**PROPOSAL TUGAS AKHIR**

**PEMISAHAN SUMBER SUARA TERCAMPUR BERDASARKAN *BINAURAL CUES: INTERAURAL TIME DIFFERENCE* DAN *INTERAURAL INTENSITY DIFFERENCE***

****

**MIFTA NUR FARID**

**NRP : 2409 100 012**

**DOSEN PEMBIMBING :**

**Dr. Dhany Arifianto ST., M.Eng.**

**PROGRAM STUDI S1 TEKNIK FISIKA**

**JURUSAN TEKNIK FISIKA**

**FAKULTAS TEKNOLOGI INDUSTRI**

**INSTITUT TEKNOLOGI SEPULUH NOPEMBER**

**SURABAYA**

**2013**

**LEMBAR PENGESAHAN**

**PROPOSAL TUGAS AKHIR**

**JURUSAN TEKNIK FISIKA FTI-ITS**

1. Judul :Pemisahan Sumber Suara Tercampur Berdasarkan *Binaural Cues: Interaular Time Difference* dan *Interaural Intensity Difference*
2. Bidang Studi : Rekayasa Akustik dan Fisika Bangunan
3. a. Nama : Mifta Nur Farid

b. NRP : 2409100012

c. Jenis Kelamin : Laki-laki

1. Jangka Waktu : 6 bulan
2. Pembimbing I : Dr. Dhany Arifianto ST., M.Eng.
3. Usulan Proposal ke : I
4. Status : Baru

Surabaya, November 2013

Pengusul,

Mifta Nur Farid

NRP 2409100012

Menyetujui

|  |
| --- |
| Pembimbing I,  Dr. Dhany Arifianto, ST., M.Eng  NIP. 19731007 199802 1 001 |

Mengetahui

Kepala Laboratorium

Rekayasa Akustik dan Fisika Bangunan,

Dr. Dhany Arifianto, ST., M.Eng

NIP. 19731007 199802 1 001

1. **Judul**

Pemisahan Sumber Suara Tercampur Berdasarkan *Binaural Cues: Interaular Time Difference* dan *Interaural Intensity Difference*

**Mata kuliah pilihan yang diambil**

|  |  |
| --- | --- |
| * Teknik Sistem Audio * Analisa Getaran Mekanik * Akustik Kelautan |  |
|  |  |

1. **Pembimbing**

Pembimbing 1 : Dr. Dhany Arifianto, ST., M.Eng.

1. **Latar Belakang**

Dalam suatu percakapan dengan latar belakang suara lain, seseorang mampu memfokuskan pendengarannya pada seorang lawan bicaranya meskipun bunyi musik latar cukup keras dan disertai percakakapan beberapa orang lainnya. Fenomena ini dikenal dengan *the cocktail party effect.[1,2]* Istilah "*cocktail party processing*" diciptakan dalam sebuah studi awal terhadap the cocktail party effect, dalam studi ini menggambarkan bahwa *binaural hearing* memberikan kontribusi penting dalam analisa pendengaran yang memungkinkan kita untuk memisahkan dan melokalisir sumber suara.[2]

Dalam *binaural hearing,* jika posisi suatu sumber suara tidak berada dalam bidang simetri vertikal (bidang median) maka salah satu telinga akan dibayangi oleh kepala sedangkan telinga lain terbuka penuh terhadap sumber bunyi. Akibatnya terjadi perbedaan tingkat tekanan bunyi yang terdengar pada kedua telinga (*Interaural Level Difference*) serta perbedaan waktu tempuh sumber suara terhadap kedua telinga tersebut (*Interaural Time Difference)*. Dari kedua efek tersebut dapat dilakukan pemisahan sumber suara yang disebut *Binaural Model* atau *Binaural Processing.*

*Binaural model* atau *binaural processing* adalah teknik pemisahan suara yang mengeksplorasi informasi perbedaan waktu tempuh (interaural time difference) dan perbedaan level amplitudo (interaural time difference) bunyi, yang disebabkan oleh efek dari kepala, torso dan telinga bagian luar[1]. Deliang Wang[3] melakukan penelitian tentang pemisahan suara yang berdasarkan pada lokalisasi bunyi. Pemisahan suara dilakukan dengan memperkirakan *time-frequency binary mask* yang diperoleh dari analisa ITD dan ILD terhadap *Relative Strength* dari sinyal target.

