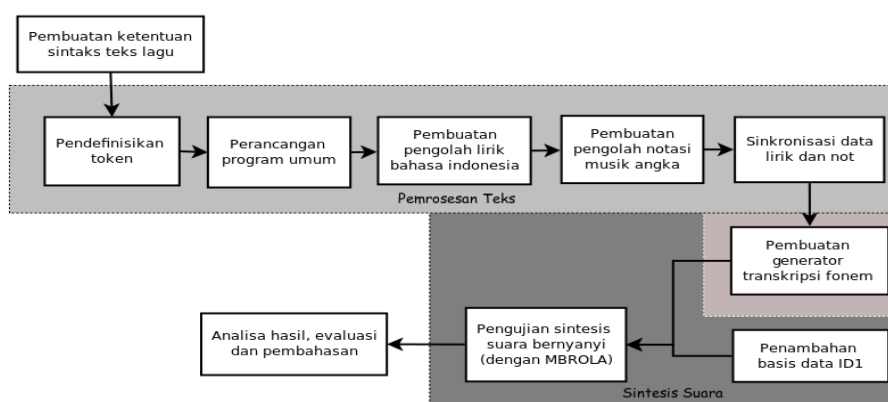
kebutuhan yang ada (seperti pada sarana belajar musik, pengembangan teknologi untuk seni musik, dan potensi lainnya), menjadi menarik untuk mengembangkan topik semacam ini.

2. METODE PENELITIAN

Rancangan umum dari sistem yang digunakan pada penelitian ini berupa sistem sintesis suara bernyanyi dengan 2 bagian utama: pemrosesan teks dan sintesis suara. Diagram blok dari rancangan sistem bisa dilihat pada Gambar 1. Penjabaran dari proses perancangan sistem pada Gambar 1 bisa dilihat pada Gambar 2. Pada diagram tersebut, blok-blok kegiatan yang akan dilakukan pada penelitian dikelompokkan menjadi kelompok pembuatan sistem pemrosesan teks dan kelompok pembuatan sistem sintesis suara.



Gambar 1 Diagram blok rancangan sistem



Gambar 2 Diagram blok tahapan pembuatan sistem

Masukan yang digunakan pada penelitian ini berupa teks notasi musik angka dan lirik berbahasa Indonesia. Masukan dibuat dalam format TXT dengan format khusus untuk komponen utama seperti lirik, not, nada dasar dan tempo. Komponen pendukung (seperti judul atau catatan tambahan) bisa dituliskan di luar format khusus sehingga datanya tidak akan diproses oleh sistem. Format khusus yang digunakan adalah sebagai berikut:

- a. Teks yang merupakan lirik lagu berada diantara tanda <L dan >L
- b. Teks yang merupakan notasi musik angka berada diantara tanda <N dan >N
- c. Tanda nada dasar, tempo, dan unit pecacahan waktu menggunakan tanda khusus
- d. Baris teks diluar lirik, not, dan tanda khusus dimulai dengan tanda <<

Teks lirik pada file TXT kemudian perlu mengikuti aturan yang dibuat dalam penelitian ini. Aturan yang dibuat pada penelitian ini mengacu pada aturan terkait ejaan untuk bahasa Indonesia sebagaimana pada [8]. Aturan untuk penulisan lirik antara lain:

- a. Setiap suku kata yang mewakili sebuah nada harus dipenggal dengan karakter *hyphen* (“-”). Jika ada 2 suku kata mewakili satu not yang sama, maka tidak dipenggal. Contoh: In-do-ne-sia

- b. Penulisan untuk huruf diakritik (ê) yang dilafalkan [ə] dituliskan dengan huruf ‘e’ diikuti tanda “^”. Contoh: Se[^]-nin se[^]-la-sa

Penulisan notasi musik angka pada kemudian menggunakan sintaks yang disederhanakan. Penyederhanaan bertujuan untuk memungkinkan notasi musik angka dituliskan dalam format TXT. Sintaks penyederhanaan notasi musik angka menggunakan sejumlah karakter sebagaimana ditunjukkan pada Tabel 1. Aturan penulisan ini sebagian besar aturannya mengadopsi dari hasil penelitian [9] dengan sejumlah perubahan kecil untuk melakukan penyesuaian dengan aturan penulisan lirik. Gambar 3 menunjukkan contoh penulisan teks lagu menggunakan sintaks penyederhanaan untuk format TXT dari teks musik konvensional.

