



TESIS – TE142599

**MODIFIKASI SKALA WAKTU PADA REKAMAN SUARA
MENGUNAKAN METODE WAVEFORM SIMILARITY OVERLAP
AND ADD (WSOLA)**

**NANDA SAPUTRI
07111550052008**

DOSEN PEMBIMBING

**Prof. Dr. Ir. Yoyon Kusnendar Suprpto M.Sc.
Dr. Diah Puspito Wulandari, S.T., M.Sc.**

PROGRAM MAGISTER

**BIDANG KEAHLIAN JARINGAN CERDAS MULTIMEDIA
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018**



TESIS – TE

**MODIFIKASI SKALA WAKTU PADA REKAMAN SUARA
MENGUNAKAN METODE WAVEFORM SIMILARITY OVERLAP
AND ADD (WSOLA)**

NANDA SAPUTRI
07111550052008

DOSEN PEMBIMBING

Prof.Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc.
Dr. Diah Puspito Wulandari, S.T., M.Sc.

PROGRAM MAGISTER
BIDANG KEAHLIAN JARINGAN CERDAS MULTIMEDIA
DEPARTEMEN TEKNIK ELEKTRO
FAKULTAS TEKNOLOGI ELEKTRO
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2018

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Teknik (M.T)
di
Institut Teknologi Sepuluh Nopember

oleh:

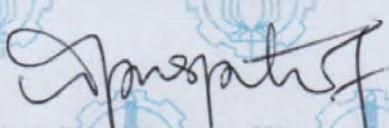
Nanda Saputri
NRP. 07111550052008

Tanggal Ujian : 26 Juni 2018
Periode Wisuda : September 2018

Disetujui oleh:


1. Prof. Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc
NIP: 195409251978031001

(Pembimbing I)


2. Dr. Dian Puspito Wulandari, ST., M.Sc.
NIP: 198012192005012001

(Pembimbing II)


3. Dr. Eko Mulyanto Yuniarno, S.T., M.T.
NIP: 196806011995121009

(Penguji)


4. Dr. Supeno Mardhi Susiki Nugroho, ST., M.T.
NIP: 197003131995121001

(Penguji)

Dekan Fakultas Teknologi Elektro


Dr. Tri Arief Sardjono, S.T., M.T.
NIP. 19700212 199512 1 001

Halaman ini sengaja dikosongkan

PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul “**MODIFIKASI SKALA WAKTU PADA REKAMAN SUARA MENGGUNAKAN METODE WAVEFORM SIMILARITY OVERLAP AND ADD (WSOLA)**” adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, Juni 2018

Nanda Saputri

NRP. 07111550052008

Halaman ini sengaja dikosongkan

**MODIFIKASI SKALA WAKTU PADA REKAMAN SUARA
MENGUNAKAN METODE METODE WAVEFORM SIMILARITY
OVERLAP AND ADD (WSOLA)**

Nama mahasiswa :Nanda Saputri
NRP :07111550052008
Pembimbing :1. Prof.Dr. Ir.Yoyon Kusnedar Suprpto,M.Sc.
2. Dr. Diah Puspito Wulandari, S.T., M.Sc.

ABSTRAK

Setiap manusia memiliki kemampuan berbeda beda dalam mendengarkan dan melafalkan ucapan.Orang melafalkan ucapan dengan cepat atau lambat. Begitu juga kemampuan mendengar ada yang dapat mendengar dengan normal dan ada yang pendengarannya menurun yang disebabkan oleh faktor keturunan, usia, penyakit dan sebagainya. Agar suara yang cepat dan bising tersebut dapat terdengar jelas, pada umumnya pendengar menggunakan cara konvensional untuk mengatasi permasalahan tersebut dengan menggunakan aplikasi atau *tape recorder* untuk memperlambat rekaman suara.

Pada penelitian ini dilakukan proses *time stretching*, yaitu pergeseran kerapatan waktu sinyal suara tanpa mengubah frekuensi dasar menggunakan metode WSOLA. Menghitung panjang maksimal waktu yang dapat digeser dengan frekuensi tetap dan suara masih terdengar dengan baik dan membandingkan dari metode WSOLA dan PSOLA tersebut yang mana bisa menghasilkan suara yang lebih baik. Diharapkan penelitian ini dapat membantu pendengaran kita untuk mendengar suara lebih baik atau jelas walaupun kecepatan suaranya dirubah.

Dari hasil penelitian menunjukkan bahwa tidak memberi nilai toleransi pada metode WSOLA mengakibatkan frekuensi maksimum suara menjadi berbeda dengan suara asli. Dari pengujian pergeseran kerapatan waktu sinyal suara saron lima slendro menggunakan WSOLA dapat dilakukan modifikasi skala waktu sinyal suara saron dan suara manusia dengan tetap mempertahankan frekuensi aslinya dengan error rata-rata suara saron 0.847% dan suara manusia 5,094%.

Kata kunci: (Time stretching, WSOLA, Time Scale Modification)

Halaman ini sengaja dikosongkan

TIME SCALE MODIFICATION ON VOICE RECORD USING WAVEFORM SIMILARITY OVERLAP AND ADD (WSOLA) METHOD

By :Nanda Saputri
Student Identity Number :07111550052008
Supervisors :1. Prof.Dr. Ir.Yoyon Kusnedar Suprpto,M.Sc.
2. Dr. Diah Puspito Wulandari, S.T., M.Sc.

ABSTRACT

Every people has a different ability to listen and pronounce speech. People recite speech quickly or slowly. Likewise, there is a listening ability that can hear normally and there is a decreased hearing caused by heredity, age, disease and so forth. So that the fast and noisy sound can be heard clearly, listeners generally use conventional methods to overcome these problems by using an application or tape recorder to slow down the sound recording.

In this research, the time stretching process is used, which is the time noise signal density shift without changing the basic frequency using WSOLA method. Calculating the maximum length of time that can be shifted with a fixed frequency and sound still sounds well and compares with the WSOLA and PSOLA methods which can produce better sound. It is hoped that this research can help our hearing to hear sounds better or clearer even though the speed of the voice is changed.

The results of the study showed that not giving the tolerance value on the WSOLA method resulted in the maximum frequency of sound being different from the original sound. From testing the time density shift of the saron five slendro sound signal using WSOLA can be modified the time scale of the saron sound signal and human voice while maintaining the original frequency with an average error of saron 0.847% and 5.094% human voice.

Keywords: (Time stretching, WSOLA, Time Scale Modification)

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Syukur Alhamdulillah dipanjatkan kepada Allah SWT, berkat rahmat dan karunia-Nya tesis ini dapat diselesaikan. Penulis menyampaikan ucapan terima kasih yang tak terhingga kepada semua pihak, yang telah membantu dalam menyelesaikan tesis ini. Secara khusus penulis menyampaikan ucapan terima kasih yang tulus kepada :

1. Prof. Dr. Ir. Yoyon Kusnendar Suprpto, M.Sc dan Dr. Diah Puspito Wulandari, ST., M.Sc selaku dosen pembimbing yang telah mengarahkan, memberi koreksi, dan motivasi dalam tesis ini.
2. Dr. Eko Mulyanto Yuniarno, S.T., M.T, selaku koordinator bidang keahlian Jaringan Cerdas Multimedia Program Studi Teknik Elektro.
3. Bapak Dewan penguji selaku dosen penguji yang telah memberikan saran dan kritik dalam tesis ini.
4. Bapak-bapak dosen pengajar di Program Studi Teknik Elektro, bidang keahlian Jaringan Cerdas Multimedia.
5. Ayahanda, Ibunda, suami serta keluarga yang menjadi motivasi penulis untuk sukses dan selalu memberikan dukungan baik materi maupun non-materi, perhatian serta doa tanpa henti kepada penulis.
6. Terima Kasih Kepada Teman-Teman seangkatan JCM 2015 dan 2016, yang selalu mendoakan dan menyemangati penulis.
7. Semua pihak yang telah banyak membantu proses penyelesaian tesis ini.

Penulis menyadari bahwa penulisan tugas akhir ini masih banyak keterbatasan dan kekurangan. Untuk itu kritik serta saran yang sifatnya membangun sangat penulis harapkan. Akhirnya penulis ucapkan semoga tugas akhir ini dapat berguna bagi pembaca dan bagi penulis khususnya.

Surabaya, Mei 2018

Penulis

Halaman ini sengaja dikosongkan

DAFTAR ISI

LEMBARAN PENGESAHAN.....	ixv
PERNYATAAN KEASLIAN TESIS	vii
ABSTRAK	ix
ABSTRACT.....	xi
KATA PENGANTAR	xiii
DAFTAR ISI	xv
DAFTAR GAMBAR	xvii
DAFTAR TABEL.....	xix
DAFTAR NOMENKLATUR.....	xxi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan	3
1.4 Batasan Masalah	3
1.5 Mamfaat	3
1.6 Sistematika Penulisan.....	3
BAB 2 KAJIAN PUSTAKA.....	5
2.1 Kajian Penelitian Terkait	5
2.2 Audio.....	6
2.3 Pengolahan Sinyal Digital.....	8
2.4 Time Stetching.....	10
2.5 Proses Sampling.....	13
2.6 Frame Blocking.....	13
2.7 Time Streching Algorithm.....	15
2.7.1 OLA(Overlapp and Add.....	15
2.7.2 SOLA(Shynchronous Overlap and Add).....	19
2.7.3 PSOLA(Pitch Shynchronous Overlap and add).....	19
2.7.4 WSOLA(Waveform Similarity Overlapp and Add).....	21
2.8 Window.....	24

2.8.1	Window Hamming.....	25
2.8.1	Window Rectangular.....	25
2.8.1	Window Triangular.....	25
2.9	Fast Fourier Transform(FFT).....	26
2.10	PESQ(Perceptual Evaluation Speech Quality.....	28
2.11	Korelasi.....	29
BAB 3 METODOLOGI PENELITIAN.....		31
3.1	Perancangan Sistem.....	32
3.2	Pelaksanaan Uji Coba.....	36
3.3	Evaluasi Kerja.....	36
3.4	Hasil yang Diharapkan.....	37
3.5	Dokumentasi.....	37
BAB 4 HASIL PENELITIAN.....		39
4.1	Sinyal Masukan.....	39
4.2	WSOLA.....	39
4.3	Pengaruh nilai toleransi.....	41
4.4	Pengaruh Window.....	43
4.5	Pengaruh Jarak antar Frame.....	49
4.6	Perbandingan Metode WSOLA dan PSOLA.....	51
4.7	Mean Opinion Score (MOS).....	55
BAB 5 PENUTUP.....		59
5.1	KESIMPULAN.....	59
5.2	SARAN.....	59
DAFTAR PUSTAKA.....		61
LAMPIRAN.....		63
BIODATA PENULIS.....		71

DAFTAR GAMBAR

Gambar 2.1 Gelombang suara.....	7
Gambar 2.2 Proses sampling.....	8
Gambar 2.3 Pengubahan dari sinyal kontinyu ke sinyal diskret	9
Gambar 2.4 Blok Diagram Sistem Pengolahan Sinyal Digital	10
Gambar 2.5 Prinsip Dasar Prosedur Time Scale Modification	11
Gambar 2.6 Proses peregangan dan pengompresan frame sinyal	14
Gambar 2.7 Diskontinyuitas sinyal ketika tahap peregangan dan pengkompresan dalam domain waktu.....	15
Gambar 2.8 Time Stretching : Overlap Add	16
Gambar 2.9 Artefak khas yang terjadi saat memilih frame sintesis ym sama dengan frame analisis xm	16
Gambar 2.10 Prinsip TSM berdasarkan overlap-add (OLA).....	18
Gambar 2.11 Time Stretching: Synchronous Overlapp and Add	19
Gambar 2.12 Pitch analysis and block window	20
Gambar 2.13 Depiction of the PSOLA Algorithm	21
Gambar 2.14 Prinsip Kesamaan Bentuk Gelombang Overlap-Add (WSOLA)....	22
Gambar 2.15 Sinyal sinus dalam domain frekuensi.....	27
Gambar 2.16 Analisis FFT	27
Gambar 3.1 <i>Prosedure time scale modification</i>	32
Gambar 3.2 Proses metode WSOLA	33
Gambar 3.3 Sinyal Masukan.....	34
Gambar 4.1 Sinyal asli berwarna merah dan sinyal hasil WSOLA yang diberi skala faktor 0.5 berwarna biru	39
Gambar 4.2 Sinyal asli berwarna merah dan sinyal hasil WSOLA yang diberi skala faktor 2 berwarna biru.	40
Gambar 4.3 Frekuensi yang dihasilkan dari saron5 asli, saron5 skala faktor 0.5 dan saron5 skala faktor 2	40
Gambar 4.4 pengaruh nilai toleransi WSOLA pada saron1sl.wav	42
Gambar 4.5 Grafik pengaruh nilai toleransi pada tatatertib.wav	43

Gambar 4.6	Garis merah merupakan transform hilbert sinyal hasil pergeseran, garis biru adalah sinyal asli	44
Gambar 4.7	Bentuk sinyal	45
Gambar 4.8(a)	Pola sinyal windowhamming dan triangular menggunakan transform hilbert dengan jarak overlap frame 100%.....	46
Gambar 4.8(b)	Pola sinyal windowhamming dan triangular menggunakan transform hilbert dengan jarak overlap frame 100%.....	46
Gambar 4.9	Bentuk sinyal hasil dari WSOLA yang diberi jarak terendah antar frame 1.25 ms.	50
Gambar 4.10	Bentuk sinyal dari penggunaan jarak antar frame 125% dengan diberi window rectangular,hamming dan triangular	51
Gambar 4.11	Grafik perbedaan frekuensi WSOLA dan PSOLA.....	53
Gambar 4.12	Grafik hasil Kuisisioner metode WSOLA berdasarkan penggunaan window dan jarak overlap.....	56

DAFTAR TABEL

Tabel 4.1 Pengujian Skala waktu WSOLA berdasarkan hasil frekuensi	41
Tabel 4.2 PESQ WSOLA dengan skala faktor 2	47
Tabel 4.3 PESQ WSOLA dengan skala faktor 0.5	47
Tabel 4.4 PESQ WSOLA pada suara tatatertib1.wav dengan skala faktor 2	48
Tabel 4.5 PESQ WSOLA pada suara tatatertib1.wav dengan skala faktor 0.5 ..	48
Tabel 4.6 PESQ WSOLA pada suara goodfriend.wav dengan skala faktor 2	48
Tabel 4.7 PESQ WSOLA pada suara goodfriend.wav dengan skala faktor 0.5..	49
Tabel 4.8 Pengujian pergeseran waktu PSOLA berdasarkan hasil frekuensi.....	52
Tabel 4.9 Perbandingan metode WSOLA dan PSOLA berdasarkan frekuensi.....	53
Tabel 4.10 Perbandingan metode WSOLA dan PSOLA berdasarkan frekuensi pada skala factor ekstrim.....	54

Halaman ini sengaja dikosongkan

DAFTAR NOMENKLATUR

1	T	:	Periode
2	F	:	frekuensi
3	x	:	sinyal masukan
4	M	:	Iterasi
5	N	:	panjang sampel
6	R	:	Posisi bingkai
7	x_m	:	Bingkai analisis
8	y_m	:	Bingkai sintesis
9	H_a	:	Analisis hopsize
10	H_z	:	Sintetis hopsize
11	w	:	Window
12	ω_0	:	Frekuensi sudut fundamental
13	L	:	Panjang perbingkai
14	R	:	Panjang pergeseran
15	Δ	:	Nilai toleransi

Halaman ini sengaja dikosongkan

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Setiap manusia memiliki keunikan suara dan frekuensi yang berbeda dalam melafadkan ucapan. Dari suku tertentu, ada manusia yang melafadkan ucapan dengan cepat dan keras. Dan sebaliknya, ada manusia yang melafadkan ucapan dengan pelan dan lambat. Sama halnya dengan pendengaran manusia, ada yang dapat mendengar dengan normal dan ada yang pendengarannya menurun yang disebabkan oleh faktor keturunan, usia, penyakit dan sebagainya. Contohnya pada saat mendengarkan suatu pidato, di saat yang sama orang yang duduk paling belakang akan sulit menangkap pidato orang tersebut apabila pengucapannya terlalu cepat dan terburu-buru. Pada saat mendengar suatu rekaman suara yang suaranya cepat misalnya sebuah lagu sehingga sulit untuk mengetahui lirik yang dinyanyikan. Faktor umur yang tua juga mempengaruhi penurunan pada pendengaran sehingga rekaman sebuah suara pun susah didengarkan jika dilafadkan dengan cepat. Agar dapat mewujudkan dari tujuan komunikasi, biasanya pendengar menggunakan *tape recorder* atau aplikasi lainnya untuk mengulangi rekaman tersebut.

Pada dunia pembelajaran, studi telah menunjukkan bahwa mendengarkan dua kali bahan ajar yang telah dipercepat oleh faktor dua lebih efektif daripada mendengarkan mereka sekali pada kecepatan normal (Sticht, 1969). Seperti pada pembelajaran bahasa asing, memperlambat rekaman pembicara asing bisa menjadi cara yang baik untuk memfasilitasi belajar bahasa asing. Kemudian, ketika kemampuan siswa meningkat, tingkat pembicara dapat ditingkatkan secara bertahap. Pada pembelajaran untuk orang buta, rekaman pidato dapat menjadi alternatif untuk membaca bagi orang buta, meskipun biasanya orang dapat membaca dengan lebih cepat daripada yang dapat berbicara. Dengan modifikasi skala waktu yang tepat, tingkat rekaman pembicaraan dapat ditingkatkan sambil mempertahankan kejelasan. (Jordi, 2002)

Pada dunia hiburan dan musik, Seringkali soundtrack disiapkan secara independen dari gambar yang seharusnya diiringi dan karena itu perlu disinkronkan antara soundtrack dan gambar. Komposer musik yang sering bekerja dengan materi yang direkam sebelumnya suka mengontrol waktu dan pitch secara terpisah, ini membutuhkan penyesuaian antara waktu dan pitch agar menghasilkan suara musik seperti yang diinginkan yang biasanya membutuhkan alat komposer. Konduktor orkestra virtual dengan tingkat realisme yang tinggi membutuhkan tidak hanya volume dan instrumentasi tetapi juga tempo orkestra untuk mengendalikan rekaman audio dan video asli dari orkestra nyata.(Borchers,2001). Pada penyiaran, untuk menjaga kualitas soundtrack pada konversi antara format video (25 atau 30 fps) dan bioskop (24fps) agar sesuai dengan standar penyiaran membutuhkan modifikasi skala waktu berkualitas tinggi. (Pallone,1999).

Dari beberapa permasalahan suara yang sulit didengar, pada umumnya pendengar menggunakan cara konvensional untuk mengatasi permasalahan tersebut dengan menggunakan aplikasi atau alat yang ada. Aplikasi atau alat tersebut, dihasilkan suara yang tidak diinginkan pendengar karena suara yang dihasilkan sudah berubah frekuensinya yang disebabkan oleh berubahnya panjang periode waktu pada suara yang dihasilkan. Pada tesis ini dilakukan proses modifikasi skala waktu, yaitu pergeseran kerapatan waktu suara tanpa mengubah frekuensi awal dengan menggunakan metode WSOLA. Metode modifikasi skala waktu diaplikasikan kedalam suara manusia dan suara alat musik saron dari gamelan Jawa dan dicari tingkat akurasi frekuensi pergeseran yang dihasilkan.Mencari pengaruh dari parameter metode WSOLA terhadap frekuensi dan suara yang dihasilkan.