Pada penelitian tersebut[3], belum dilakukan pemisahan suara tercampur dari sumber suara yang berdekatan yaitu pada azimuth 0o, 10o, 20o dan 30o serta sumber suara yang berada tepat pada posisi azimuth terjauh yaitu azimuth 0o dan 180o yang menyebabkan kebingungan arah (*ambiguity)*. Maka pada penelitian ini, akan dilakukan pemisahan suara dari input binaural dengan 2 sensor mikrofon dari dua sumber dengan azimuth yang berdekatan dan yang terjauh terhadap sensor menggunakan metode *Joint* ITD-ILD.

1. **Permasalahan**

Berdasarkan uraian pada latar belakang dirumuskan beberapa permasalahan yang harus terjawab:

* + 1. Bagaimana memisahkan sumber suara tercampur dari *input binaural* berdasarkan *binaural cue:* ITD dan ILD?
    2. Bagaimana kualitas hasil pemisahan sumber suara tercampur dari *input binaural* berdasarkan *binaural cue*: ITD dan ILD secara objektif (PESQ dan SNR) dan subjektif (*Intelligibility Index - % cirrect word* dan *Mean Opinion Score* ?

1. **Batasan Masalah**

Agar pembahasan tidak meluas dan menyimpang dari tujuan dan rumusan masalah, akan diberikan beberapa batasan permasalahan dari penelitian ini, yaitu sebagai berikut:

1. Sinyal suara yang digunakan adalah suara perempuan dan laki-laki berbahasa indonesia.
2. Data HRTF yang digunakan adalah *CIPIC HRTF DATABASE[6]* dengan azimuth 0o, 10o, 20o, 30o dan 180o.
3. Pencampuran sinyal suara dilakukan dengan melakukan konvolusi antara HRTF dengan sumber suara yang digunakan.
4. Pengolahan data dibatasi dengan menggunakan teknik ITD dan ILD menggunakan software matlab.
5. **Tujuan**

Tujuan dari dilakukannya penelitian ini adalah untuk memisahkan 2 suara tercampur dari *input binaural* dengan menggunakan teknik *binary masking* yang diperoleh dari analisa pola ITD dan ILD.

1. **Tinjauan Pustaka**
   * 1. Tris Atmaja, Bagus. 2012. “**Pemisahan Sumber-Sumber Suara Tercampur Dari Input Binaural*”.***Pada penelitian ini, pemisahan sumber suara tercampur dari *input binaural* menggunakan metode ICA, ICABM, FDCW, FastICA, dan FastICABM. Hasil yang didapatkan adalah metode FastICA dengan binary mask merupakan metode yang terbaik dengan nilai PESQ sebesar 2.74.
2. **Teori Penunjang**

Kata kunci: ***Head Relative Transfer Function, Interaural Time Difference, Interaural Level Difference, Ideal Binary Mask, Relative Strength.***

**1. *Head-Related Impulse Response***

Untuk menemukan tekanan bunyi yang dihasilkan oleh suatu sumber sembarang pada gendang telinga manusia, diperlukan respon impuls dari sumber tersebut sampai ke gendang telinga. Respon impuls ini disebut dengan *Head Related Impulse Response* (HRIR). Jadi, definisi dari HRIR adalah suatu respon impuls gendang telinga manusia yang berfungsi untuk menyaring bunyi yang datang padanya dari suatu posisi sehingga dihasilkan tekanan bunyi tertentu pada gendang telinga tadi[4].

Proses sintesis binaural dapat dirumuskan dalam suatu sistem Linear Time-Invariant (LTI). Jika sinyal masukan monaural disimbolkan dengan xm(n), sinyal-sinyal binaural yang dihasilkan pada kedua telinga manusia diperoleh dari konvolusi berikut:

(1)

(2)

dimana y adalah suatu vektor kolom dari sinyal-sinyal binaural dan h adalah suatu vektor kolom sepasang HRIR yang digunakan dalam sintesis binaural. Persamaan (2) menunjukkan bahwa dan adalah masing-masing HRIR telinga kiri dan HRIR telinga kanan serta dan adalah masing-masing tekanan bunyi yang sampai pada gendang telinga kiri dan kanan. Jadi dan merupakan hasil penyaringan sumber bunyi monoaural yang dilakukan oleh masing-masing HRIR telinga kiri dan kanan.