Program menyediakan pemeriksa *error* dalam sintaks penulisan teks lagu format TXT. Kesalahan yang dibaca berupa tidak lengkapnya jumlah ketukan dalam 1 bar atau adanya penulisan yang tidak dikenali (menyalahi aturan sintaks). Pemeriksaan ini dilakukan dengan fitur menentukan unit pecahan waktu yang bisa diisi dengan nilai 2/4, 3/4, 4/4, 6/8, 9/8, dan 12/8, dan akan menentukan berapa ketuk yang harus ada pada setiap bar-nya.

Tabel 1 Sintaks penyederhanaan notasi musik angka untuk *file* teks

Karakter	Fungsi	Contoh
0-7	Sebagai not angka	3 5 2 3
	Sebagai batas bar	3 . . . 2 . (45) 3
.	Tanda birama untuk meneruskan ketukan not	3 . 4 .
()	Tanda birama untuk membagi ketukan not menjadi setengahnya	(35) 5 (.2) 3
(t)	Tanda birama untuk membagi ketukan not menjadi sepertiganya	3 (t 234) 5 7
[]	Tanda untuk menyanyikan beberapa not dalam 1 buah silabel	3 [4 5 6]
,	Tanda turun 1 oktaf	7, 1 2 3
`	Tanda naik 1 oktaf	6 7 1' 3'
_	Tanda naik setengah nada (kromatik)	3 4 4/ 5
}	Sebagai bar penutup	4 5 3 . }
#D=	Menentukan kunci nada dasar	#D= G
#MM=	Menentukan nilai tempo (satuan waktu setiap ketukan nada dengan satuan bpm atau <i>beat per minute</i>)	#MM = 120 .
#B=	Menentukan nilai unit pecahan waktu (ketentuan jumlah ketukan pada setiap bar)	#B = 4/4

Sintesis suara bernyanyi pada akhirnya dapat dilakukan dengan memberi masukan *file* PHO bersama *diphone inventory* ID1 (basis data rekaman *diphone* untuk bahasa Indonesia) pada aplikasi MBROLA. Aplikasi MBROLA digunakan untuk melakukan implementasi sintesis suara dengan metode penyambungan *diphone*. Sebagaimana hasil penelitian [10], metode penyambungan *diphone* memiliki tingkat kealamian kejelasan yang relatif baik. Perintah untuk aplikasi MBROLA dapat dilakukan dengan membuat *script* untuk mengakses *command line*

dengan perintah `mbrola diphone_database phonetic_file output_file`. Keluaran yang didapat pada tahap ini adalah suara bernyanyi dengan format WAV.

3. HASIL DAN PEMBAHASAN

Pengujian terhadap hasil sistem akan dilakukan dengan terlebih dahulu mensintesis suara bernyanyi dari 4 teks lagu sederhana (Pengujian 1), kemudian dilanjutkan dengan memperdengarkan lagu ke 30 orang responden umum (Pengujian 2), dan 2 orang responden khusus dengan kompetensi pada bidang seni bernyanyi (Pengujian 3). Pengujian dengan melibatkan 30 orang responden umum bertujuan untuk mengukur tingkat kesesuaian hasil sintesis suara bernyanyi dengan naskah lagunya dan tingkat keramahan sistem bagi masyarakat umum. Pengujian dengan melibatkan 2 orang responden khusus bertujuan untuk mengukur kesesuaian hasil sintesis suara bernyanyi dengan naskah lagunya dari sudut pandang pelaku seni bernyanyi yang dianggap memiliki pengetahuan lebih terkait seni bernyanyi.

Ringkasan dari hasil Pengujian 1 yang bertujuan untuk melihat kinerja dari fitur-fitur utama dari sistem dapat dilihat pada Tabel 1. Hasil visualisasi Pengujian 1.4 tentang penggunaan tanda legato sebagai contoh hasil pengujian dapat dilihat pada Gambar 4.

Burung Hantu

Do = C M.M. = 60 4/4 NN

— —

||: 0 5̣ 1 . 7̣ | 1 3 1 . | 5̣ 3 2 . 1 | 2 3 1 . |

Ma - ta - ha - ri ter - be - nam ha - ri mu - lai ma - lam

— —

| 0 5̣ 1 . 7̣ | 1 3 1 . | 5̣ 3 2 . 1 | 2 3 1 . |

Ter - de - ngar bu - rung han - tu su - a - ra nya mer - du

— —

| 0 5 3 . | 0 5 3 . | 0 5 4 . 3 | 4 5 3 . ||

ku - ku ku - ku ku - ku - ku - ku - ku - ku

<< Burung Hantu

#Do=C. #B=4/4 #MM=60.