1.2 Rumusan Masalah

1. Rekaman suara jika diperlambat frekuensinya akan turun dan jika dipercepat frekuensinya akan naik, hal ini membuat suara sulit untuk didengar dengan jelas.

2. Bagaimana mengetahui pengaruh dari parameter metode WSOLA terhadap frekuensi dan suara yang dihasilkan.

1.3 Tujuan

Modifikasi skala waktu pada rekaman suara jika diperlambat atau dipercepat, namun frekuensinya tetap menggunakan metode WSOLA. Mengetahui pengaruh dari parameter metode WSOLA terhadap frekuensi dan suara yang dihasilkan.

1.4 Batasan Masalah

1. Metode yang digunakan adalah WSOLA
2. Rekaman suara dalam format wav
3. Data suara saron dan suara manusia

1.5 Mamfaat

Modifikasi skala waktu rekaman suara menggunakan metode WSOLA ini diharapkan membantu pendengaran kita untuk mendengar suara lebih baik atau jelas walaupun kecepatan suaranya dirubah.

1.6 Sistematika Penulisan

Penelitian ini disusun menjadi lima bab sebagai berikut :

Bab I Pendahuluan

Pada bab ini akan dibahas mengenai latar belakang, perumusan masalah, tujuan, manfaat penelitian, pembatasan masalah, dan sistematika penulisan dari penelitian ini.

Bab II Tinjauan Pustaka

Pada bab ini akan dibahas mengenai teori-teori penunjang yang akan digunakan untuk mengatur rekaman suara menggunakan metode WSOLA yaitu penelitian terkait dan teori *time stretching*, time scale modification, WSOLA

Bab III Metodologi Penelitian

Bab ini berisi tahapan proses perancangan untuk proses Time stretching

Bab IV Hasil Penelitian dan Pembahasan

Bab ini berisi data pengamatan hasil pengujian, perhitungan tingkat keberhasilan metoda yang digunakan, dan analisa dari data hasil pengujian program.

Bab V Kesimpulan dan Saran

Pada bab ini berisi kesimpulan dari penelitian dan saran-saran yang perlu dilakukan untuk pengembangan di masa mendatang.

BAB 2

KAJIAN PUSTAKA

Mengatur panjang rekaman suara menggunakan metode WSOLA memerlukan beberapa pemahaman teori dasar. Pemahaman teori ini sangat bermanfaat untuk memahami cara kerja dari sistem ini. Bab ini akan dibahas tentang teori tentang *time stretching*, pengolahan sinyal suara dan WSOLA. Beberapa aspek yang tidak berkaitan langsung dengan tugas akhir ini tidak akan dibahas dalam bab ini sehingga pembaca dianjurkan untuk mengacu ke buku referensi yang dipakai untuk pemahaman lebih.

2.1 Kajian Penelitian Terkait

Prasetio,2014,merancang sebuah sistem perbaikan suara manusia terhadap referensi musik dengan menggunakan metode phase vocoder.Di bagi menjadi dua tahap utama yaitu pitch detection dan pitch correction. Disini menggeser frekuensi dengan nilai waktu tetap.

Kong,2015 mengimplementasi real-time dari vocoder fase pada chip TI TMS320C5535 DSP. Tahap pembuatan dapat dibagi menjadi dua bagian: Fast Lifting Wavelet Transform (FLWT) digunakan untuk deteksi pitch, dan Time-Domain pitch Synchronous Tumpang tindih dan Tambahkan metode (TD-PSOLA) untuk koreksi pitch. Sebuah implementasi dari algoritma disediakan yang dioptimalkan untuk prosesor yang kurang dari unit aritmatika floating-point.

Chellam,2015 menggabungkan emosi yang diinginkan ke dalam suara netral, metode pemrosesan sinyal yang digunakan dalam karya ini untuk memodifikasi parameter suara prosodi dalam domain waktu, baik dalam beberapa kata atau seluruh ucapan. Analisis awal dilakukan, dengan membandingkan suara netral dengan pidato bahagia dan sedih. Berdasarkan pengamatan dari analisis, parameter dari sinyal suara yang bervariasi menggunakan teknik TD-PSOLA. Parameter, singkat energi waktu, durasi, dan pitch kontur dalam suara netral

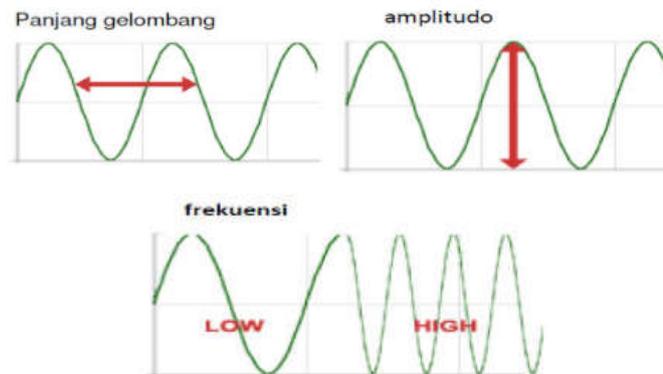
dimodifikasi dan selanjutnya dianalisis secara kuantitatif untuk menentukan kombinasi modifikasi parameter yang lebih baik untuk mensintesis suara emosional dari suara netral.

(Kail,2005)Mengembangkan sistem pakar secara sederhana untuk pemantauan phonocardiographic (PCG) di unit perawatan koroner dan di rumah. Tujuan penelitian adalah untuk menguji kinerja transformasi wavelet pada kelompok gagal jantung dengan lingkungan yang bising. Untuk deteksi dan interpretasi bunyi jantung dan murmur, dua analisis metoda dilakukan. Sebagai pertama, dengan algoritma Waveform Similarity Overlap-and-Add (WSOLA), pemisahan suara yang lebih baik dicapai untuk penentuan standar emas untuk interpretator kardiologis.

(Grofit,2008) perceptually significant transient sections (PSTs) dideteksi menggunakan Mel frekuensi cepstrum ukuran nonstasioneritas dan korelasi silang yang dinormalkan, dengan fungsi ambang waktu yang bervariasi. pendeteksian PSTs yang akurat dalam kerangka WSOLA memungkinkan untuk mencapai kualitas yang lebih tinggi dari musik skala waktu, yang dikonfirmasi oleh tes pendengaran subjektif.

2.2 Audio

Semua gelombang suara pasti memiliki tiga sifat penting untuk kerja audio, yaitu meliputi : panjang gelombang, amplitudo dan frekuensi. Gelombang suara adalah gelombang yang dihasilkan dari sebuah benda yang bergetar. Sebagai contoh, senar gitar yang dipetik, gitar akan bergetar dan getaran ini merambat di udara, atau air, atau material lainnya. Satu-satunya tempat dimana suara tak dapat merambat adalah ruangan hampa udara. Gelombang suara ini memiliki lembah dan bukit, satu buah lembah dan bukit akan menghasilkan satu siklus atau periode. Siklus ini berlangsung berulang-ulang, yang membawa pada konsep frekuensi. Gelombang suara dapat juga ditunjukkan dalam suatu grafik standar x versus y seperti ditunjukkan Gambar 2.1.



Gambar 2.1 Gelombang suara (wahyudi,2015)

Gambar 2.1 memungkinkan untuk memvisualisasi gelombang dengan sudut pandang matematis, menghasilkan kurva yang dikenal sebagai bentuk gelombang. Periode gelombang adalah jarak antar titik gelombang dan titik ekuivalen pada fasa berikutnya. Amplitudo adalah kekuatan atau daya gelombang sinyal. Tinggi gelombang yang bisa dilihat sebagai grafik. Gelombang yang lebih tinggi diinterpretasikan sebagai volume yang lebih tinggi, sehingga dinamakan amplifier untuk perangkat yang menambah amplitude. Frekuensi adalah jumlah getaran yang terjadi dalam waktu satu detik. Diukur dalam hertz atau siklus per detik. Getaran gelombang suara semakin cepat, frekuensi semakin tinggi. Frekuensi lebih tinggi diinterpretasikan sebagai jalur lebih tinggi. Sesuai dengan Persamaan 2.1.

$$F = \frac{1}{T} \tag{2.1}$$

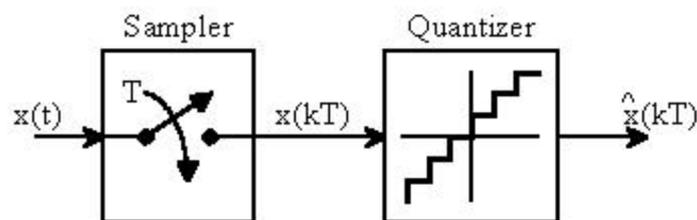
Dari Persamaan 2.1, apabila nilai periode sinyal suara (T) diperbesar atau diperlambat maka frekuensi suara (F) yang dihasilkan berubah menjadi kecil. Dan sebaliknya, bila nilai periode sinyal suara (T) diperkecil atau dipercepat maka frekuensi suara (F) yang dihasilkan berubah menjadi besar.

Telinga manusia dapat mendengar bunyi antara 20 Hz hingga 20 kHz (20.000 Hz) sesuai batasan sinyal audio. Karena pada dasarnya sinyal suara adalah sinyal yang dapat diterima oleh telinga manusia. Angka 20 Hz sebagai frekuensi suara terendah yang dapat didengar, sedangkan 20 KHz merupakan frekuensi tertinggi yang dapat didengar. Gelombang suara bervariasi sebagaimana variasi tekanan

media perantara seperti udara. Suara diciptakan oleh getaran dari suatu obyek, yang menyebabkan udara disekitarnya bergetar. Getaran udara ini kemudian menyebabkan gendang telinga manusia bergetar, yang kemudian oleh otak dianggap sebagai suara.

2.3 Pengolahan Sinyal Digital

Proses pengolahan sinyal digital, diawali dengan proses pencuplikan sinyal masukan yang berupa sinyal kontinyu. Proses ini mengubah representasi sinyal yang tadinya berupa sinyal kontinyu menjadi sinyal diskrete. Proses ini dilakukan oleh suatu unit ADC (*Analog to Digital Converter*). Unit ADC ini terdiri dari sebuah bagian *Sample/Hold* dan sebuah bagian quantiser seperti pada Gambar 2.2.



Gambar 2.2 Proses sampling (Virdaga, 2013)

Unit sample/hold merupakan bagian yang melakukan pencuplikan orde ke-0, yang berarti nilai masukan selama kurun waktu T dianggap memiliki nilai yang sama. Pencuplikan dilakukan setiap satu satuan waktu yang lazim disebut sebagai waktu cuplik (*sampling time*). Bagian quantiser akan merubah menjadi beberapa level nilai, pembagian level nilai ini bisa secara *uniform* ataupun secara *non-uniform* missal pada *Gaussian quantiser*.

Unjuk kerja dari suatu ADC bergantung pada beberapa parameter, parameter utama yang menjadi pertimbangan adalah :

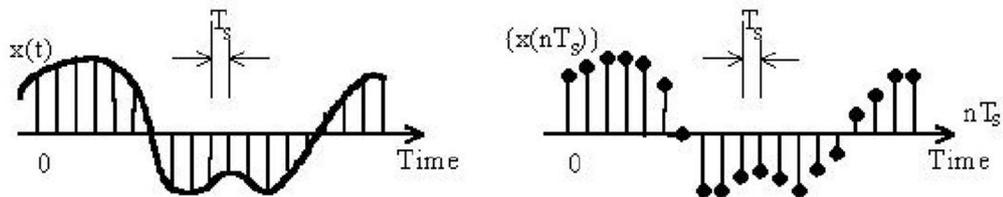
1. Kecepatan maksimum dari waktu cuplik.
2. Kecepatan ADC melakukan konversi.

3. Resolusi dari *quantizer*, misal 8 bit akan mengubah menjadi 256 tingkatan nilai.

4. Metoda kuantisasi akan mempengaruhi terhadap kekebalan *noise*.

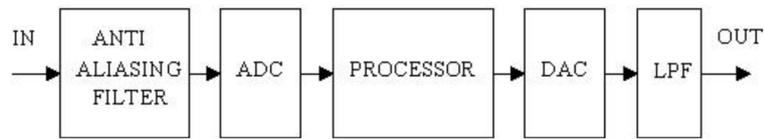
Sinyal input asli yang tadinya berupa sinyal kontinyu, $x(T)$ akan dicuplik dan dikuantisasi sehingga berubah menjadi sinyaldiskret $x(kT)$. Dalam representasi yang baru inilah sinyal diolah.

Keuntungan dari metoda ini adalah pengolahan menjadi mudah dandapat memanfaatkan program sebagai pengolahnya. Dalam prosesampling ini diasumsikan kita menggunakan waktu cuplik yangsama dan konstan, yaitu T_s . Parameter cuplik ini menentukan dari frekuensi harmonis tertinggi dari sinyal yang masih dapat ditangkapoleh proses cuplik ini. Frekuensi sampling minimal adalah 2 kalidari frekuensi harmonis dari sinyal.Untuk mengurangi kesalahan cuplik maka lazimnyadigunakan filter *anti-aliasing* sebelum dilakukandiproses cuplikan. Filter ini digunakan untuk meyakinkan bahwa komponen sinyal yang dicuplik adalah benar-benar yang kurang daribatastersebut. Sebagai ilustrasi, proses pencuplikan suatu sinyaldigambarkan pada Gambar 2.3.



Gambar 2.3 Pengubahan dari sinyal kontinyu ke sinyal diskret. (Virdaga,2013)

Setelah sinyal diubah representasinya menjadi deretan datadiskrete, selanjutnya data ini dapat diolah oleh prosesor menggunakan suatu algoritma pemrosesan yang diimplementasikan dalam program. Hasil dari pemrosesan akan dilewatkan ke suatuDAC (*Digital to Analog Converter*) dan LPF (*Low Pass Filter*) untuk dapat diubah menjadi sinyal kontinyu kembali. Secara garis besar, blok diagram dari suatu pengolahan sinyal digital adalahseperti pada Gambar 2.4 :



Gambar 2.4 Blok Diagram Sistem Pengolahan Sinyal Digital. (Virdaga, 2013)

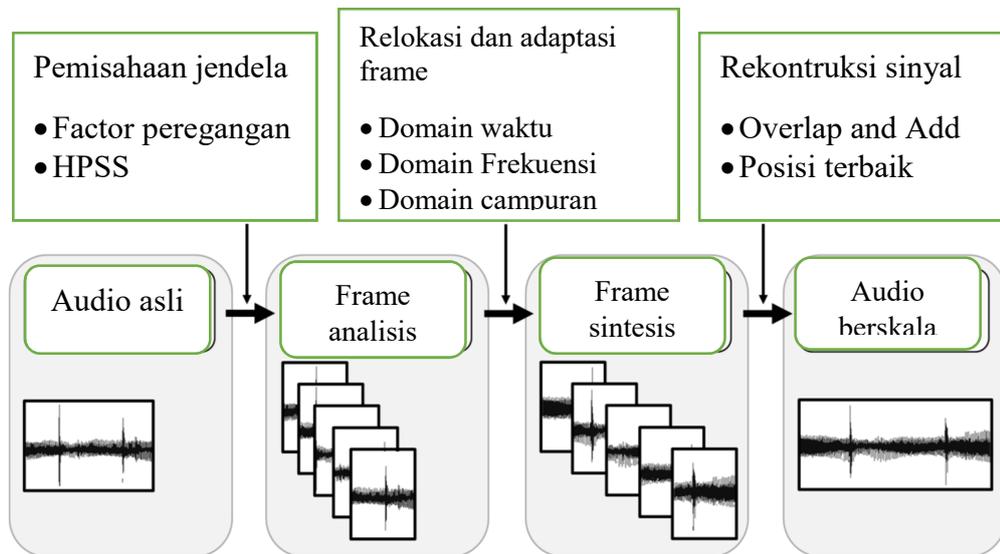
Keluaran sinyal diskret inilah nantinya akan diproses dalam metode WSOLA.

2.4 Time Stretching

Time scale atau *time stretching* adalah proses memperlambat atau mempercepat sinyal audio tanpa mengubah *pitch*. Aplikasi dari skala waktu yang digeser ini banyak, contohnya seperti membaca teks untuk orang buta, dan belajar bahasa asing. Skala waktu dilakukan dengan membagi sinyal menjadi frame tetap yang tumpang tindih. Frame tumpang tindih ini kemudian bergeser sesuai dengan tujuan keseluruhan (mempercepat atau memperlambat) dan dikombinasikan untuk memberikan output direkonstruksi. (Dave, 2000)

Time-scale modification (TSM) adalah tugas memanipulasi sinyal audio seperti yang kedengarannya seolah-olah isinya dilakukan di tempo yang berbeda. TSM menemukan aplikasi misalnya dalam musik remixing di mana ia digunakan untuk mengatur kecepatan pemutaran rekaman yang ada sehingga mereka dapat dimainkan secara bersamaan pada tempo yang sama (Hiromi, Alexis). bidang lain dari aplikasi adalah penyesuaian aliran audio dalam video klip. Misalnya, saat membuat video gerakan lambat, TSM dapat digunakan untuk menyinkronkan materi audio dengan konten visual. (Werner, 1993)

Semua dalam rangka melestarikan lapangan sinyal audio ketika peregangan atau penekanan durasi, banyak prosedur TSM mengikuti strategi fundamental umum seperti pada Gambar 2.5.



Gambar 2.5 Prinsip Dasar Prosedur Time Scale Modification (Driedger, 2014)

Mengingat waktu diskrit sinyal audio asli, langkah pertama strategi ini adalah untuk membagi sinyal menjadi frame analisis singkat dengan panjang tetap. Kerangka analisis diberi jarak oleh sejumlah sampel tetap, disebut analisis hopsize $H_a \in \mathbb{N}$. Untuk mencapai modifikasi skala waktu yang sebenarnya, frame analisis kemudian temporal direlokasi untuk memiliki hopsize sintesis $H_s \in \mathbb{N}$. Ini bingkai relokasi hasil modifikasi dari durasi sinyal dengan faktor peregangan $\alpha = H_s/H_a$. Namun, hanya melapiskan frame analisis dimodifikasi biasanya menghasilkan artefak yang tidak diinginkan seperti diskontinuitas fase atau fluktuasi amplitudo. Untuk mencegah jenis artefak, frame analisis yang disesuaikan dengan bentuk frame sintesis, sebelum rekonstruksi sinyal keluaran skala waktu yang dimodifikasi. Cara untuk mendapatkan frame sintesis dari frame analisis adalah perbedaan utama antara prosedur TSM yang berbeda.

Masukan dari prosedur TSM adalah sinyal audio diskrit waktu $x: Z \rightarrow R$, dengan sampel yang sama dengan sampling rate F_s . Perhatikan bahwa meskipun sinyal audio biasanya memiliki panjang terbatas $L \in N$ sample $x(r)$ untuk $r \in [0: L-1]: = \{0, 1, \dots, L-1\}$, demi kesederhanaan, kita memodelkannya untuk mendapatkan dukungan yang tidak terbatas dengan mendefinisikan $x(r) = 0$ untuk $r \in Z \setminus [0: L-1]$.