Pasangan transformasi Fourier dari HRIR dikenal sebagai Head-Related Transfer Function(HRTF). HRTF didefinisikan sebagai fungsi alih (*transfer function)* sistem penyaringan sumber bunyi oleh gendang telinga kiri dan gendang telinga kanan dalam kawasan frekuensi. Jika HRIR dilambangkan h(n) maka HRTF dilambangkan H(ejω).

HRTF adalah fungi transfer dari kedua telinga pendengar dalam reproduksi bunyi. Karena persepsi bunyi setiap orang berbeda dengan persepsi orang lain, maka HRTF dari masing-masing telinga seseorang bersifat sangat khas. Dengan kata lain, pasangan HRTF seseorang berbeda dengan pasangan HRTF orang lain, oleh karena itu dikatakan HRTF bersifat individual.

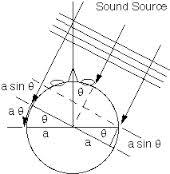
HRTF mencakup semua faktor fisik dari penentuan posisi sumber bunyi. Jika HRTF untuk telinga kiri dan HRTF untuk telinga kanan diperoleh, suatu sumber monaural dapat disintesis dengan menggunakan kedua HRTF tadi menjadi sinyal-sinyal binaural. HRTF merupakan fungsi yang rumit dari empat variabel: tiga koordinat ruang dan frekuensi. Dalam koordinat bola, jika jarak lebih dari satu meter, sumber bunyi dikatakan berada dalam medan jauh dan HRTF bertambah kecil dengan bertambahnya jarak. Kebanyakan pengukuran HRTF dilakukan di dalam medan jauh, yang pada intinya mengakibatkan HRTF menjadi suatu fungsi dari azimut dan frekuensi.

**2. *Azimuth Cues***

Lord Rayleigh[5] dalam teorinya, terdapat dua *cue* utama untuk azimuth yakni selisih waktu *interaural* (ITD) dan selisih *level interaural* (ILD).

***a. Interaural Time Difference***

ITD adalah perbedaan waktu tempuh suatu bunyi antara telinga kiri dan telinga kanan. Posisi-posisi sumber bunyi pada bidang median memiliki ITD mendekati nol, dimana untuk kepala yang simetris sempurna tidak ada ITD pada bidang tersebut. Semakin jauh posisi sumber bunyi dari bidang median maka semakin besar ITD yang diperoleh pada posisi tersebut. ITD maksimum terjadi pada posisi-posisi ekstrem persis di sebelah salah satu telingan dengan elevasi 0o. Dengan demikian ITD merupakan fungsi dari azimuth pada bidang-bidang dengan elevasi tetap. ITD dapat juga dihitung dari waktu tunda hasil korelasi silang maksimum antara sepasang HRIR.



**Gambar 1**. Ilustrasi untuk Menentukan ITD.

Lord Rayleigh memiliki suatu penjelasan yang sederhada untuk ITD. Bunyi merambat dengan kecepatan sebesar c, lalu gelombang bunyi dari suatu sumber mengenai kepala *spheric* dengan jari-jari a dari suatu arah yang ditentukan oleh sudut azimuth Ɵ. Maka bunyi akan datang pada telinga kanan sebelum telinga kiri, karena gelombang akan merambat jarak ekstra sebesar (aƟ + a sin Ɵ) untuk mencapai telinga kiri. Jika jarak ekstra itu dibagi dengan kecepatan bunyi c, formula sederhana untuk ITD berikut ini yang juga akurat diperoleh:

(3)

***b. Interaural Level Difference***

Lord Rayleigh juga mengamati bahwa gelombang-gelombang bunyi yang datang dibelokkan oleh kepala. Persamaan gelombang benar-benar diselesaikan untuk menunjukkan bagaimana suatu gelombang bidang dibelokkan oleh bola keras. Solusinya menunjukkan bahwa sebagai tambahan untuk selisih waktu, terdapat juga suatu selisih signifikan antara level-level sinyal pada kedua telinga yaitu ILD.

ILD didefinisikan sebagai perbedaan *level* atau *magnitude* dalam satuan dB di kawasan frekuensi antara sepasang HRTF magnitude yang berasal dari kedua telinga. Untuk suatu posisi sumber bunyi, diperoleh nilai ILD dari masing - masing komponen frekuensi. Pada frekuensi-frekuensi rendah,dimana panjang gelombang bunyinya relatif panjang terhadap diameter kepala, terdapat selisih yang sangat kecil dalam tekanan bunyi pada kedua telinga. Tetapi pada frekuensi-frekuensi tinggi, dimana panjang gelombangnya pendek, akan terdapat suatu selisih sebesar 20 dB atau lebih. Hal ini disebut efek bayangan kepala *(headshadow effect)*, dimana telinga jauh berada di dalam bayangan bunyi dari kepala.