<N | 0 5, 1 (. 7,) | 1 3 1 . | 5, 3 2 (. 1) | 2 3 1 . | >N

<L Ma-ta-ha-ri te^r-be^nam ha-ri mu-lai ma-lam >L

<N | 0 5, 1 (. 7,) | 1 3 1 . | 5, 3 2 (. 1) | 2 3 1 . | >N

<L Te^r-de^-ngar bu-rung han-tu su-a-ra-nya me^r-du >L

<N | 0 5 3 . | 0 5 3 . | 0 5 4 (. 3) | 4 5 3 . |} >N

<L Ku-ku ku-ku ku-ku-ku-ku-ku >L

Gambar 3 Notasi musik angka format TXT untuk lagu Burung Hantu

Pengujian terhadap 30 orang responden umum dilakukan dengan meminta responden melakukan kegiatan-kegiatan sebagai berikut:

1. Mendengarkan 5 buah lagu hasil sintesis dan menilai kesesuaiannya dengan naskah setiap lagu yang dijadikan bahan uji (Pengujian 2.1)
2. Mendengarkan pengenalan singkat cara penggunaan sistem selama 5 menit, kemudian mencoba menulis sintaks teks lagu dari notasi angka yang diberikan untuk mensintesis lagu tersebut secara mandiri. (Pengujian 2.2)

Pada Pengujian 2.1, responden akan mendapatkan lembar naskah lagu dengan kolom penilaian pada setiap kata pada lagu tersebut. Jika suara hasil sintesis pada suatu kata (dari segi nada maupun pengucapan) dianggap sesuai dengan naskah lagunya, maka responden dapat mengisi kolom penilaian dengan nilai 1, sedangkan nilai 0 bisa diberikan jika suara hasil sintesis dianggap tidak sesuai dengan naskah lagunya. Hasil dari Pengujian 2.1 bisa dilihat pada Tabel 2.

Pada tahap pengujian ini kemudian ditemukan sejumlah bentuk kegagalan sintesis suara. Sejumlah bentuk kegagalan sintesis suara bernyanyi yang ditemukan oleh responden kemudian dapat diklasifikasikan ke dalam 3 bentuk utama sebagai berikut

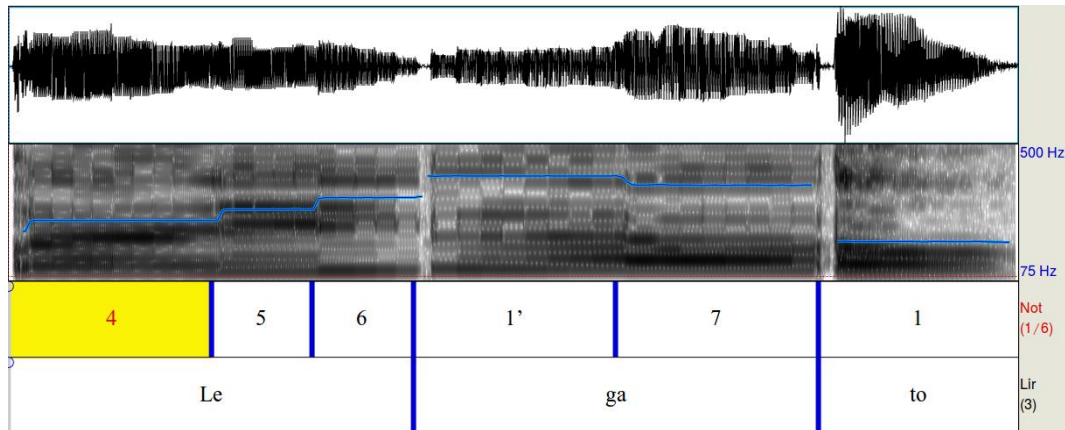
1. Perubahan pelafalan pada suatu fonem (misalnya fonem e mejadi I)
2. Penambahan atau pengurangan fonem pada suatu silabel
3. Terjadinya *noise* berupa suara dengung

Pengambilan data pada Pengujian 2.2 kemudian dilakukan dengan menghitung jumlah *error* yang terjadi selama proses sintesis lagu oleh responden. Data rekaman kesalahan sintaks dari Pengujian 2.2 terhadap 30 orang responden dapat dilihat pada Tabel 3. Pada tabel tersebut, dapat dilihat bahwa dalam pengujian ini terdapat 11 orang responden yang berhasil menuliskan sintaks teks lagu tanpa melakukan kesalahan sintaks, sedangkan 19 orang lainnya memerlukan kompilasi ulang hingga 3 kali untuk mengkonversi teks lagu ke transkripsi fonem tanpa melakukan kesalahan. Hasil ini menunjukkan bahwa sintaks teks lagu yang digunakan sudah cukup sederhana dan mudah dipahami oleh 30 orang responden umum.