Langkah pertama dari prosedur TSM adalah membagi x menjadi analisis singkat adalah frame $x_m, m \in Z$, masing-masing memiliki panjang sampel N (dalam literatur, kerangka analisis kadang-kadang juga disebut butiran (Zölzer, 2002). Bingkai analisis diberi jarak dengan analisis hopsize H_a :

$$x_m(r) = \begin{cases} x(r + mH_a), & \text{if } r \in \left[-\frac{N}{2} : \frac{N}{2} - 1\right], \\ 0, & \text{otherwise} \end{cases} \quad (2.2)$$

Keterangan :

$x(r)$ =sinyal masukan, m =iterasi,

H_a = analisis hopsize, x_m =bingkai analisis

N = panjang sampel, r =posisi bingkai

Pada langkah kedua, frame-frame ini direlokasi pada poros waktu berkenaan dengan hopsize sintesis spesifik H_s . Relokasi ini memperhitungkan modifikasi aktual dari skala waktu sinyal input dengan faktor peregangan $\alpha = H_s / H_a$. Karena seringkali diinginkan untuk memiliki tumpang tindih frame yang direlokasi secara spesifik, sintesis hopsize H_s lebih tinggi dijaga (pilihan umum adalah $H_s = N / 2$ or $H_s = N / 4$) dimana analisis hopsize diberikan oleh $H_a = H_s / \alpha$. Namun, hanya melapiskan bingkai yang tertunda yang tumpang tindih akan menghasilkan artefak yang tidak diinginkan seperti diskontinuitas fase pada batas bingkai dan fluktuasi amplitudo.

Oleh karena itu, sebelum rekonstruksi sinyal, kerangka analisis disesuaikan untuk membentuk bingkai sintesis y_m . Pada tahap akhir, kerangka sintesis dilapiskan untuk merekonstruksi sinyal keluaran modifikasi skala waktu yang sebenarnya $y: Z \rightarrow R$ dari prosedur TSM:

$$y(r) = \sum_{m \in Z} y_m(r - mH_s) \quad (2.3)$$

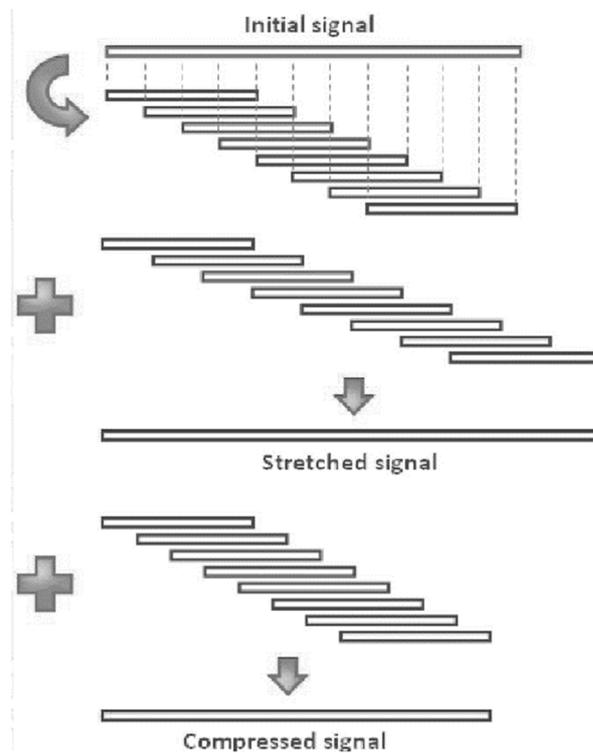
Pada Persamaan (2.3) $y(r)$ adalah sinyal keluaran, y_m bingkai sintesis, dan $H_s =$ sintesis hopsize. Meskipun strategi fundamentalisme yang tepat adalah hal yang sangat mudah, maka ada banyak pilihan dan pilihan desain yang dapat memengaruhi kualitas perseptual dari sinyal keluaran modifikasi skala waktu. Beberapa metode TSM akan dibahas pada bab selanjutnya.

2.5 Proses Sampling

Sinyal suara merupakan sinyal yang tidak terbatas dalam domain waktu (*infinite time interval*). Suara alat musik dan manusia akan menghasilkan sinyal analog yang terus kontinyu. Untuk keperluan pemrosesan dalam metode WSOLA maka sinyal suara harus dibentuk dalam potongan-potongan waktu yang terbatas (*finite time interval*). Karena itu sinyal yang ada dipotong-potong dalam slot-slot interval waktu tertentu.

2.6 Frame blocking

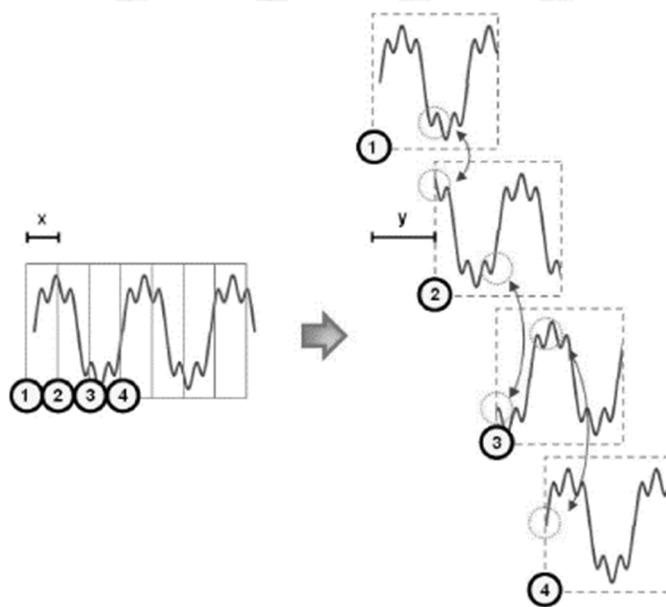
Dalam teknik ini, sinyal akan dibagi menjadi beberapa bagian kecil. sinyal ucapan dibagi menjadi urutan frame di mana setiap frame dapat dianalisis secara independen dan diwakili oleh vektor fitur tunggal karena setiap frame seharusnya memiliki sifat stasioner. Untuk membuat frame blocking adalah menggunakan jendela 20-25 ms diterapkan pada interval 10 ms (kecepatan frame 100 frame / s dan tumpang tindih antara jendela yang berdekatan sekitar 50%), Frame-frame ini mengalami overlap sebanyak 50% satu sama lain seperti pada Gambar 2.6.



Gambar 2.6 Proses frame blocking, peregangan dan pengompresan frame sinyal (virdaga,2014)

Pada Gambar 2.6, Tahap pertama sinyal asli (*initial signal*) yang utuh dipotong – potong menjadi blok-blok frame yang panjang framenya sama dan jarak antar satu framenya ke yang lainnya juga sama. Kemudian frame–frame ini diberi jarak berbeda untuk meregangkan (*stretched*) dan mengkompres (*compressed*) sinyal output dalam domain waktu.

Permasalahan timbul ketika frame–frame diberi jarak, karena menyebabkan diskontinuitas. Hal ini dapat dilihat pada Gambar 2.7 dimana jarak diantara frame pada mulanya sebesar x menjadi y . Hal ini akan menimbulkan adanya glitches yang apabila disuarakan akan terdengar oleh manusia. Diskontinuitas sinyal ucapan di tepi setiap frame dapat dikurangi dengan menerapkan pemakaian window ke masing-masing frame. Jendela yang paling umum digunakan adalah jendela Hamming. Namun pada penelitian ini dicoba dengan menggunakan beberapa window yaitu, hamming, triangular dan rectangular.

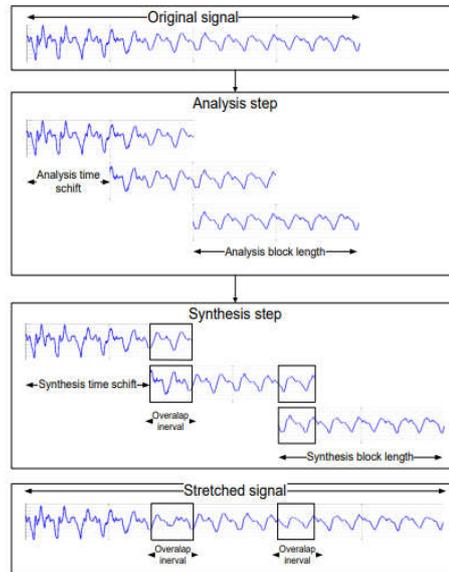


Gambar 2.7 Diskontinuitas sinyal ketika tahap peregangan dan pengkompresan dalam domain waktu.(Virdaga,2014)

2.7 Time stretching Algorithm

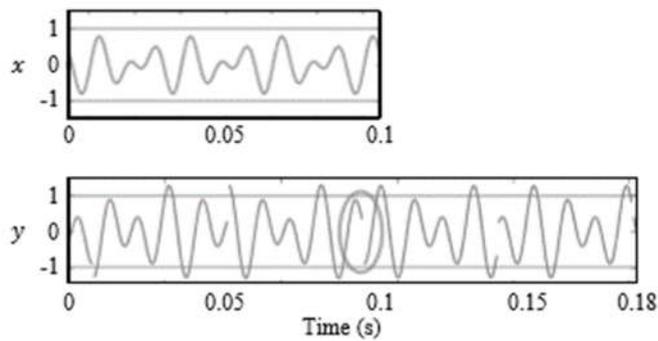
2.7.1 *Overlap and Add (OLA)*

Metode *overlap and add* (OLA) adalah metode awal untuk semua algoritma TSM. Algoritma OLA tidak menganalisis isi dari sinyal input, jadi itu berarti bahwa titik-titik sambungan frame adalah konstan, mewakili permulaan dari frame-frame sintesis yang diubah. Waktu interval tumpang tindih dihitung sebagai campuran dua bingkai. Bingkai pertama *fade-out* dan yang kedua *fade-in* dalam periode tumpang tindih. Gambar. 2.8 menyajikan skema proses peregangan sinyal menggunakan algoritma OLA. Blok asli dipisahkan dengan pergeseran waktu H_a sampel. Blok kemudian direposisi dengan pergeseran waktu $H_s = \alpha H_a$. Blok tumpang tindih sekarang tertimbang oleh *fade-in* dan fungsi *fade-out* dan frame-frame dijumlahkan. Akhirnya blok baru yang bersambung menghasilkan waktu sinyal yang utuh kembali.



Gambar 2.8 *Time Stretching : Overlap Add* (Adam Kupryjanow,2009)

Dalam OLA, frame sintesis dihitung dengan hanya windowing frame analisis dengan jendela fungsi w dan tidak memproses mereka lebih jauh. Karena titik konstan sambungan frame sintesis, sinyal OLA yang diregangkan mengandung distorsi yang tidak menyenangkan yang disebabkan oleh diskontinuitas fase dan waktu, lihat Gambar 2.9.



Gambar 2.9 Artefak khas yang terjadi saat memilih frame sintesis yang sama dengan frame analisis x_m . Sinyal masukan x diregangkan dengan faktor $\alpha = 1,8$. Sinyal

keluaran y menunjukkan diskontinuitas (oval biru) dan fluktuasi amplitudo (ditunjukkan dengan garis biru).(Driedger,dkk,2015)

Prosedur TSM dasar harus memberlakukan transisi yang mulus antara frame dan juga mengkompensasi fluktuasi amplitudo yang tidak diinginkan. Gagasan tentang prosedur TSM overlap-add (OLA) adalah menerapkan fungsi jendela pada kerangka analisis, sebelum rekonstruksi sinyal keluaran y . Tugas fungsi jendela adalah menghapus diskontinuitas bentuk gelombang mendadak pada batas-batas frame analisis (Driedger,dkk,2015). Pilihan khas untuk w adalah fungsi jendela Hann seperti pada Persamaan (2.4) dan (2.5):

$$w(r) = \begin{cases} 0.5 \left(1 - \cos \left(\frac{2\pi \left(r + \frac{N}{2} \right)}{N-1} \right) \right), & \text{if } r \in \left[-\frac{N}{2} : \frac{N}{2} - 1 \right], \\ 0, & \text{otherwise.} \end{cases} \quad (2.4)$$

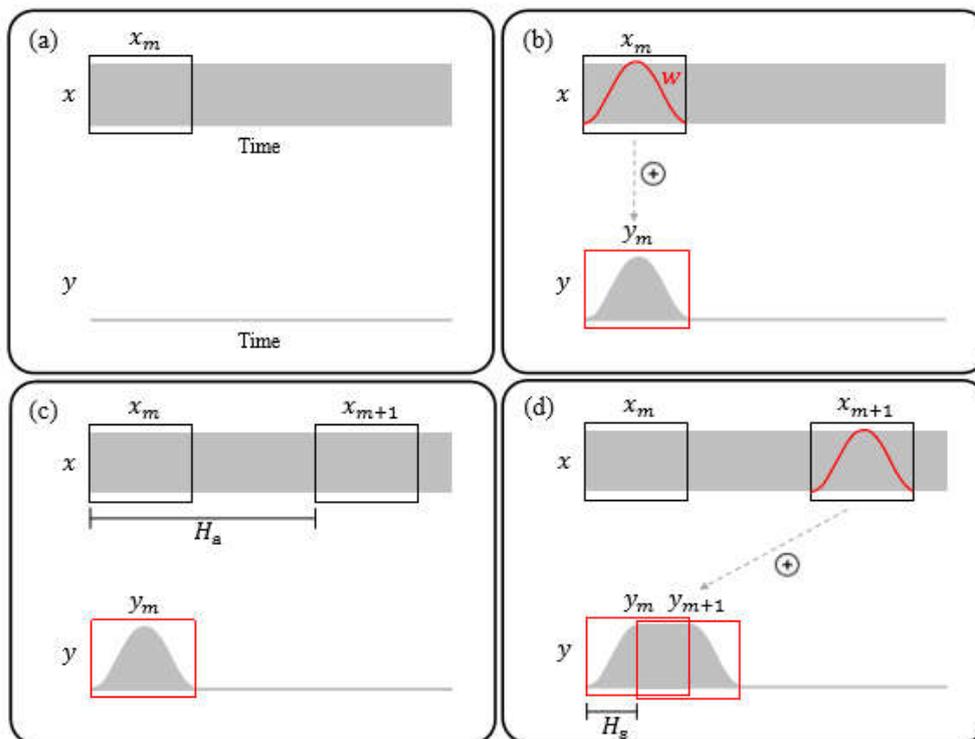
$$\sum_{n \in Z} w \left(r - n \frac{N}{2} \right) = 1, \quad (2.5)$$

untuk semua $r \in Z$. Prinsip prosedur OLA iteratif divisualisasikan pada Gambar 2.4. Untuk indeks frame $m \in Z$, pertama-tama kita menggunakan Persamaan (2.2) untuk menghitung kerangka analisis x_m (Gambar 2.4a). Kemudian, kita mendapatkan bingkai sintesis y_m seperti pada Persamaan (2.6):

$$y_m(r) = \frac{w(r)x_m(r)}{\sum_{n \in Z} w(r-nH_s)} \quad (2.6)$$

Nominator Persamaan (2.6) merupakan windowing aktual dari kerangka analisis dengan mengalikannya secara pointwise dengan fungsi jendela yang diberikan. Denominator menormalkan frame dengan jumlah fungsi jendela yang tumpang tindih, yang mencegah fluktuasi amplitudo pada sinyal output. Perhatikan bahwa, ketika memilih jendela Hann dan $H_s = N / 2$, penyebut selalu dikurangi menjadi satu dengan Persamaan (2.5). Ini adalah kasus pada Gambar 2.10b dimana amplitudo frame sintesis tidak diskalakan sebelum ditambahkan ke sinyal keluaran y . Melanjutkan ke frame analisis berikutnya x_{m+1} , (Gambar 2.10c), frame ini lagi

berjendela, tumpang tindih dengan bingkai sintesis sebelumnya, dan ditambahkan ke sinyal output (Gambar 2.10d). Perhatikan bahwa Gambar 2.10 memvisualisasikan kasus di mana sinyal asli dikompres ($H_a > H_s$). Peregangan sinyal ($H_a < H_s$) bekerja dengan cara yang persis sama. Dalam kasus ini, kerangka analisis tumpang tindih ke tingkat yang lebih besar daripada bingkai sintesis.



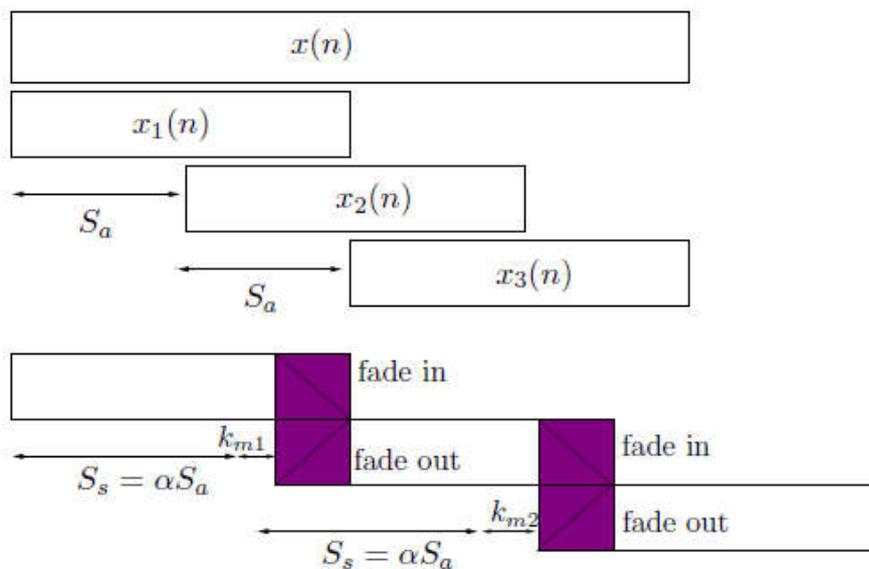
Gambar 2.10. Prinsip TSM berdasarkan overlap-add (OLA). (a) Masukan sinyal audio x dengan bingkai analisis x_m . Sinyal keluaran y dibuat secara iteratif; (b) Penerapan fungsi jendela Hann ke kerangka analisis x_m yang menghasilkan bingkai sintesis y_m ; (c) Kerangka analisis berikutnya $x_m + 1$ memiliki jarak sampel H_a yang ditentukan dari x_m ; (d) Tumpang tindih-tambahkan dengan menggunakan hopsize sintesis spesifik H_s . (Driedger,dkk,2015)

OLA adalah contoh prosedur TSM domain waktu dimana modifikasi pada kerangka analisis diterapkan secara murni dalam domain waktu. Secara umum, prosedur

TSM domain-waktu tidak hanya efisien tetapi juga menjaga timbre dari sinyal input ke tingkat yang tinggi.

2.7.2 SOLA (Synchronous Overlap and Add)

Synchronous Overlap and Add sintesis sangat mirip dengan prosedur OLA umum yang disajikan sebelumnya. Perbedaan utama antara keduanya adalah bahwa SOLA bergantung pada teknik korelasi untuk memperbaiki algoritma waktu-peregangan. Ketika blok dialihkan oleh faktor waktu α kesamaan di daerah interval *overlap* dicari untuk diskrit-waktu lag kesamaan maksimal. Hal ini kesamaan maksimum blok *overlap* kemudian ditimbang dengan fungsi fade-out fade-in dan lagi dijumlahkan sampel-by-sampel. Sebuah penggambaran ini diberikan pada Gambar 2.11.



Gambar 2.11 Time Stretching : Synchronous Overlap and Add

2.7.3 PSOLA (Pitch Synchronous Overlap and Add)

Pitch Synchronous Overlap and Add menggunakan hipotesis bahwa suara masukan ditandai dengan pitch. Ini mengeksplorasi pengetahuan dari pitch untuk sinkronisasi segmen waktu dengan benar untuk menghindari diskontinuitas pitch. Algoritma PSOLA pada dasarnya dibagi menjadi dua langkah: tahap pertama

analisis segmen suara input dan ekstrak informasi pitch, dan tahap kedua sintesis waktu versi membentang oleh *overlap* dan menambahkan segmen waktu diekstraksi dengan tahap analisis.

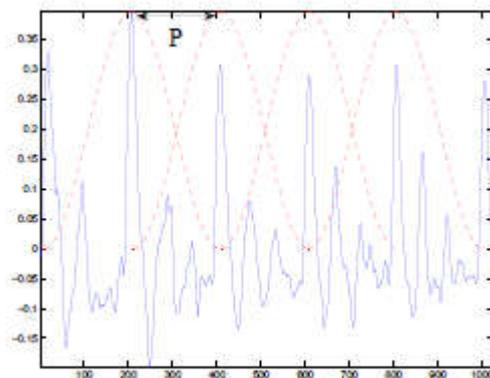
Algoritma Analisis:

1. Tentukan periode pitch sinyal input dan dari waktu instants (tanda pitch) t_i .
2. Ekstrak segmen berpusat di masing-masing pitch mark t_i dengan menggunakan jendela Hanning dengan panjang $L_i = 2P(t_i)$. Periode pitch dua ini memastikan bahwa fade-in dan fade-out dapat terjadi.

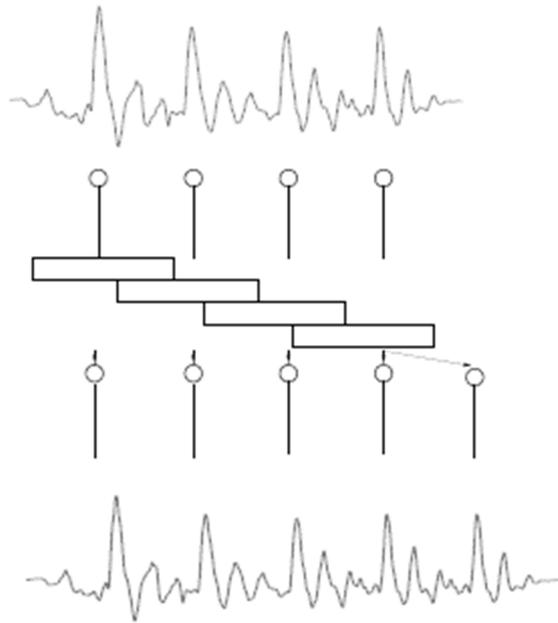
Algoritma sintesis :

1. Pilih segmen analisis yang meminimalkan waktu jarak $|\alpha_k - t_k|$
2. *Overlap and add* segmen yang dipilih. Perhatikan bahwa ini akan menghasilkan beberapa segmen masukan yang berulang $\alpha > 1$ dan beberapa segmen yang dibuang $\alpha < 1$.
3. Menentukan waktu berikutnya instan $t_k + 1$ di mana segmen sintesis berikutnya akan dipusatkan.

Algoritma PSOLA digambarkan pada Gambar 2.12 dan Gambar 2.13



Gambar 2.12 PSOLA : Analisis pitch dan *block window*



Gambar 2.13 Depiction of the PSOLA algorithm.

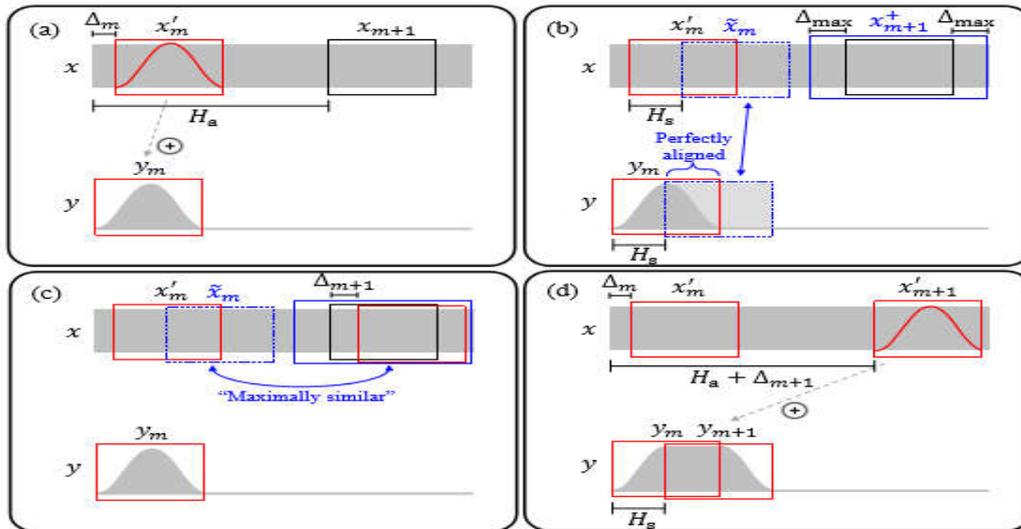
2.7.4 WSOLA (Waveform Similarity Overlap-Add)

Salah satu cara untuk menghindari diskontinuitas fase seperti yang diperkenalkan oleh OLA adalah memilih frame analisis sehingga sintesis berturut frame lebih cocok bersama-sama saat menambahkan mereka. Bingkai analisis windowed disalin dari posisi tetap pada sinyal input ke posisi tetap pada sinyal output. Dengan kata lain, sinyal input tidak memengaruhi prosedur. Salah satu strategi domain waktu untuk mengurangi artefak pada fase loncatan yang disebabkan oleh OLA. *The Waveform Similarity Overlap-Add algoritma* (WSOLA) (Verhelst,1993) mencapai ini dengan memperkenalkan toleransi analisis posisi *frame* Δ_{Max} .

Ide utamanya adalah menyesuaikan bingkai sintesis berturut-turut sedemikian rupa sehingga struktur periodik dalam bentuk gelombang frame sejajar di daerah yang tumpang tindih. Pola periodik pada sinyal input dipelihara dalam output. Asumsikan bahwa pada iterasi m maka posisi kerangka analisis x_m digeser oleh

sampel $\Delta_m \in [-\Delta_{max}: \Delta_{max}]$. Kami menyebut bingkai ini kerangka analisis x'_m yang di sesuaikan.

$$x'_m(r) = \begin{cases} x(r + mH_a + \Delta_m), & \text{if } r \in \left[-\frac{N}{2}; \frac{N}{2} - 1\right], \\ 0, & \text{otherwise.} \end{cases} \quad (2.7)$$



Gambar 2.14. Prinsip Kesamaan Bentuk Gelombang Overlap-Add (WSOLA). (a) Masukan sinyal audio x dengan kerangka analisis x'_m yang disesuaikan. Bingkai itu sudah dilipat dan disalin ke sinyal keluaran y ; (b, c) Retrieval bingkai dari daerah bingkai diperluas x'_{m+1} (kotak biru solid) yang sama mirip dengan bentuk alami \tilde{x}_m (kotak biru putus-putus) dari kerangka analisis yang disesuaikan x'_m ; (d) Kerangka analisis yang disesuaikan x'_{m+1} dilambangkan dan disalin ke sinyal keluaran y . (Drieger, 2015)

Kerangka analisis yang disesuaikan x'_m dilambangkan dan disalin ke sinyal keluaran y seperti OLA. Sekarang kita perlu menyesuaikan posisi frame analisis berikutnya x_{m+1} . Pada tahapan ini masalah optimasi dibatasi yang bertujuan untuk menemukan indeks pergeseran yang optimal $\Delta_{m+1} \in [-\Delta_{max}: \Delta_{max}]$ sehingga struktur periodik dari kerangka analisis yang disesuaikan x'_{m+1} secara optimal disesuaikan dengan struktur kerangka sintesis y_m yang sebelumnya disalin pada

daerah tumpang tindih saat melapiskan kedua frame pada sintesis hopsize H_s . Kerangka ideal x_m^{\sim} dari kerangka analisis x_m' yang disesuaikan (kotak biru putus-putus pada Gambar 5b) akan menjadi pilihan optimal untuk kerangka analisis yang disesuaikan x_{m+1}' dalam skenario yang tidak dibatasi. (Driedger,2015)

Kerangka ideal x_m^{\sim} seperti pada Persamaan (2.8) :

$$x_m^{\sim}(r) = \begin{cases} x(r + mH_a + \Delta m + H_s), & \text{if } r \in \left[-\frac{N}{2}; \frac{N}{2} - 1\right], \\ 0, & \text{otherwise} \end{cases} \quad (2.8)$$

Ini adalah kasusnya, karena frame analisis yang disesuaikan x_m' dan frame sintesis y_m pada dasarnya sama (hingga windowing). Akibatnya, struktur kerangka ideal x_m^{\sim} selaras dengan struktur kerangka sintesis y_m saat melapiskan dua frame pada sintesis hopsize H_s (Gambar 5b). Namun, karena batasan $\Delta_{m+1} \in [-\Delta_{max}; \Delta_{max}]$, frame x_{m+1}' yang disesuaikan harus berada di dalam area bingkai real x_{m+1}^+ (kotak biru solid pada Gambar 5b) seperti pada Persamaan (2.9):

$$x_{m+1}^+(r) = \begin{cases} x(r + (m + 1)H_a), & \text{if } r \in \left[-\frac{N}{2} - \Delta_{max}; \frac{N}{2} - 1 + \Delta_{max}\right], \\ 0, & \text{otherwise} \end{cases} \quad (2.9)$$

Tujuannya adalah untuk mengambil frame yang disesuaikan x_{m+1}' sebagai frame dalam x_{m+1}^+ yang gelombangnya paling mirip dengan x_m^{\sim} , kita harus mendefinisikan ukuran yang menguantifikasi kesamaan dua bingkai. Satu pilihan yang mungkin untuk matrik ini adalah korelasi silang seperti pada Persamaan(2.10):

$$c(q, p, \Delta) = \sum_{r \in Z} q(r)p(r + \Delta) \quad (2.10)$$

dari sinyal q dan sinyal p digeser oleh sampel $\Delta \in Z$. Kita kemudian dapat menghitung indeks pergeseran optimal Δ_{m+1} , yang memaksimalkan korelasi silang x_m^{\sim} dan x_{m+1}^+ dengan persamaan Δ_{m+1} seperti pada persamaan (2.11):

$$\Delta_{m+1} = \underset{\Delta \in [-\Delta_{max}; \Delta_{max}]}{\operatorname{argmax}} c(x_m^{\sim}, x_{m+1}^+, \Delta), \quad (2.11)$$

Indeks pergeseran Δ_{m+1} menentukan posisi kerangka analisis yang disesuaikan x'_{m+1} di dalam area bingkai real x_{m+1}^+ (Gambar 5c). Akhirnya sama seperti OLA, kita menghitung kerangka sintesis y_{m+1} seperti pada Persamaan (2.12):

$$y_{m+1}(r) = \frac{w(r)x(r+(m+1)H_a+\Delta_{m+1})}{\sum_{n \in \mathbb{Z}} w(r-nH_s)} \quad (2.12)$$

dan gunakan Persamaan (2.3) untuk merekonstruksi sinyal keluaran y (Gambar 5d). Dalam prakteknya, kita memulai proses optimasi iteratif pada indeks frame $m = 0$ dan asumsikan $\Delta_0 = 0$.

2.8 Window

Konsep windowing sangat penting dalam merancang sistem LTI maupun dalam analisis spektrum sinyal[3]. Windowing adalah operasi pengambilan sinyal $x[n]$ dan mengalikannya dengan sinyal window $w[[n]$ dengan selang waktu terbatas. Yaitu pada persamaan :

$$p[n] = x[n]w[n] . \quad (2.13)$$

$P[n]$ adalah selang waktu terbatas.

Pentingnya windowing dalam analisis spektrum berasal dari fakta bahwa dalam aplikasi yang sangat banyak diperlukan untuk menghitung transformasi Fourier dari sinyal yang telah diukur. Selama dalam Selamatdalam praktis kita dapat mengukur sinyal $x[n]$ hanya melewati interval waktu yang terbatas (time window), sinyal aktual yang tersedia untuk analisis spektrum adalah pada persamaan (2.14)

$$p[n] = \begin{cases} x\{n\}, & -M \leq n \leq M \\ 0, & \text{selain itu} \end{cases} \quad (2.14)$$

Di mana $-M \leq n \leq M$ adalah time window. Dengan demikian, $p[n]=x[n]w[n]$. Ada beberapa jenis windowing, dalam paper ini akan digunakan 3 window yaitu Hamming,, rectangular dan triangular.

2.8.1 *window Hamming*

Richard W. Hamming mengamati bahwa lobus sisi jendela rectangular dan hanning adalah fase terbalik relatif satu sama lain, sehingga kombinasi linear dari keduanya akan cenderung menyebabkan mereka saling membatalkan. Dia mencari kombinasi linear yang meminimalkan amplitudo lobus sisi maksimum dan muncul dengan formulasi berikut, yang mewakili kosinus dibangkitkan pada alas persegi panjang (Carnegie,):

$$w[n] = \begin{cases} 0,54 - 0,46 \cos\left(\frac{2\pi n}{M}\right), & 0 \leq n \leq M \\ 0, & \end{cases} \quad (2.15)$$

2.8.2 *window rectangular*

Jendela persegi panjang kadang-kadang dikenal sebagai jendela Dirichlet. Ini adalah jendela yang paling sederhana, mengambil sepotong sinyal tanpa modifikasi lain sama sekali, yang mengarah ke diskontinuitas di titik akhir (kecuali jika sinyal kebetulan tepat untuk panjang jendela, seperti yang digunakan dalam pengujian multitone, misalnya) . Sisi-lobus pertama hanya 13 dB lebih rendah dari lobus utama, dengan sisanya jatuh sekitar 6 dB per oktaf. (Pallone,1999)

$$w[n] = \begin{cases} 1, & 0 \leq n \leq M \\ 0, & \text{selain itu} \end{cases} \quad (2.16)$$

2.8.3 *window Triangular*

window triangular disebut juga dengan window Bartlett karena Karena jendela Bartlett dapat dianggap telah diperoleh dari mengkonvolusi dua ukuran setengah lebar window rectangular persegi, transformasinya mudah mengkuadratkan transformasi jendela persegi panjang:

$$w[n] = \begin{cases} 1 - \left| \left(\frac{2n}{M} \right) - 1 \right|, & 0 \leq n \leq M \\ 0, & \text{selain itu} \end{cases} \quad (2.17)$$

2.9 Fast Fourier Transform (FFT)

Deret fourier dipakai sebagai perangkat untuk menghitung spektrum dari sinyal periodik. Untuk sinyal bukan periodik, perangkat yang digunakan adalah transformasi fourier. Definisi dari deret Fourier eksponensial kompleks digambarkan dalam Persamaan 2.18:

$$y(t) = \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega_0 t} \quad (2.18)$$

Keterangan:

- $\omega_0 = 2\pi f_0 = 2\pi/T_0$ disebut frekuensi sudut fundamental
- c_n = koefisien Furier, ditulis dalam Persamaan 2.19
- t_0 = parameter yang bebas dipilih

$$c_n = \frac{1}{T_0} \int_{t_0}^{t_0+T_0} y(t) e^{-jn\omega_0 t} dt \quad (2.19)$$

Sedangkan untuk transformasi fourier yang umum digunakan untuk merubah sinyal dari domain waktu ke domain frekuensi digambarkan dalam Persamaan 2.20:

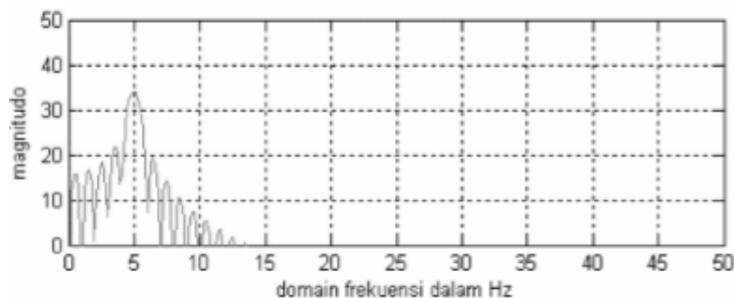
$$X(\omega) = F\{x(t)\} = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \quad (2.20)$$

Persamaan 2.20 merupakan bentuk transformasi Fourier yang siap dikomputasi secara langsung dari bentuk sinyal $x(t)$. Dimana $x(t)$ merupakan fungsi yang tidak periodik terhadap waktu t . Jika $x(\omega)$ diketahui, maka dapat diperoleh nilai $x(t)$ dari persamaan invers transformasi fourier seperti pada Persamaan 2.21:

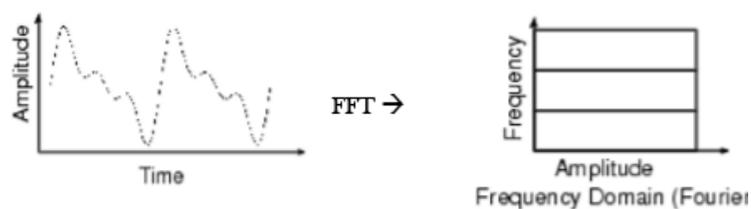
$$x(t) = f^{-1}\{X(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} x(\omega) e^{j\omega t} dt \quad (2.21)$$

FFT merupakan metode yang sangat efisien untuk menghitung DFT secara komputasional. Sebagai akibatnya penginterpretasian hasil FFT hanya memerlukan pemahaman dari DFT. Efisiensi komputasional FFT muncul dari kepandaian menyusun kembali suku-suku dalam DFT sedemikian hingga suku-suku yang sama hanya dihitung sekali.

Untuk mengetahui frekuensi suatu sinyal diperkenalkan metode yang dikenal sebagai *Fast Fourier Transform* atau disingkat sebagai FFT. Transformasi Fourier ini dilakukan untuk mentransformasikan sinyal dari domain waktu ke domain frekuensi. Salah satu algoritma FFT yang terkenal diciptakan oleh Cooley dan Tukey. Proses FFT dengan algoritma Cooley-Tukey dilakukan dengan cara membagi sinyal masukan menjadi bagian-bagian kecil. Kemudian sistem melakukan Transformasi *Fourier* terhadap bagian-bagian kecil tersebut, lalu pada akhirnya hasil Transformasi *Fourier* yang sudah dilakukan, digabungkan menjadi satu. Contoh hasil FFT adalah seperti yang terlihat pada Gambar 2.15 dan model analisisnya ditunjukkan dalam Gambar 2.16:



Gambar 2.15 Sinyal sinus dalam domain frekuensi



Gambar 2.16 Analisis FFT (fera,2018)

Penghitungan FFT dilakukan pada 44100 sample suara yang telah diambil dengan perintah `fft`. Untuk mencari frekuensi dasar sinyal masukan, maka dilakukan pencarian frekuensi yang memiliki magnitudo terbesar pada spektrum FFT. Nilai frekuensi dibutuhkan sebagai pembanding antara rekaman suara asli dengan rekaman suara yang sudah diberikan metode WSOLA.

2.10 PESQ (*Perceptual Evaluation of Speech Quality*)

PESQ adalah standar internasional untuk pengukuran kualitas obyektif obyektif dan itu memprediksi hasil tes mendengarkan subyektif pada sistem telepon. PESQ adalah hasil integrasi dari *Perceptual Analysis Measurement System* (PAMS) dan PSQM99, versi yang disempurnakan dari *Perceptual Speech Quality Measur*(PSQM) (Ebem,2009). *International Telecommunication Union* (ITU-T) telah menerima PESQ sebagai rekomendasi P.862 dan merupakan metode pengukuran persepsi obyektif yang paling dikenal. Untuk mengukur kualitas bicara, PESQ menggunakan model sensorik untuk membandingkan sinyal asli yang belum diolah dengan versi terdegradasi pada output dari sistem komunikasi (Malden,2004).