Tabel 1 Ringkasan hasil uji sistem pengolahan teks dan sintesis suara

	Aspek yang diuji	Hasil transkripsi fonem	Hasil sintesis suara
Pengujian 1.1	Fungsi dasar, konfigurasi tempo dan konfigurasi nada dasar	Transkripsi fonem sesuai teks lagu, durasi ketukan sesuai tempo, dan nilai frekuensi sesuai nada dasar	Frekuensi dominan bisa terdeteksi dengan baik pada fonem vokal terakhir, dan tidak pada fonem lainnya
Pengujian 1.2	Penggunaan tanda pecahan waktu	Tanda pecahan waktu hingga $1/4$ dan $1/6$ dikonversi dalam milisekon dengan baik	Frekuensi dominan dapat ditemui hingga pecahan waktu terkecil
Pengujian 1.3	Penggunaan tanda kromatik dan tanda oktaf	Not yang dimaksud bisa dikonversi ke nilai frekuensi (satuan Hertz) dengan pembulatan ke bilangan bulat	Frekuensi dominan dapat ditemui perbedaannya pada interval setengah nada
Pengujian 1.4	Penggunaan tanda legato	Transkripsi fonem dapat membunyikan beberapa not pada sebuah fonem	Frekuensi dominan dapat ditemui sesuai not-not yang ada pada sebuah fonem

```
<<Pengujian 1.4
#Do=A #B=4/4 #MM=60.
<N [4 (5 6)] [1' 7] 1 |} >N
<L Le-ga-to >L
```



Gambar 4 Grafik suara hasil Pengujian 1.4

Tabel 2 Data akurasi sintesis dari Pengujian 2.1

Persentase suara silabel terdengar jelas (%)					
	Lagu A	Lagu B	Lagu C	Lagu D	Lagu E
Mean	87.47	74.67	89.02	77.91	79.47
SDEV	2.34	5.73	4.44	6.30	6.00
81.71% ±6.24					

Pengujian terhadap 2 responden khusus (Pengujian 3) kemudian dilakukan dengan pengambilan data secara kuantitatif dan kualitatif. Pengambilan data secara kuantitatif dilakukan dengan meminta responden mendengarkan 2 buah lagu hasil sintesis dan menilai kesesuaiannya dengan naskah setiap lagu yang dijadikan bahan uji. Pengujian secara kualitatif kemudian dilakukan dengan meminta komentar responden terkait desain umum sistem dan bagian hasil sintesis yang dianggap tidak sesuai dengan naskah lagu.

Responden khusus yang terlibat pada Pengujian 3 adalah

a. Bapak Petrus Martuty (Responden A)

Pelatih vokal *full-time* untuk kelompok paduan suara Solafide, Gereja Kristen Indonesia Gejayan, Yogyakarta (aktif)

b. Saudara Alfonsus Andaru (Responden B)

Pelatih Paduan Suara Mahasiswa, Universitas Gadjah Mada (periode 2015-2016)

Kedua orang responden ini mendengarkan 5 buah lagu sebagaimana dilakukan sebelumnya pada Pengujian 2.1. Setelah mendengarkan lagu-lagu tersebut dan memperhatikan lembar teks lagu yang digunakan untuk masing-masing lagu, didapatkan sejumlah tanggapan terkait hasil sintesis suara bernyanyi pada penelitian ini.

Tabel 3 Jumlah kesalahan sintaks pada Pengujian 2.2

	Urutan <i>recompile</i>				Urutan <i>recompile</i>				Urutan <i>recompile</i>		
	1	2	3		1	2	3		1	2	3
1	7	0	0	11	7	0	0	21	0	0	0
2	5	0	0	12	1	0	0	22	0	0	0
3	1	0	0	13	8	0	0	23	0	0	0
4	8	1	0	14	8	1	0	24	0	0	0
5	1	0	0	15	1	0	0	25	0	0	0
6	1	0	0	16	8	0	0	26	0	0	0
7	3	1	0	17	3	1	0	27	0	0	0
8	7	1	0	18	3	0	0	28	0	0	0
9	6	0	0	19	8	7	0	29	0	0	0
10	8	1	0	20	0	0	0	30	0	0	0