PESQ terdiri dari model perseptual dan kognitif digunakan untuk membandingkan output perangkat yang diuji dengan sinyal ucapan input atau sinyal referensi . Filosofi dasar yang digunakan dalam pengembangan metode pengukuran berdasarkan persepsi obyektif diberikan di atas gambar 2.1 (Ebem,2009). Dengan membandingkan referensi dan sinyal tes, dapat memperoleh skor kualitas dan skor ini mirip dengan subyektif "*Mean Opinion Score*" (MOS) yang merupakan tes panel. Skor PESQ dikalibrasi menggunakan database tes subyektif yang besar. PESQ menggabungkan banyak perkembangan dari model sebelumnya untuk menilai kualitas ucapan seperti PSQM dan *Measuring Normalizing Blocks*(MNB) (ITU-T-P.861). Perkembangan ini telah menyebabkan PESQ digunakan dengan percaya diri untuk menilai kualitas ujaran akhir-ke-akhir serta efek dari elemen-elemen individual seperti pengkodean lossy, dan pendeteksian aktivitas suara.

Kunci untuk proses ini adalah transformasi dari keduanya asli dan sinyal terdegradasi ke representasi internal yang analog dengan representasi psikofisik sinyal audio dalam sistem pendengaran manusia, dengan memperhitungkan frekuensi perseptual (*Bark*) dan kenyaringan (*Sone*). Ini dicapai dalam beberapa tahap: penyelarasan waktu, tingkat keselarasan ke tingkat pendengaran yang dikalibrasi, pemetaan frekuensi waktu, frekuensi warping, dan skala pengerasan kompres. PESQ mempertimbangkan sumber degradasi sinyal yang berbeda seperti

distorsi pengkodean, kesalahan, kehilangan paket, penundaan dan penundaan variabel, dan penyaringan dalam komponen jaringan analog. PESQ tidak memperhitungkan efek subjektif dari perubahan level dalam jaringan, gempa, dan efek penundaan bolak-balik pada percakapan.

2.11 Korelasi

Korelasi merupakan salah satu alat matematika yang biasa digunakan pada pengolahan sinyal untuk analisis fungsi atau deret nilai, seperti sinyal domain waktu. Korelasi biasanya digunakan untuk menemukan pola pengulangan dari suatu sinyal, seperti menentukan periode sinyal atau untuk identifikasi frekuensi dasar dari sinyal yang tidak memiliki komponen frekuensi tetapi memiliki frekuensi yang harmonis.(Risny,2010). Korelasi terdiri dari dua macam, yaitu autokorelasi, dan korelasi silang. Autokorelasi merupakan korelasi silang antara suatu sinyal dengan sinyal itu sendiri. Misalnya sinyal $f(t)$, autokorelasi kontinyu $R_{ff}(\tau)$ selamanya diartikan sebagai integral korelasi silang kontinyu dari $f(t)$ itu sendiri.

$$\begin{aligned} R_{ff}(\tau) &= f * (-\tau) \odot f(\tau) = \int_{-\infty}^{\infty} f(t + \tau) f * (t) dt \\ &= \int_{-\infty}^{\infty} f(t) f * (t - \tau) dt \end{aligned} \quad (2.22)$$

Pada Persamaan 2.22 , dimana $f *$ menunjukkan conjugate dari bilangan kompleks serta \odot menunjukkan konvolusi. Untuk fungsi real, $f * = f$.

Autokorelasi diskrit ditunjukkan dengan Persamaan 2.17:

$$R_{xx}(j) = \sum_n (x_n)(x_n^* - j) \quad (2.23)$$

Korelasi dua buah sinyal, dimana untuk setiap sampel n akan diperoleh $X_1(k)$ sebagai hasil sinyal pertama dari $x_1(n)$ dan $X_2(k)$ sebagai hasil sinyal kedua dari $x_2(n)$ untuk setiap frekuensi k, dapat dihitung melalui Persamaan 2.24:

$$R_{x_1x_2}(k) = \frac{1}{N-|k|} \sum_{i=0}^{N-1} X_1(i + |k|) X_2(i) \quad (2.24)$$

Nilai korelasi yang diperoleh akan digunakan untuk mencari sinyal yang mirip yang nantinya akan diakan diberi skala waktu dan dikeluarkan pada sinyal output hasil WSOLA.

BAB 3

METODOLOGI PENELITIAN

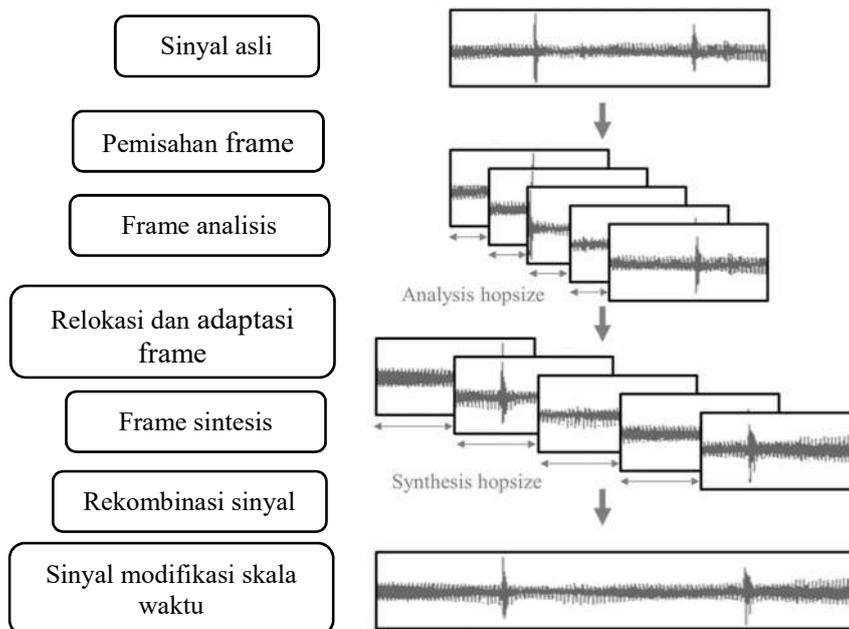
Pada bab ini dijelaskan garis besar langkah-langkah penelitian Metodologi penelitian ini disusun agar penelitian fokus pada rencana awal.

Penelitian ini yaitu tentang modifikasi skala waktu menggunakan *Waveform Similarity overlap and Add* (WSOLA), dimana metode ini mempertahankan frekuensi rekaman suara walaupun skala waktu dipercepat atau diperlambat. Tahap pertama yang dilakukan adalah pengambilan data yang bersumber dari perangkat instrument gamelan yang ada di Elektro Budoyo ITS. Instrumen gamelan yang digunakan pada penelitian ini adalah saron. Karena nada-nada dari set gamelan diciptakan dengan frekuensi dasar berbeda-beda sehingga nantinya gamelan akan memiliki nada dasar yang bervariasi, maka tahap selanjutnya adalah proses analisis *Fast Fourier Transform* (FFT) yang digunakan untuk mengetahui frekuensi dasar saron yang digunakan sebagai data penelitian, yang kemudian frekuensi dasar ini bisa digunakan sebagai acuan pada sinyal referensi yang merupakan sinyal sintetik.

Tahap selanjutnya adalah memodifikasi skala waktu rekaman suara menggunakan *Waveform Similarity Overlap and Add* (WSOLA). *Waveform Similarity Overlap and Add* (WSOLA) digunakan untuk mempertahankan frekuensi rekaman suara walaupun skala waktu dipercepat atau diperlambat. Sinyal keluaran dari penggunaan metode WSOLA akan dilihat frekuensi maksimum dari sinyal tersebut dan akan dibandingkan dengan frekuensi sinyal masukan. Untuk mengoptimalkan kinerja dari system maka dilakukan optimasi dengan menganalisa parameter-parameter yang berpengaruh pada system. Selanjutnya dari analisa tersebut, didapatkan nilai terbaik. Nilai terbaik yang dimaksud adalah yang mempunyai nilai *Mean Square Error* (MSE) terendah. Sehingga parameter-parameter tersebut bisa dijadikan *variable control* untuk mengoptimalkan kinerja

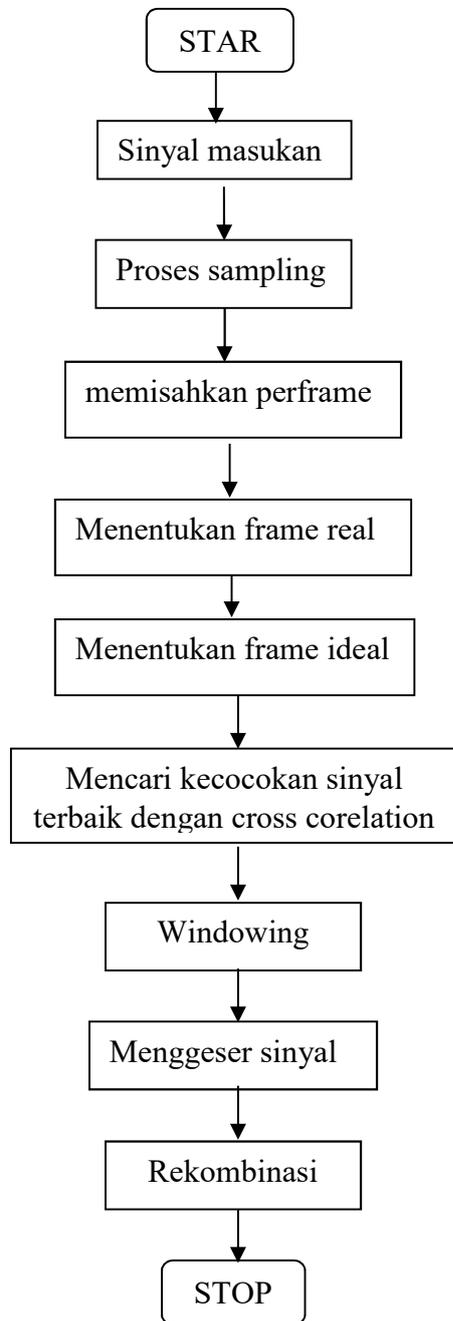
system. Sistematika umum dari metodologi penelitian secara umum ini bisa dilihat pada Gambar 3.1.

3.1 Perancangan Sistem



Gambar 3.1 *Prosedure time scale modification*

Gambar 3.1 merupakan gambar prosedur dari proses time scaling modification secara umum. Pada awalnya sinyal input di pecah menjadi frame-frame. Setiap frame yg telah di pecah tersebut memiliki nilai jarak pergeseran yang disebut analisis hop size yang di miliki antara 1 frame dengan frame selanjutnya. Kemudian diberi nilai skala jarak untuk tumpang tindih sesuai yang kita inginkan yang disebut nilai sintesis hop size. Setelah frame- frame tersebut bergeser, frame yang tadinya terpisah dinyatukan kembali, sehingga membentuk sinyal yang utuh kembali dalam keadaan sudah bergeser. Proses metode WSOLA dapat dilihat pada Gambar 3.2 danb erikut akan dijelaskan lebih rinci:

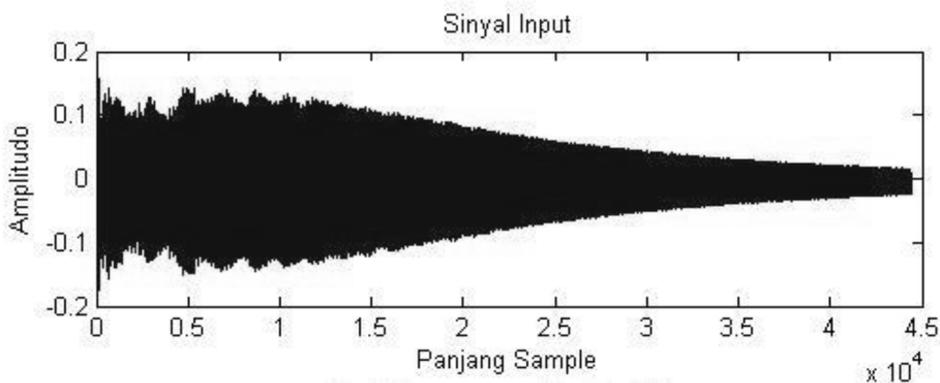


Gambar 3.2 Proses metode WSOLA

Tahap 1 Sinyal masukan :

1. Sinyal input

Pada percobaan ini, digunakan suara pada umumnya yaitu suara saron dan suara manusia. Tiap-tiap suara yang dihasilkan dapat dilakukan dengan berbagai macam variasi. Pada gambar 3.3 merupakan salah satu contoh gambar sinyal *input*. Parameter yang menjadi nilai dimasukkan seperti nilai panjang frame analisis, pergeseran frame analisis, frame offset maksimum dan tipe window.



Gambar 3.3 Sinyal masukan

2. Proses Sampling

Pada proses awal dari pergeseran kerapatan waktu suara ini adalah program sampling. Yang mana persyaratan frekuensi sampling menurut teorema Shannon harus samadengan atau melebihi 2 kali frekuensi sinyal yang disample. Pada sistem ini menggunakan $F_s = 44100$ Hz, karena frekuensi tersebut telah sesuai dengan teori sampling Nyquist yaitu $f_{\text{sampling}} \geq 2 f_{\text{sinyal}}$. Parameter panjang frame analisis (L_m), pergeseran frame analisis (R_m), dan frame offset maksimum atau nilai toleransi (deltamax) perlu dikonversi dari satuan msec ke satuan sampel pada tingkat sampling sampel f_s per detik, dengan menggunakan relasi:

1. $L = \text{round}(L_m * f_s / 1000)$ (sampel),
2. $R = \text{round}(R_m * f_s / 1000)$ (sampel),
3. $\text{delta} = \text{round}(\text{deltamax} * f_s / 1000)$ (sampel).

3. Frame Blocking

Proses frame blocking untuk mempermudah dalam penganalisaan sinyal pada proses pengambilan fitur. Pada proses ini dilakukan pemotongan sinyal dalam slot-slot tertentu agar memenuhi 2 syarat yaitu linear dan time invariant. Setiap potongan tersebut disebut frame. Pada penelitian ini sinyal suara dipotong sepanjang yang sesuai diinginkan peneliti, maka panjang perframe menjadi inputan. Contohnya sinyal suara dipotong 40 milidetik. Jadi dalam satu frame terdapat 1764 sampel dari 44100 sampling rate yang ada. Setiap satu frame dengan frame lainnya memiliki jarak pergeseran sebesar yang diinginkan peneliti, contohnya 10 milidetik yang dikonversi kedalam sample menjadi 160 sample. Biasanya jarak pergeseran ini disebut hop size analysis.

4. Frame real

Sebuah array dari sampel L panjang nominal. Array ini menjelaskan perubahan kecepatan dengan menggeser nilai frame awal dengan sampel $\alpha * R$ (dari frame sebelumnya), seperti yang dibutuhkan untuk mengubah tingkat sinyal. Untuk menjelaskan kesalahpahaman sampel ucapan pada batas-batas bingkai, karena perubahan tingkat, susunan x_{real} yang ditarik oleh sampel δ pada bagian bawah dari balok dan bagian dari susunan. Ukuran x_{real} array yang diperluas adalah sampel $L + 2 * \delta$.

5. Frame ideal

sebuah array dengan panjang sampel L yang idealnya tumpang tindih ditambahkan pada frame saat ini tanpa perubahan kecepatan, yaitu untuk setiap frame, sampel awal untuk array x_{ideal} digeser oleh sampel R dari kecocokan terbaik dari frame sebelumnya.

6. Kecocokan sinyal terbaik dengan cross correlation

Untuk menyelaraskan kedua pasangan array, frame ideal dan frame real, kedua array dikorelasi silang. Indeks korelasi maksimum (maxind) digunakan sebagai indeks awal dari array frame real yang akan diperluas pada iterasi berikutnya, dan sampel ucapan dari frame real (maxind: maxind + L-1) disimpan dalam array (xadd).

7. Windowing

Menerapkan fungsi jendela pada kerangka analisis, sebelum rekonstruksi sinyal keluaran. Tugas fungsi jendela adalah menghapus diskontinuitas bentuk gelombang mendadak pada batas-batas frame analisis.

8. Mengeser sinyal

Overlap and add segmen yang dipilih. Perhatikan bahwa ini akan menghasilkan beberapa segmen masukan yang berulang $\alpha > 1$ dan beberapa segmen yang dibuang $\alpha < 1$.

9. Rekombinasi

Segmen yang terpisah kemudian dapat digabungkan dengan menempatkan pusat mereka di posisi sinyal asli dan menambahkan daerah yang tumpang tindih.

3.2 Pelaksanaan Uji Coba

Uji coba dilakukan pada sample suara gamelan saron dan manusia. Kemudian suara tersebut akan dipercepat dan diperlambat waktunya menggunakan metode WSOLA, kemudian dihitung maksimal suara tersebut dapat digeser.

3.3 Evaluasi Kerja

Evaluasi kerja dilakukan dengan menganalisis kinerja. Dari evaluasi kerja ini akan dibandingkan metode mana yang lebih baik antara WSOLA dengan metode modifikasi skala waktu lainnya dan kemudian dirumuskan

kesimpulan mengenai hasil penelitian yang telah dilakukan dan hal-hal yang perlu dikembangkan untuk penelitian selanjutnya.

3.4 Hasil yang diharapkan

Pada penelitian ini diharapkan sistem mampu melakukan pergeseran waktu dengan kondisi frekuensi tetap. Menghasilkan metode yang lebih baik dari metode yang dibandingkan.

3.5 Dokumentasi

Pada tahap ini akan dilakukan dokumentasi sistem dalam bentuk penulisan laporan hasil penelitian dari setiap tahapan yang dilakukan. Tujuan dari tahapan ini adalah menghasilkan dokumentasi tertulis dari penelitian yang telah dilakukan.

Halaman ini sengaja dikosongkan

BAB 4

HASIL PENELITIAN SEMENTARA

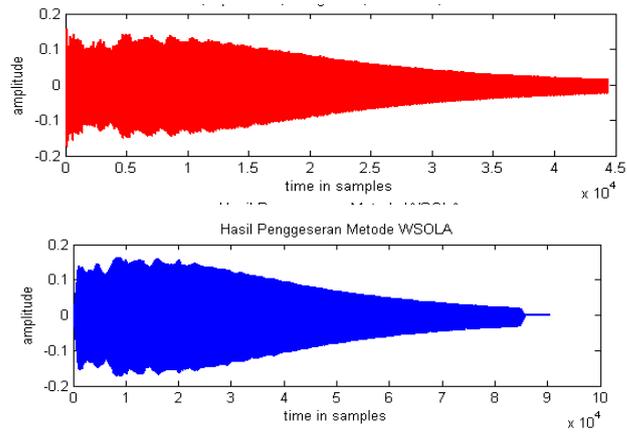
Penelitian yang telah dilakukan saat ini adalah mencoba menggunakan metode Waveform Similarity Overlap and Add(WSOLA) dan metode Pitch Synchronous Overlap and Add(PSOLA).

4.1 Sinyal Masukan

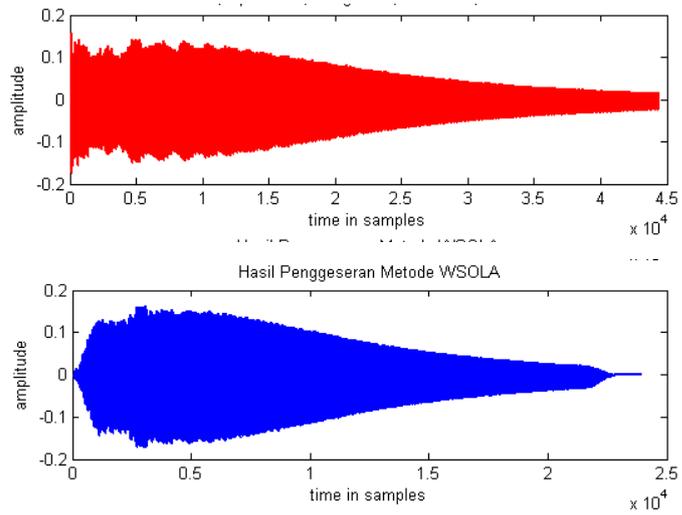
Pada pengujian ini digunakan rekaman suara sinyal gamelan saron lima slendro satu ketukan, saron satu slendro satu ketukan, saron dua slendro satu ketukan, saron, saron lima slendro dua ketukan, saron nada satu dan dua (dua ketukan beda nada) dan suara manusia. Masing –masing rekaman suara yang diuji memiliki sampling rate yang berbeda2 yaitu 44100, 48000 dan 16000. Pada rekaman suara manusia, saya merekam dengan mengabaikan noise sesuai dengan tujuan saya suara yang ingin dihasilkan memang dalam bentuk asli namun yang sudah dipercepat dan diperlambat.

4.2 WSOLA

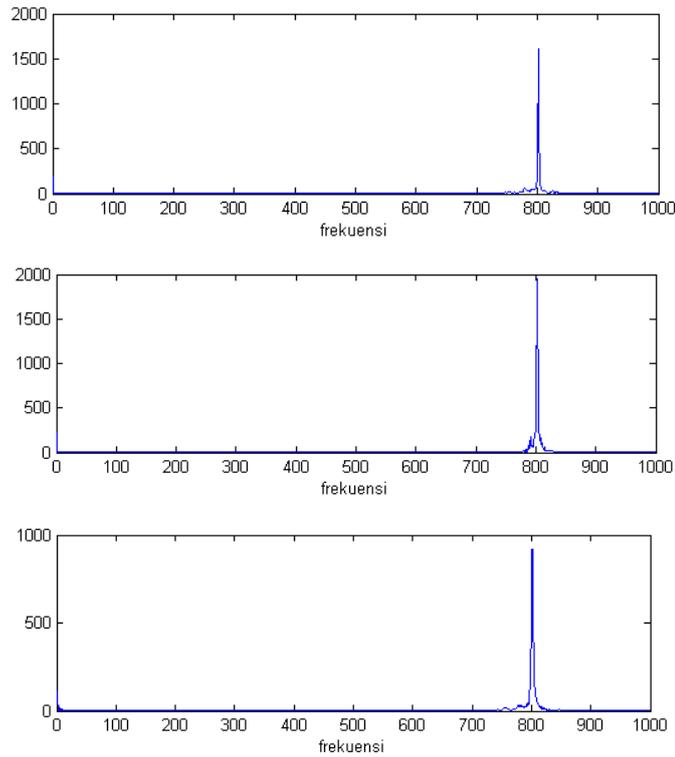
Pada WSOLA pengujian dilakukan dengan mengganti nilai parameter yang terdapat pada metode WSOLA. Ada 5 parameter pada metode WSOLA yaitu skala faktor, window, deltamax(nilai toleransi), panjang frame dan panjang pergeseran. Hasil pengujian yang dilihat adalah nilai frekuensi rekaman suara setelah pergeseran. Gambar 4.1 dan 4.2 menunjukkan bentuk sinyal hasil dari proses metode WSOLA.



Gambar 4.1 Sinyal asli (merah) dan sinyal hasil WSOLA yang diberi skala faktor 0.5(biru)



Gambar 4.2 Sinyal asli berwarna merah dan sinyal hasil WSOLA yang diberi skala faktor 2 berwarna biru



Gambar 4.3 frekuensi yang dihasilkan dari saron5 asli, saron5 skala faktor 0.5 dan saron5 skala faktor 2.

Dari Gambar 4.3 dapat dilihat frekuensi yang dihasilkan dari saron5 yang diberi metode WSOLA dengan skala faktor 0.5 dan 2 menghasilkan frekuensi yang sama dengan frekuensi rekaman saron5 yang asli yaitu frekuensi maksimumnya adalah 802 Hz. Ini membuktikan bahwa dengan mengubah skala waktu yaitu waktu yang dipercepat atau diperlambat dapat menghasilkan frekuensi yang tidak berubah dengan menggunakan metode WSOLA seperti pada Tabel 4.1.

Tabel 4.1 Pengujian Skala waktu WSOLA berdasarkan hasil frekuensi

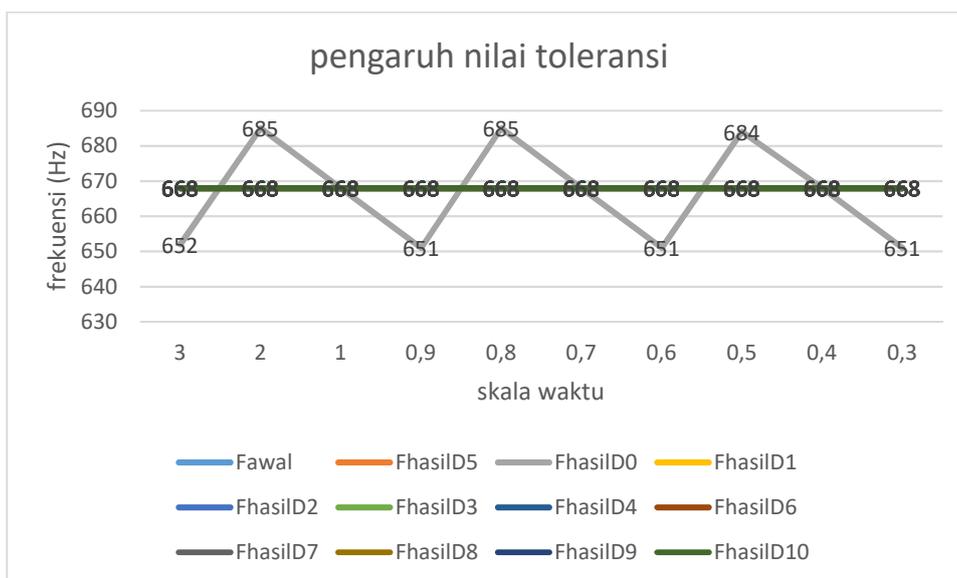
Skala faktor	Frekuensi awal (Hz)	Frekuensi Hasil (Hz)	Error frekuensi (%)
3	802	802	0
2	802	802	0
1	802	802	0

0.9	802	802	0
0.8	802	802	0
0.7	802	802	0
0.6	802	802	0
0.5	802	802	0
0.4	802	802	0
0.3	802	802	0

Tabel 4.1 adalah hasil dari rekaman suara saron5sl.wav yang dipotong sepanjang 40ms per frame, jarak tumpang tindih 20ms antar frame, dan nilai toleransi 5ms. Pada percobaan tersebut dilakukan perubahan skala waktu dari range 0.3ms-3ms menghasilkan nilai frekuensi yang sama dengan frekuensi awal.

4.3 Pengaruh nilai toleransi

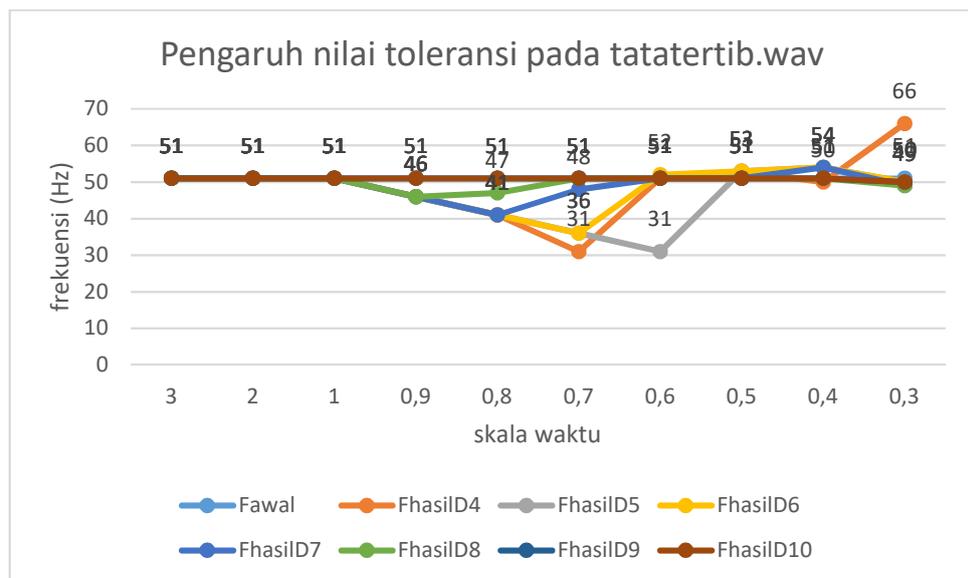
Peneliti melakukan percobaan dengan tidak memberikan nilai toleransi pada proses metode WSOLA pada rekaman suara. Hasil yang didapatkan adalah hasil frekuensi rekaman suara menjadi tidak sama dengan frekuensi asli seperti grafik yang terlihat pada Gambar 4.4.



Gambar 4.4 pengaruh nilai toleransi WSOLA pada saron1sl.wav

Pada Gambar 4.4 nilai toleransi yang dicoba antara 0-10 ms. FhasilD0,FhasilD1,...Fhasil10 adalah keterangan dari frekuensi hasil dengan nilai toleransi 0, frekuensi hasil dengan nilai toleransi 1 dan seterusnya sampai dengan nilai toleransi 10 ms. Pada grafik, FhasilD0 memiliki grafik yang naik dan turun yang artinya tidak memberikan nilai toleransi mempengaruhi frekuensi hasil dari rekaman suara.

Ini artinya parameter nilai toleransi sangat penting pada hasil metode WSOLA. Pada rekaman suara telewicara tatatertib.wav, semakin besar nilai toleransi yang diberikan akan menghasilkan frekuensi yang sama dengan rekaman suara asli. Grafik pengaruh nilai toleransi pada tatatertib.wav dapat dilihat pada Gambar 4.5

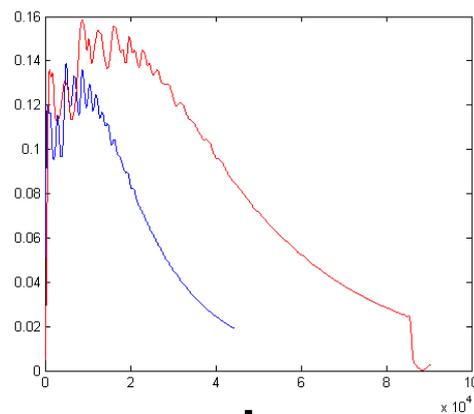


Gambar 4.5 Grafik pengaruh nilai toleransi pada tatatertib.wav

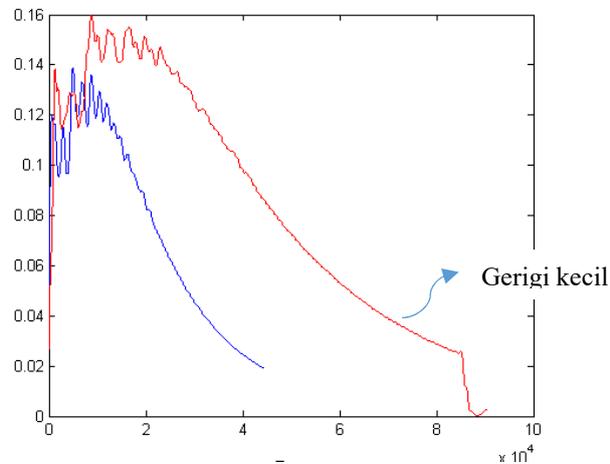
Pada Gambar 4.5 menunjukkan grafik hasil frekuensi dengan nilai toleransi 4ms dan 5 ms turun hingga frekuensi 31Hz dan naik hingga 66 Hz. Ini mengalami penurunan dan kenaikan yang sangat signifikan. Namun semakin tinggi nilai toleransi yang diberikan, grafik makin menunjukan hasil frekuensi maksimal yang semakin sama seperti sinyal masukan.

4.4 Pengaruh Window

Pengujian dilakukan pada rekaman suara gamelan saron satu slendro satu ketukan nada. Pengujian dilakukan pada keadaan jarak antar frame diberikan pergeseran sebesar 25%,50%,75%, dan 100% dari panjang frame dan diberikan jendela hamming, rectangular dan triangular pada setiap pengujiannya. Sinyal suara dipotong-potong menjadi beberapa frame, dimana setiap frame dipotong dengan panjang 40ms. Jarak overlap antar frame 10ms (25%),20ms(50%),30ms(75%), dan 40ms(100%). Pada pengujian antar jarak frame pergeseran sebesar 25%,50%,75% dengan menggunakan window hamming dan triangular menghasilkan bentuk sinyal dari transform hilbert yang lebih halus dari pada menggunakan window rectangular yang pada ujungnya sedikit berigi, lihat Gambar 4.6



(a) Tranform hilbert hamming and triangular

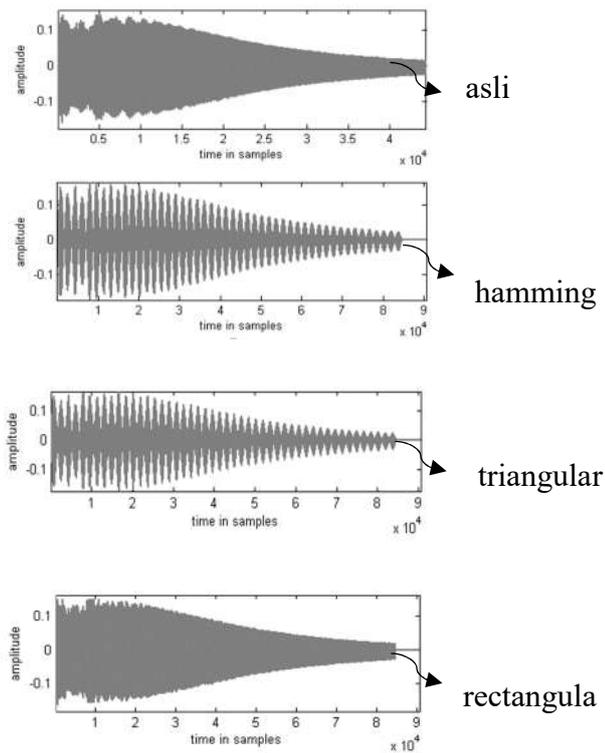


(b)transform hilbert rectangular

Gambar 4.6 Garis merah merupakan transform hilbert sinyal hasil pergeseran, garis biru adalah sinyal asli.

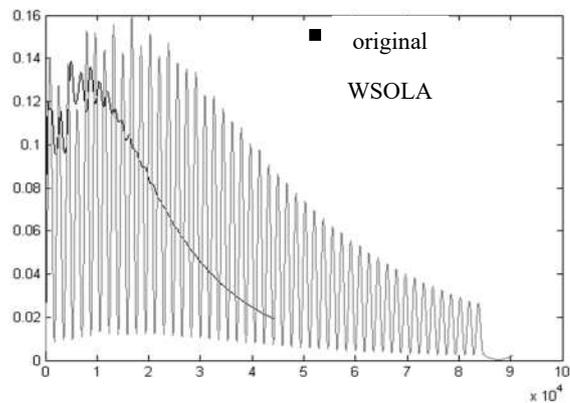
Walau hanya sedikit perbedaan yang dapat dilihat dari transform hilbert, namun pada saat kita mendengarkan suaranya menunjukan ada sedikit getaran dan ada kalanya suara tidak terdengar bergetar sama sekali.

Pada pengujian antar jarak frame pergeseran sebesar 100%, misalkan panjang frame 40ms dan jarak pergeseranya antar frame 40 ms dengan menggunakan window hamming dan triangular. Penggunaan window hamming dan triangular menghasilkan bentuk sinyal yang bergerigi dan transform hilbert yang menebal dan bergerigi. Sedangkan dengan menggunakan window rectangular, bentuk sinyal yang dihasilkan mirip dengan asli dan tidak bergerigi seperti yang ditunjukkan pada Gambar 4.7. ini artinya penggunaan window rectangular sangat tepat ketika overlap antar frame pada metode WSOLA di berikan jarak 100 % dari panjang frame. Side lobe yang dimiliki window rectangular lebih lebar dari pada hamming dan triangular, hal itu menyebabkan window rectangular masih mampu memberi sambungan yang baik pada overlap antar frame saat frame-frame tersebut di rekonstruksi kembali.

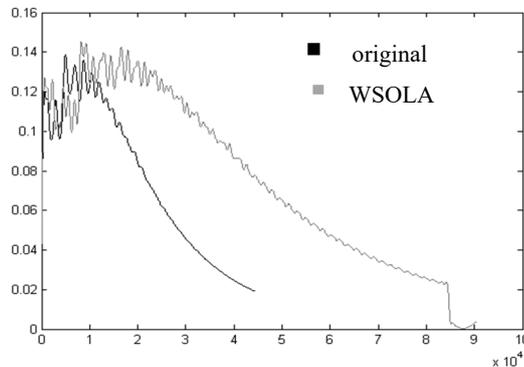


Gambar 4.7 bentuk sinyal

Pola sinyal yang dihasilkan pada penggunaan window hamming dan triangular adalah bergerigi besar dan tidak halus, jauh berbeda dari bentuk pola sinyal asli yang ditunjukkan dengan transform Hilbert. Sedangkan menggunakan window rectangular menghasilkan pola sinyal yang halus, tidak bergerigi dan tidak menebal yang di tujukan pada Gambar 4.8(a) dan 4.8(b).



Gambar 4.8(a) pola sinyal windowhamming dan triangular menggunakan transform hilbert dengan jarak overlap frame 100%



Gambar 4.8(b) pola sinyal window rectangular menggunakan transform hilbert dengan jarak overlap frame 100%

Pada penelitian ini, pengujian kualitas suara untuk mengevaluasi kinerja metode yang diusulkan menggunakan Perceptual Evaluation of Speech quality (PESQ, ITU-T P.862-2) dan Mean Opinion Score (MOS). Kualitas suara yang direkonstruksi sangat sensitif terhadap rasio data yang hilang. Dalam percobaan ini, empat jarak overlap digunakan dengan panjang masing-masing frame 40ms untuk mengevaluasi sinyal hasil metode WSOLA menggunakan window hamming, rectangular dan triangular.

Menurut hasil yang ditunjukkan pada Tabel 4.2 dan Tabel 4.3, kita dapat menemukan bahwa WSOLA menggunakan hamming dan triangular dapat mencapai nilai PESQ yang lebih tinggi dibandingkan dengan WSOLA menggunakan window rectangular pada saat overlap antar frame memiliki jarak 25%,50%,75% dan WSOLA menggunakan window rectangular memiliki nilai PESQ yang lebih tinggi dari window hamming dan triangular pada saat overlap antar frame 100%.