Secara umum kedua responden berpendapat bahwa fitur-fitur yang ditawarkan sudah berhasil menjalankan fungsinya. Akurasi nada dan tempo bisa diterima, sebagaimana teks lagu dinyanyikan dengan pembagian waktu yang tepat pada setiap nadanya (sesuai dengan hasil Pengujian 1.1 hingga Pengujian 1.4). Berkaitan dengan pengucapan kata-kata, kedua responden menyetujui bahwa adanya 3 macam bentuk kegagalan sintesis suara (seperti yang ditemukan pada Pengujian 2.1), namun secara keseluruhan lagu masih dapat dipahami.

Berkaitan dengan ketinggian nada pada suara bernyanyi, kedua responden menemukan sejumlah perbedaan pada hasil sintesis suara bernyanyi dengan suara bernyanyi yang sebenarnya. Responden A menyatakan bahwa hasil sintesis suara bernyanyi terdengar tidak alami pada nada tinggi. Hal ini kemungkinan terjadi karena basis data *diphone inventory* yang digunakan adalah rekaman suara laki-laki dengan frekuensi rendah. Responden B kemudian mengemukakan bahwa pada suara bernyanyi sebenarnya nada dengan durasi yang panjang tidak dinyanyikan dengan frekuensi yang datar, melainkan memberikan vibrasi dengan rentang tertentu. Hal ini dianggap sebagai hal yang menyebabkan hasil sintesis suara bernyanyi pada penelitian ini masih tidak seperti suara manusia saat menyanyikan not dengan durasi panjang.

Berkaitan dengan transisi antar silabel, kedua responden juga memiliki sejumlah pendapat. Responden A menilai bahwa manusia normal pada umumnya membutuhkan jeda waktu singkat untuk menarik nafas, yang umumnya dilakukan setiap awal bar pada notasi angka. Hal ini menyebabkan subjek yang belajar bernyanyi dengan menirukan hasil sintesis suara bernyanyi ini akan kelelahan karena tidak menemukan waktu untuk menarik nafas.

Hal-hal berkaitan dengan estetika seni bernyanyi kemudian menjadi komentar tambahan dari kedua responden. Responden B berpendapat bahwa ketika seseorang bernyanyi, akan ditemukan perbedaan karakter saat nada rendah dan nada tinggi. Saat penyanyi berusaha mencapai nada diluar frekuensi normalnya, maka akan ditemukan usaha yang lebih sehingga terdengar suara yang lebih bertenaga. Berkaitan korelasi nada dengan lirik, Responden B juga mengatakan bahwa pada saat melatih vokal, konteks dari lirik akan mempengaruhi penekanan pada silabel dari kata dengan makna lebih (yang pada penelitian selanjutnya bisa dilakukan dengan memvariasikan amplitudo dari hasil sintesis suara bernyanyi pada bagian tersebut).

4. KESIMPULAN

Dari Berdasarkan hasil penelitian mengenai sintesis suara bernyanyi dengan teknologi *text-to-speech* untuk notasi musik angka dan lirik bahasa Indonesia yang telah dilakukan, maka dapat diambil kesimpulan bahwa sistem sintesis suara bernyanyi berhasil dibuat dengan sejumlah fitur seperti konfigurasi nada dasar, konfigurasi unit pecahan waktu, konfigurasi tempo, penggunaan tanda partisi durasi not, penggunaan tanda oktaf, penggunaan tanda kromatik, dan penggunaan tanda legato. Sintesis suara bernyanyi pada setiap pengujian mampu menunjukkan frekuensi dominan dan ketepatan durasi sehingga interpretasi yang dihasilkan sesuai dengan teks lagu yang diberikan.

Secara umum, sistem sintesis suara bernyanyi dengan program translasi ke transkripsi fonem dan sintesis dengan metode penyambungan *diphone* berhasil menghasilkan sintesis suara bernyanyi dengan tingkat akurasi sebesar 81.71% dengan standar deviasi sebesar 6.24% berdasarkan pengujian terhadap 30 orang responden. Pada pengujian tersebut kemudian ditemukan 3 bentuk kegagalan utama pada hasil sintesis suara bernyanyi berupa perubahan pelafalan pada suatu fonem, penambahan atau pengurangan fonem pada suatu silabel, dan *noise* berupa suara dengung. Berdasarkan pendapat dari 2 orang responden ahli, hasil sintesis suara bernyanyi memiliki sejumlah perbedaan dengan suara bernyanyi yang sebenarnya seperti tidak ditemukannya suara jeda untuk menarik nafas, vibrasi pada nada panjang, dan perubahan karakter vokal di luar *range* normal penyanyi.