Tabel 4.2 PESQ WSOLA dengan skala faktor 2.

	overlap 10ms	overlap 20ms	overlap 30ms	overlap 40ms
Hamming	0.586	0.771	0.794	0.819
Triangular	0.579	0.760	0.894	0.730
Rectangular	0.346	0.724	0.457	0.944

Tabel 4.3 PESQ WSOLA dengan skala faktor 0.5.

	overlap 10ms	overlap 20ms	overlap 30ms	overlap 40ms
Hamming	2.522	2.369	2.089	1.933
Triangular	2.566	2.303	2.294	1.649
Rectangular	2.271	2.243	1.2	2.333

Pengujian juga dilakukan pada suara manusia yaitu tatatertib.wav dan goodfriend.wav dari kedua suara tersebut dicari nilai kualitas suara menggunakan software PESQ. Nilai hasil uji pada suara tatertib.wav ditunjukkan pada Tabel 4.4 dan Tabel 4.5

Tabel4.4 PESQ pada suara tatatertib1.wav menggunakan WSOLA dengan skala faktor 2.

	overlap 10ms	overlap 20ms	overlap 30ms	overlap 40ms
Hamming	1.940	1.745	1.700	0.565
Triangular	1.807	1.746	1.709	0.649
rectangular	1.945	2.589	1.673	1.698

Tabel4.5 PESQ pada suara tatatertib1.wav menggunakan WSOLA dengan skala faktor 0.5.

	overlap 10ms	overlap 20ms	overlap 30ms	overlap 40ms
Hamming	1.184	1.273	1.185	1.208
Triangular	1.188	1.292	1.194	0.876
rectangular	1.098	1.108	0.979	1.237

Tabel4.6 PESQ pada suara good friend.wav menggunakan WSOLA dengan skala faktor 2.

	overlap 10ms	overlap 20ms	overlap 30ms	overlap 40ms
Hamming	1.664	0.865	1.251	0.615
Triangular	1.775	1.085	1.237	0.579
rectangular	1.445	1.085	1.278	1.149

Tabel 4.7 PESQ pada suara good friend.wav menggunakan WSOLA dengan skala faktor 0.5.

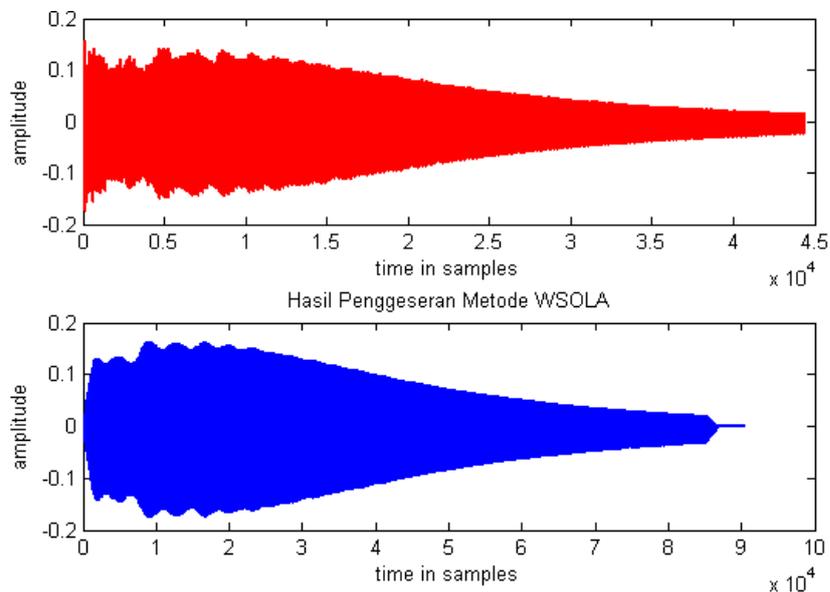
	overlap 10ms	overlap 20ms	overlap 30ms	overlap 40ms
Hamming	1.815	1.507	1.579	1.206
triangular	1.674	1.507	1.544	1.434
rectangular	1.644	1.468	1.626	1.443

Dari Tabel 4.4 dan 4.5 menunjukkan bahwa pada saat suara tatatertib1.wav yang diberi skala faktor 2 atau dipercepat dua kali menghasilkan nilai PESQ lebih tinggi saat metode WSOLA menggunakan window rectangular dari pada window hamming dan triangular kecuali pada 30 ms nilai window rectangular paling rendah. Sedangkan saat diperlambat dua kali atau diberi skala faktor 0.5 menghasilkan nilai PESQ tinggi pada window hamming dan triangular saat jarak overlap 10ms,20ms,dan30 ms.Pada jarak 40 ms window rectangular menghasilkan nilai PESQ yang lebih tinggi dari window hamming dan rectangular. Dari Tabel 4.6 dan 4.7 menunjukkan bahwa pada saat suara goodfriend.wav yang diberi skala faktor 2 atau dipercepat dua kali menghasilkan nilai PESQ lebih tinggi saat metode WSOLA menggunakan window rectangular dari pada window hamming dan triangular kecuali pada 10 ms nilai window rectangular paling rendah. Sedangkan saat diperlambat dua kali atau diberi skala faktor 0.5 menghasilkan nilai PESQ tinggi pada window hamming dan triangular saat jarak overlap 10ms dan 20ms. Pada jarak 30 dan 40 ms window rectangular menghasilkan nilai PESQ yang lebih tinggi dari window hamming dan rectangular.

4.5 Pengaruh Jarak antar Frame

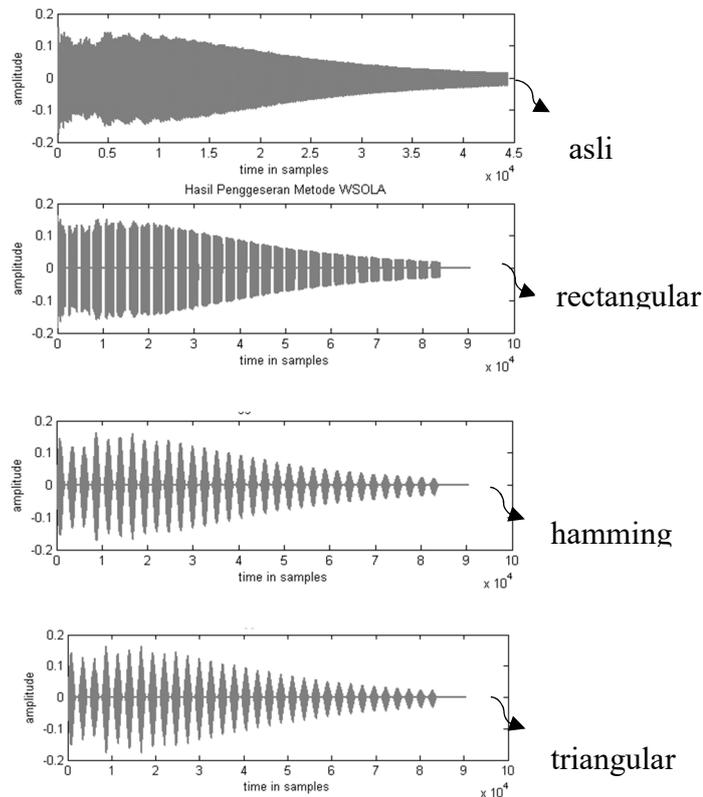
Pada Subbab 4.4 telah dilakukan percobaan menggunakan jarak antar frame antara 25%,50%,75% dan 100%. Semuanya jarak tersebut dapat menghasilkan suara dengan baik pada metode WSOLA. Walaupun setiap jarak tersebut membutuhkan

window yang berbeda untuk mendukung kinerja dari metode WSOLA untuk menghasilkan suara yang bagus. Namun apa yang terjadi dengan penggunaan jarak 0% dan lebih dari 100% pada metode WSOLA. Pada jarak antar frame di beri jarak 0% dalam artian tidak diberi jarak sama sekali, maka WSOLA akan error. Dari hasil percobaan, jarak terendah yang dapat dilakukan oleh WSOLA pada antar frame yaitu 1,25 ms dengan memberi nilai toleransi pada ujung frame 1-3ms. Sinyal hasil pergeserannya mirip dengan aslinya. Hasil bentuk sinyalnya dapat dilihat pada Gambar 4.9



Gambar 4.9 Bentuk sinyal hasil dari WSOLA yang diberi jarak terendah antar frame 1.25 ms.

Pada percobaan menggunakan jarak antar frame lebih dari 100% menghasilkan bentuk sinyal yang terputus-putus dapat dilihat pada Gambar 4.10 . percobaan dilakukan dengan memberi jarak 125 % antar framenya dan memberi window hamming,triangular dan rectangular, namun hasilnya tetap terputus, tidak mampu menyambung frame-frame tersebut karena jaraknya sudah tidak dapat dijangkau lagi oleh window-window tersebut. Hasil suaranya pun sudah pasti tidak bagus karena terputus-putus.



Gambar 4.10 Bentuk sinyal dari penggunaan jarak antar frame 125% dengan diberi window rectangular, hamming dan triangular.

4.6 Perbandingan Metode WSOLA dan PSOLA

Metode modifikasi skala waktu lain yang menjadi pembanding pada penelitian ini adalah pitch synchronous overlap and add (PSOLA). Rekaman suara yang digunakan sama dengan rekaman suara yang diuji pada metode WSOLA. Pada PSOLA, Pengujian nilai skala faktor di uji untuk melihat hasil frekuensi pergeseran. Hasil nilai frekuensi pergeseran akan dibandingkan dengan nilai frekuensi sinyal masukan. Range nilai skala faktor yang digunakan pada pengujian ini antara 3 hingga 0.3 untuk menggeser kerapatan waktu dipercepat dan diperlambat. Hasil pergeseran kerapatan waktu dengan menggunakan metode PSOLA berdasarkan hasil frekuensi sinyal suara dapat dilihat pada Tabel 4.8 dengan sinyal suara gamelan saron 5 slendro.

Tabel 4.8 Pengujian pergeseran waktu PSOLA berdasarkan hasil frekuensi

No.	Skala faktor	Frekuensi awal (Hz)	Frekuensi hasil (Hz)	Error frekuensi (%)	Square Error (Fout – Fin) ²
1	3	802	805	0.374	9
2	2	802	803	0.125	1
3	1	802	802	0	0
4	0.9	802	804	0.249	4
5	0.8	802	802	0	0
6	0.7	802	802	0	0
7	0.6	802	802	0	0
8	0.5	802	802	0	0
9	0.4	802	802	0	0
10	0.3	802	802	0	0

Mean Square Error merupakan rata-rata nilai square error

$$\text{Mean Square Error (MSE)} = \sum_{i=0}^n \frac{(F_{out} - F_{in})^2}{n} = 1.4$$

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\text{MSE}} = 1.183$$

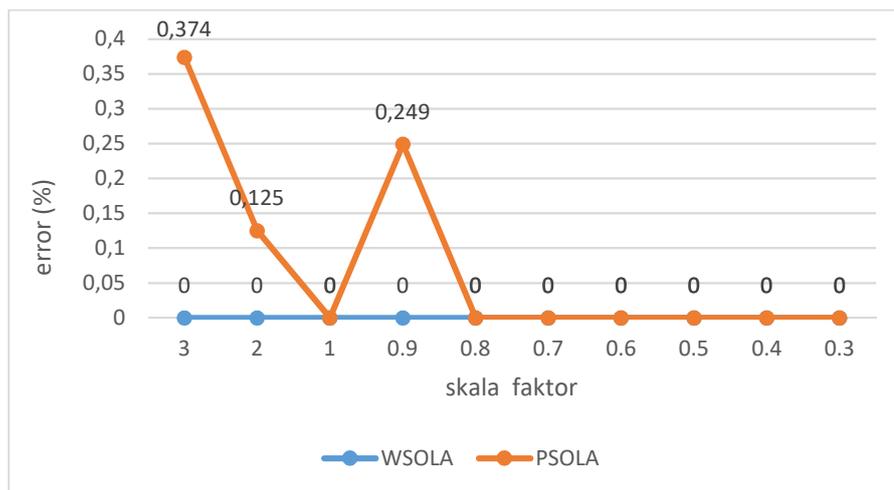
Hasil pengujian metode PSOLA seperti pada Tabel 4.6 menghasilkan frekuensi yang tidak sama dengan frekuensi masukan ketika diberi nilai skala faktor pergeserannya 3, 2 dan 0.9. Frekuensi maksimum sinyal saron 5 slendro menjadi sedikit naik dari sinyal aslinya walaupun tidak signifikan. Pada pengujian ini PSOLA memperoleh nilai MSE 1.4 dan RMSE 1.183.

Perbandingan pergeseran waktu pada metode WSOLA dan PSOLA berdasarkan frekuensi setelah dilakukan pengujian dapat ditunjukkan pada Tabel 4.9

Tabel 4.9 perbandingan metode WSOLA dan PSOLA berdasarkan frekuensi

No.	Skala faktor	Frekuensi awal (Hz)	Frekuensi hasil (Hz)	Frekuensi hasil (Hz)	Error frekuensi (%)	Error frekuensi (%)
1	3	802	805	805	0	0.374
2	2	802	803	803	0	0.125
3	1	802	802	802	0	0
4	0.9	802	804	804	0	0.249
5	0.8	802	802	802	0	0
6	0.7	802	802	802	0	0
7	0.6	802	802	802	0	0
8	0.5	802	802	802	0	0
9	0.4	802	802	802	0	0
10	0.3	802	802	802	0	0

Dari perbandingan Tabel 4.9, dapat dibuat grafik perbandingan antara metode WSOLA dan PSOLA. Grafik ditunjukkan oleh Gambar 4.11 dimana PSOLA memiliki nilai error 0.374 % sedangkan WSOLA 0%.



Gambar 4.11 Grafik perbedaan frekuensi WSOLA dan PSOLA.

Metode WSOLA dan PSOLA diuji juga dengan keadaan skala waktunya di percepat dan diperlambat 5 kali,10 kali dan 20 kali. Ini untuk melihat ketahanan dari kedua metode tersebut dengan digeser waktunya secara ekstrim masih dapat menjaga frekuensi maksimal dari suara atau tidak. Hasil frekuensi kedua metode tersebut ditunjukkan pada Tabel 4.10

Tabel 4.10 perbandingan metode WSOLA dan PSOLA berdasarkan frekuensi pada keadaan skala faktor ekstrim

No.	Skala factor	Frekuensi awal	Frekuensi WSOLA	Frekuensi PSOLA	Square Error WSOLA (Fout – Fin) ²	Square Error PSOLA (Fout – Fin) ²
1.	0.05	802	802	808	0	36
2.	0.1	802	802	783	0	361
3.	0.2	802	802	801	0	1
4.	5	802	802	806	0	16
5.	10	802	802	803	0	1
6.	20	802	802	814	0	144

$$\text{Mean Square Error (MSE) PSOLA} = \sum_{i=0}^n \frac{(F_{out} - F_{in})^2}{n} = 93,17$$

$$\text{Root Mean Square Error (RMSE)PSOLA} = \sqrt{\text{MSE}} = 9.65$$

$$\text{(MSE) WSOLA} = 0 \text{ , (RMSE)WSOLA} = \sqrt{\text{MSE}} = 0$$

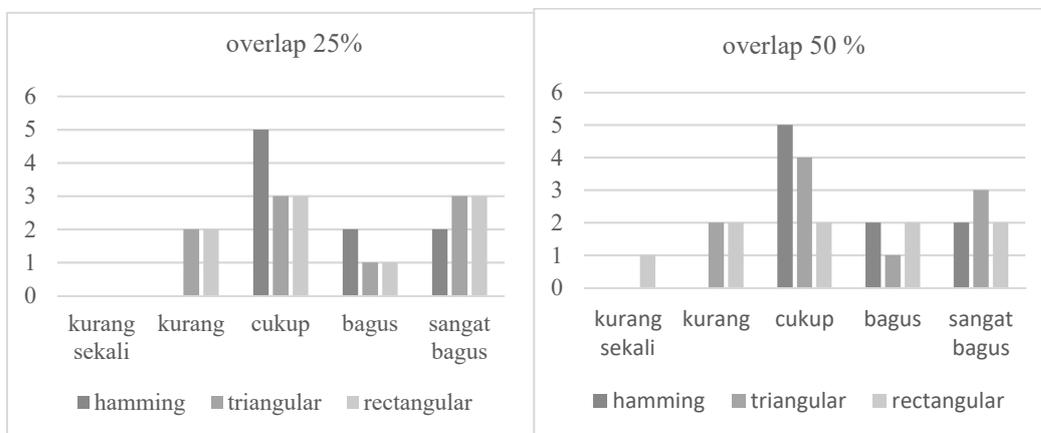
Tabel 4.11 Perbandingan nilai PESQ suara Saron5.wav

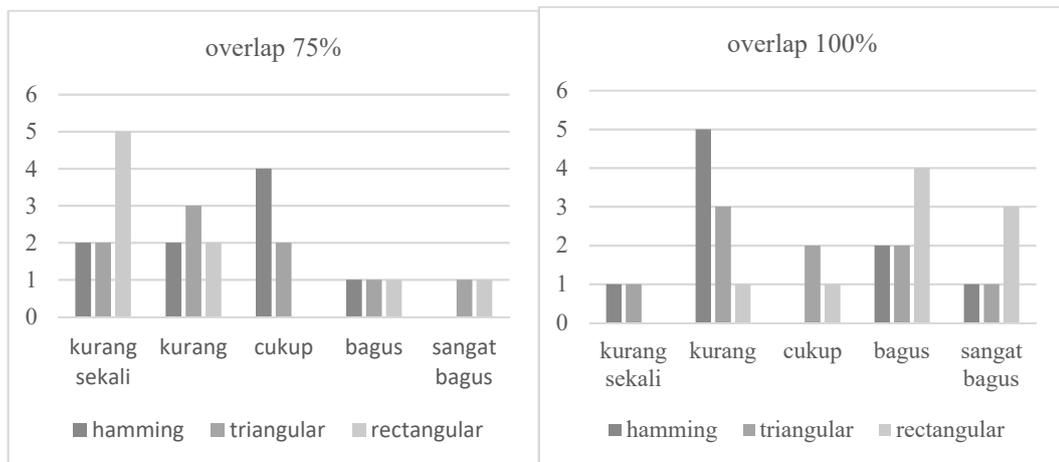
Skala faktor	PESQ WSOLA	PESQ PSOLA
3	1.055	0.198
2	0.586	0.632
1	2.338	2.166
0.9	2.927	2.671
0.8	2.631	2.608
0.7	2.587	2.495
0.6	2.546	2.321
0.5	2.521	2.166
0.4	2.379	2.020
0.3	2.089	1.923

Dari Tabel 4.11 menunjukkan nilai PESQ suara saron5 menggunakan metode WSOLA lebih tinggi dibandingkan dengan menggunakan metode PSOLA.

4.7 Mean Opinion Score(MOS)

Setelah membandingkan hasil pergeseran pada sinyal suara dengan menggunakan metode WSOLA dan PSOLA, maka dapat dilakukan kuisisioner dengan metoda MOS (*Mean Opinion Score*) untuk membandingkan hasil suara pergeseran kedua metode Time Stretching ke pendengar / responder. Pada kuisisioner, terdapat sepuluh responder yang berasal dari mahasiswa Institut Teknologi Sepuluh November. Penilaian responder terhadap kuisisioner terdiri dari lima kategori, yaitu kurang sekali, kurang, cukup, bagus, dan sangat bagus. Dan penilaian berdasarkan tiga macam, yaitu kekerasan, kejelasan dan kesamaan frekuensi suara hasil pergeseran dengan suara masukan. Berikut grafik dari hasil kuisisioner suara saron menggunakan metode WSOLA berdasarkan penggunaan window dan jarak overlap yang ditunjukkan pada Gambar 4.12





Gambar 4.12 Grafik hasil Kuisisioner metode WSOLA berdasarkan penggunaan window dan jarak overlap

Pada Gambar 4.12 grafik menunjukkan hasil penilaian subjektif dari para koresponden. Hasil kuisisioner menyatakan pada saat jarak overlap 25%, penggunaan window hamming lebih bagus dari triangular dan rectangular. Window triangular dan rectangular berada pada posisi yang seimbang. Jarak overlap 50%, koresponden menyatakan penggunaan window haming dan triangular lebih bagus, namun penggunaan window rectangular juga masih bagus. Pada jarak overlap 75 % penggunaan window rectangular sangat kurang baik dan penggunaan window hamming dan triangular lebih bagus. Terakhir pada jarak overlap 100 %, koresponden menyatakan penggunaan window rectangular lebih bagus dari pada window hamming dan triangular.

Halaman ini sengaja dikosongkan

BAB V

PENUTUP

5.1 KESIMPULAN

Berdasarkan hasil pengujian pada bab IV, maka dapat ditarik beberapa kesimpulan, yaitu

1. Dengan menggunakan metode WSOLA, dapat dilakukan modifikasi skala waktu sinyal suara saron dan suara manusia dengan tetap mempertahankan frekuensi aslinya dengan error rata-rata suara saron 0.847% dan suara manusia 5,094%.
2. Nilai toleransi yang merupakan salah satu nilai parameter WSOLA jika tidak diberi maka akan menghasilkan frekuensi yang tidak sama dengan frekuensi sinyal masukkan.
3. Pada percobaan suara manusia semakin tinggi nilai toleransi yang diberikan akan menghasilkan frekuensi yang semakin sama dengan frekuensi masukan.
4. Overlap antar frame dengan jarak 25%,50% dan 75% pada instrument sarondan suara manusia menggunakan metode WSOLA bagus menggunakan window hamming dan triangular sedangkan pada jarak 100% bagus menggunakan window rectangular. Ini ditunjukkan dari hasil bentuk sinyal, PESQ dan kuisioner.
5. Jarak antar frame yang lebih besar dari panjang frame akan menghasilkan diskontinuitas, bentuk sinyal terputus- putus dan suara yang buruk.

5.2 SARAN

1. Pada penelitian selanjutnya agar tidak hanya mengambil nilai frekuensi maksimum.
2. Pada penelitian selanjutnya agar menggunakan data yang lebih variasi misalnya data detak jantung.

Halaman ini sengaja dikosongkan

DAFTAR PUSTAKA

Alexis Moinet, Thierry Dutoit, and Thierry Dutoit Alexis Moinet, "Audio time-scaling for slow motion sports videos," in Proceedings of the 16th International Conference on Digital Audio Effects (DAFx), Maynooth, Ireland, September 2013

A. R. Virdaga, Pergeseran Kerapatan Waktu Sinyal Suara Menggunakan Metode PSOLA. Institut Teknologi Sepuluh Nopember. Publication on 2014.

Borchers, J. O., Samminger, W., and Mühlhäuser, M., Conducting a realistic electronic orchestra, UIST 2001 14th Annual Symposium on User Interface Software and Technology, Orlando, Florida, USA 2001.

Chellam,janani," Prosodic modification of speech to incorporate happy and sad emotions", Energy Information and Communication(ICCPEIC), 2015.

Carnegie Mellon, "Notes on FIR Filter Design Using Window Function", Signals and System (18396).

Dave Cliff, "Hang the DJ: Automatic sequencing and seamless mixing of dance-music tracks," Tech. Rep., HP Laboratories Bristol, 2000.

D. U. Ebem, "Validating Perceptual Objective Listening Quality Assessment Methods on the tonal Language Igbo", Delft University of Technology, june 2009

E Kail , S Khoo' r,dkk. *Expert System for Phonocardiographic Monitoring of Heart Failure Patients Based on Wavelet Analysis*. Computers in Cardiology IEEE. Publish on 2005.

Fera P.Ayu, Deteksi Dan Analisis Gelombang Alpha Pada Sinyal Eeg Terhadap Rangsang Suara Musik Menggunakan Transformasi Wavelet. Institut Teknologi Sepuluh Nopember. Publication on 2018.

Hiromi Ishizaki, Keiichiro Hoashi, and Yasuhiro Takishima, "*Full-automatic DJ mixing system with optimal tempo adjustment based on measurement function of user discomfort*," in Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Kobe, Japan, 2009,pp. 135–140.

J.Driedger and Muller Meinard, "*TSM Tolboox: Matlab Implementations of time-scale modification algorithms*," Int. Conference on Digital Audio Effects (DAFx-14), Erlangen, Germany, September 1-5, 2014.

Jordi, B.S., *Audio Time-Scale Modification in the Context of Professional Audio Post-production*, Universitat Pompeu Fabra, Barcelona. Publish on 2002

K. Terry, "Phase Vocoder Implementation with FLWT and TD-PSOLA", Stanford University. 2015.

Malden Electronics Ltd and Psytechnics Ltd, "Speech Quality Assessment," 2004.

Pallone, G., Boussard, P., Daudet, L., Guillemain, P., and Kronland-Martinet, R., A wavelet based method for audio-video synchronization in broadcasting applications, *Proceedings of the Digital Audio Effects Workshop (DAFx99)*, Trondheim 1999.

Prasetyo Rudi, "Perancangan sistem perbaikan nada suara manusia dengan menggunakan metode phase vocoder terhadap nada referensi musik", Jurusan teknik elektro, Universitas Diponegoro, Semarang: 2014

Pallone, G., Boussard, P., Daudet, L., Guillemain, P., and Kronland-Martinet, R., A wavelet-based method for audio-video synchronization in broadcasting applications, *Proceedings of the Digital Audio Effects Workshop (DAFx99)*, Trondheim 1999

Sticht, T. G., Comprehension of repeated time-compressed recordings, *The Journal of Experimental Education*, vol. 37, no. 4, 1969

S. Grofit, Y. Lavne. *Time-Scale Modification of Audio Signals Using Enhanced WSOLA With Management of Transients*. *IEEE Transactions On Audio, Speech, And Language Processing*. Publish on 2008.

Werner Verhelst and Marc Roelands, "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Minneapolis, USA, 1993.

X. Huang, A. Agero, and H. Hon, *Spoken Language Processing: A Guide to theory, Algorithm, and System Development*, Prentice Hall, New Jersey, 2001

Zölzer, U. *DAFX: Digital Audio Effects*; John Wiley & Sons, Inc.: New York, NY, USA, 2002

LAMPIRAN

Keterangan:

Alpha = skala faktor

Fawal= Frekuensi awal

FD5= frekuensi hasil pada deltamax bernilai 5

METODE WSOLA

SARON2SL.wav

alpha	Fawal	FD5	FD0	FD1	FD2	FD3	FD4	FD6	FD7	FD8	FD9	FD10
3	668	668	652	668	668	668	668	668	668	668	668	668
2	668	668	685	668	668	668	668	668	668	668	668	668
1	668	668	668	668	668	668	668	668	668	668	668	668
0,9	668	668	651	668	668	668	668	668	668	668	668	668
0,8	668	668	685	668	668	668	668	668	668	668	668	668
0,7	668	668	668	668	668	668	668	668	668	668	668	668
0,6	668	668	651	668	668	668	668	668	668	668	668	668
0,5	668	668	684	668	668	668	668	668	668	668	668	668
0,4	668	668	668	668	668	668	668	668	668	668	668	668
0,3	668	668	651	668	668	668	668	668	668	668	668	668

SARON5SL.wav

alpha	Fawal	FD5	FD0	FD1	FD2	FD3	FD4	FD6	FD7	FD8	FD9	FD10
3	581	581	1	581	581	581	581	581	581	581	581	581
2	581	581	561	581	581	581	581	581	581	581	581	581
1	581	581	581	581	581	581	581	581	581	581	581	581
0,9	581	581	1	581	581	581	581	581	581	581	581	581
0,8	581	581	565	581	581	581	581	581	581	581	581	581
0,7	581	581	607	580	581	581	581	581	581	581	581	581
0,6	581	581	549	581	581	581	581	581	581	581	581	581
0,5	581	581	591	581	581	581	581	581	581	581	581	581
0,4	581	581	1	581	581	581	581	581	581	581	581	581
0,3	581	581	575	581	581	581	581	581	581	581	581	581

SARON5 dua nada

alpha	Fawal	FD0	FD1	FD2	FD3	FD4	FD6	FD7	FD8	FD9	FD10	FD5
3	802	805	802	802	802	802	802	802	802	802	802	802
2	802	804	803	802	802	802	802	802	802	802	802	802
1	802	802	802	802	802	802	802	802	802	802	802	802
0,9	802	822	802	802	802	802	802	802	802	802	802	802
0,8	802	792	802	802	802	802	802	802	802	802	802	802
0,7	802	812	802	802	802	802	802	802	802	802	802	802
0,6	802	782	802	802	802	802	802	802	802	802	802	802
0,5	802	802	802	802	802	802	802	802	802	802	802	802
0,4	802	822	802	802	802	802	802	802	802	802	802	802
0,3	802	792	802	802	802	802	802	802	802	802	802	802

Nada saron1 dan saron2 di gabung

alpha	Fawal	FD5	FD0	FD1	FD2	FD3	FD4	FD6	FD7	FD8	FD9	FD10
3	668	668	653	668	668	668	668	668	668	668	668	668
2	668	668	685	667	667	668	668	668	668	668	668	668
1	668	668	668	668	668	668	668	668	668	668	668	668
0,9	668	668	651	667	667	668	668	668	668	668	668	668
0,8	668	668	685	667	668	668	668	668	668	668	668	668
0,7	668	668	668	668	668	668	668	668	668	668	668	668
0,6	668	668	651	669	669	668	668	668	668	668	668	668
0,5	668	668	684	668	669	669	668	668	668	668	668	668
0,4	668	668	668	668	668	668	668	668	668	668	668	668
0,3	668	668	651	668	668	668	668	668	668	668	668	668

Tatatertib1.wav

alpha	Fawal	FD5	FD4	FD6	FD7	FD8	FD9	FD10
3	51	51	51	51	51	51	51	51
2	51	51	51	51	51	51	51	51
1	51	51	51	51	51	51	51	51
0,9	51	46	46	46	46	46	51	51
0,8	51	41	41	41	41	47	51	51
0,7	51	36	31	36	48	51	51	51
0,6	51	31	51	52	51	51	51	51
0,5	51	53	53	53	51	51	51	51

0,4	51	54	50	54	54	51	51	51
0,3	51	50	66	50	49	49	50	50

WSOLA Good_friends.wav

alpha	Fawal	Fhasil
3	727	1049
2	727	1050
1	727	725
0,9	727	728
0,8	727	728
0,7	727	727
0,6	727	725
0,5	727	726
0,4	727	727
0,3	727	727

METODE PSOLA

SARON5SL.wav

alpha	Fawal	Fhasil
3	802	805
2	802	803
1	802	802
0,9	802	804
0,8	802	802
0,7	802	801
0,6	802	803
0,5	802	802
0,4	802	802
0,3	802	802

Tatatertib1.wav

alpha	Fawal	Fhasil
3	51	101
2	51	101
1	51	51

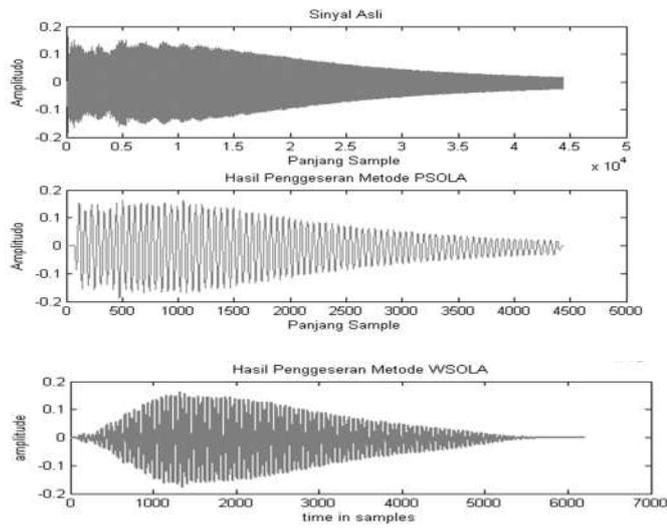
0,9	51	46
0,8	51	41
0,7	51	37
0,6	51	31
0,5	51	26
0,4	51	21
0,3	51	16

Good_friend.wav

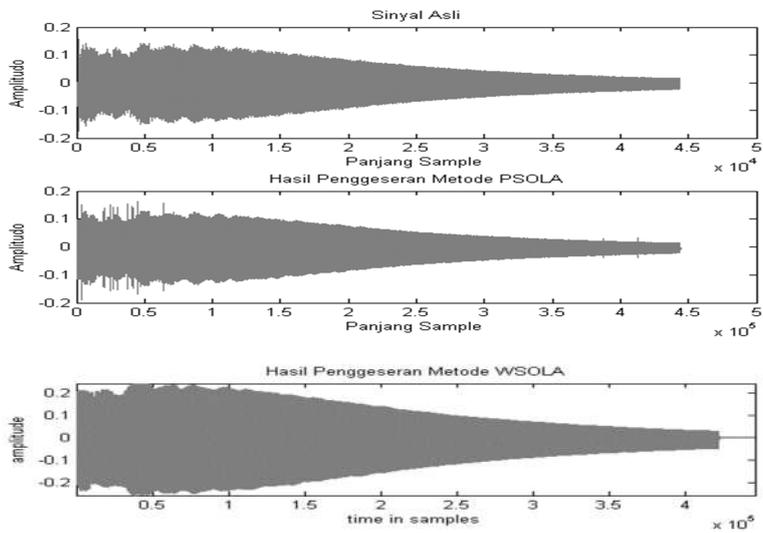
alpha	Fawal	Fhasil
3	727	1080
2	727	1059
1	727	730
0,9	727	1037
0,8	727	726
0,7	727	726
0,6	727	729
0,5	727	730
0,4	727	728
0,3	727	730

SARON5sl

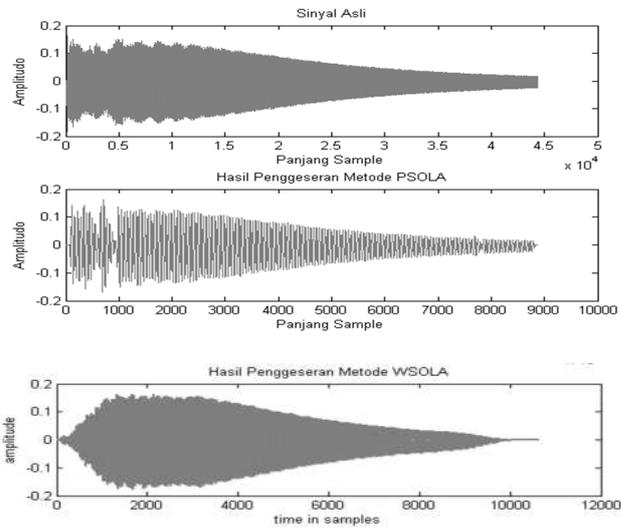
Dipercepat 10 FPSOLA=803 Hz , WSOLA=802Hz



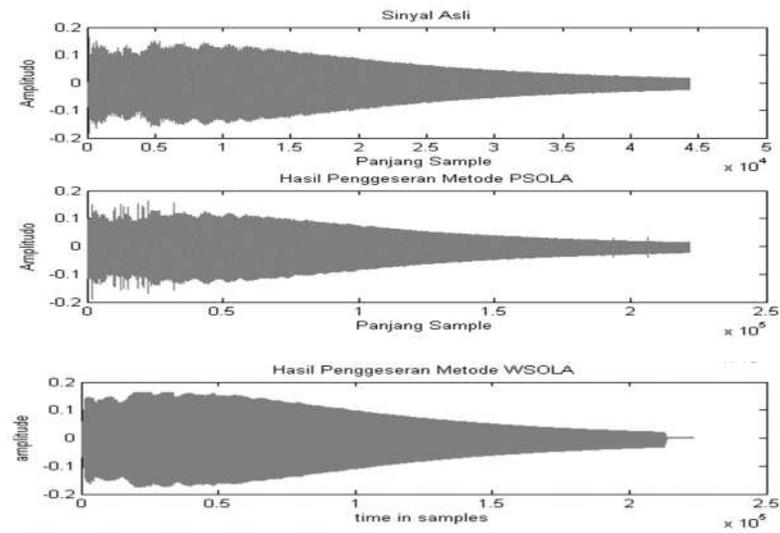
PSOLA diperlambat 10x f=783 WSOLA=802Hz



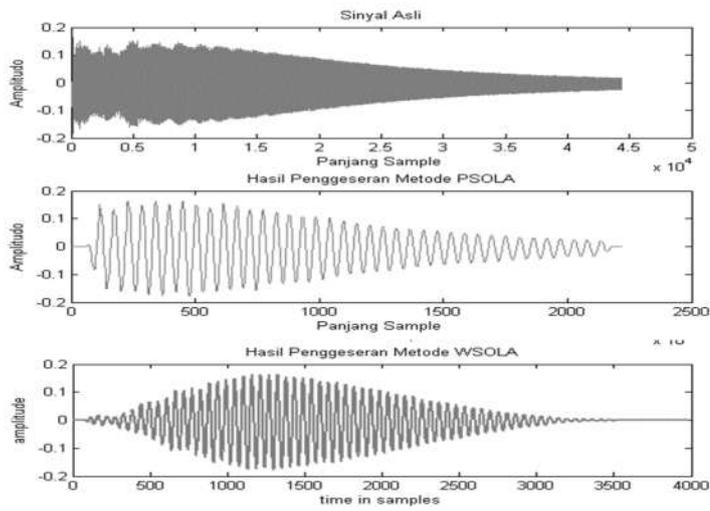
Dipercepat 5x $f=806$ WSOLA=802Hz



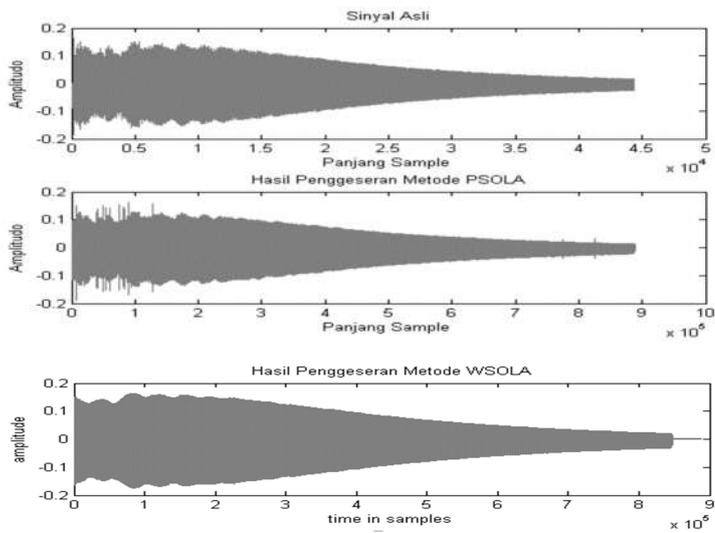
Diperlambat 5x FPSOLA=801 FWSOLA=802Hz



Dipercepat 20x fPSOLA=814



Diperlambat 20x FPSOLA=806, FWSOLA=802



Halaman ini sengaja dikosongkan

BIODATA PENULIS



I. DATA PRIBADI

Nama lengkap : Nanda Saputri
Tempat,Tanggal Lahir : Bayu, 20 November 1991\
Alamat : Jl. Ismail Ali, Kampun Jawa Baru,
Lhokseumawe, Aceh
Kebangsaan/Suku : Indonesia/ Aceh
Agama : Islam
Email : nandasaputri24@gmail.com

II. RIWAYAT PENDIDIKAN

SD : SD Negeri Bertingkat Lancang Garam
SMP : SMP Negeri 1 Lhokseumawe
SMA : SMA Negeri 1 Lhokseumawe
D4 : Politeknik Negeri Lhokseumawe/ Teknik Informatika

Halaman ini sengaja dikosongkan