Aturan penulisan teks lagu bisa dikatakan *user-friendly* dengan jumlah kompilasi ulang paling banyak untuk menuliskan sintaks tanpa *error* adalah 3 kali pada pengujian penulisan 8 bar pertama teks lagu Himne Gadjah Mada oleh 30 orang responden.

5. SARAN

Pada penelitian mendatang, dapat dilakukan analisis terkait korelasi kecenderungan kegagalan sintesis dengan komposisi kata dan durasi pada teks lagu. Pengembangan lain yang bisa dilakukan terkait penelitian ini adalah upaya untuk menciptakan karakter suara hasil sintesis yang lebih alami dengan menciptakan suara jeda untuk menarik nafas, vibrasi pada nada panjang, dan perubahan karakter vokal di luar *range* normal penyanyi.

DAFTAR PUSTAKA

- [1] R. Hermawan, A. E. Putra, and C. Atmaji, "Metode Stereo Cancellation untuk Vocal Removal pada Lagu Pop, Rock, dan Jazz", IJEIS (Indonesian Journal of Electronics and Instrumentations System), Vol.4, No.1, April 2014, pp. 69~78 ISSN: 2088-3714, [Online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/4223/3477> [Accessed: 20-Dec-2017]
- [2] D. K. Wardhana and T. K. Priyambodo, "Pengoreksi Nada Menggunakan Mikrokontroler untuk Meningkatkan Kualitas Suara pada Biola", IJEIS (Indonesian Journal of Electronics and Instrumentations System), Vol.6, No.1, April 2016, pp. 59~68 ISSN: 2088-3714, [Online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/10772/8067> [Accessed: 20-Dec-2017]
- [3] A. Yuniar and B. N. Prastowo, "Optimasi Purwarupa Kendali Virtual Instrumen Musik Drum Berbasis Sensor Akselerometer dan LDR", IJEIS (Indonesian Journal of Electronics and Instrumentations System), Vol.3, No.2, October 2013, pp. 127~136 ISSN: 2088-3714, [Online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/3887/3176> [Accessed: 20-Dec-2017]

- [4] H. Y. Gu and J. K. He, "Singing-Voice Synthesis Using Demi-Syllable Unit Selection", Proceedings of the 2016 Indonesian Conference on Machine Learning and Cybernetics, Jeju, South Korea, 10-13 July, 2016, [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7872965> [Accessed: 20-Dec-2017]
- [5] K. Apriyanti and T. Y. Widodo, "Implementasi Optical Character Recognition Berbasis Backpropagation untuk Text to Speech Perangkat Android", IJEIS (Indonesian Journal of Electronics and Instrumentations System), Vol.6, No.1, April 2016, pp. 13~24 ISSN: 2088-3714, [Online]. Available: <https://jurnal.ugm.ac.id/ijeis/article/view/10767/8062> [Accessed: 20-Dec-2017]
- [6] I. Ramli, N. Jamil, N. Seman, and N. Ardi, "An improved pitch contour formulation for Malay language storytelling Text-to-Speech (TTS)", Proceedings of the 2016 IEEE Industrial Electronics and Applications Conference (IEACon), Kota Kinabalu, 2016, pp. 250-255. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8067387> [Accessed: 20-Dec-2017]
- [7] Y. Suyanto, Subanar, A. Harjoko, and S. Hartati, "An Intonation Speech Synthesis Model for Indonesian Using Pitch Pattern and Phrase Identification", Journal of Signal and Information Processing, 5, 80-88, [Online]. Available: https://file.scirp.org/pdf/JSIP_2014073015180798.pdf [Accessed: 20-Dec-2017]
- [8] Kemendikbud, 2016, Pedoman Umum Ejaan Bahasa Indonesia, Edisi Keempat, Grasindo, Jakarta
- [9] Y. Suyanto, "The Numbered Musical Notation Editor in Plain Text", Proceedings of the 2015 International Conference on Science and Technology, Yogyakarta, Indonesia, 11-13 November, 2015
- [10] S. Lukose and S. S. Upadhyaya, "Text to speech synthesizer-formant synthesis", Proceedings of the 2017 International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, 2017, pp. 1-4. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7947945> [Accessed: 20-Dec-2017]