Seperti halnya perubahan ITD terhadap azimut di bidang horisontal, ILD dari setiap frekuensi bertambah besar ketika bunyi mendekati kedua telinga. ILD maksimum pada bidang horisontal terjadi di sekitar azimut 90o dan 270o. Di sini dapat diamati bahwa ILD merupakan fungsi dari azimut di bidang horisontal. Karena efek difraksi dan bayangan kepala, pada komponen-komponen frekuensi lebih tinggi pada umumnya terjadi ILD yang lebih besar

Teori Duplex menambahkan bahwa ILD dan ITD saling melengkapi. Pada frekuensi-frekuensi rendah (di bawah sekitar 1,5 kHz), ada sedikit informasi ILD, tetapi ITD menggeser gelombang itu sebagian dari satu periode, yang dengan mudah dapat dideteksi. Padafrekuensi-frekuensi tinggi (di atas kira-kira 1,5 kHz), terdapat kebingungan dalam ITD karena ada pergeseran sebanyak beberapa periode, tetapi ILD menutupi kebingungan arah (*ambiguity)* ini. Teori Duplex dari Rayleigh mengatakan bahwa ILD dan ITD yang diambil bersama-sama memberikan informasi lokalisasi sepanjang jangkauan frekuensi yang terdengar[4].

**3. Ideal Binary Masked**

Masking terhadap suara tercampur dan menentukan secara biner apakah sinyal tersebut memiliki sinyal target berdasarkan *relative strength.*[3]

***A.* Binary Masked:**



(4)

**B. Relative Strength**

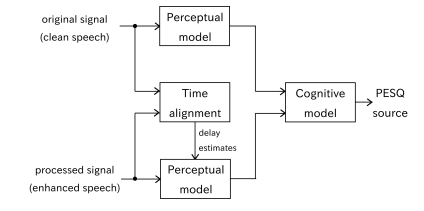


(5)

dimana s adalah sinyal dan n adalah noise.

**PESQ**

PESQ adalah standar ukuran objektif untuk mengetahui kualitas suara khususnya di bidang telekomunikasi. Metode ini diusulkan oleh ITU-T Reccomendation P.862. PESQ mengukur degradasi suara yang diterima terhadap suara aslinya, sehingga dibutuhkan referensi dan suara output sistem untuk dibandingkan satu sama lain.[1]



**Gambar 2**. Skema penghitungan PESQ[1]

**SNR**

SNR adalah perbandingan (ratio) antara kekuatan sinyal (*signal strength)* dengan kekuatan derau (*noise level).* Semakin besar nilai SNR maka kualitas sinyal semakin bagus. Secara matematis SNR dapat ditulis sebagai berikut.

SNR = 10 Log ()2 (6)

Dimana :

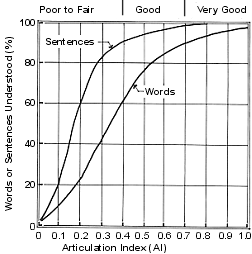
SNR = *Signal to noise ratio* (dB)

As = Amplitudo sinyal (m)

An = Amplitudo *noise* (m)

**Speech Intelligibility (% Correct Word)**

Dalam tes subjektif banyak cara mengukur kualitas suara, selain MOS ada pula %Correct Words. Penilaian %Correct Words didasari pada seberapa banyak (dalam presentase) kata yang bisa ditulis kembali setelah didengarkan. Dari sini dapat dilihat seberapa baik pesan yang diterima oleh pendengar, semakin sedikit kata yang salah maka semakin baik kualitas suara dari sisi pesan yang diterima. Oleh karena itu jenis kata yang diberikan harus diperhatikan betul semisal kata-kata yang diberikan adalah kata-kata umum. Cara melakukan percobaan ini adalah, pendengar diberikan stimuli berupa kalimat sederhana yang tidak boleh terlalu panjang. Mengikuti peraturan ITU-T tentang subjektif tes, kalimat yang diberikan berdurasi sekitar 2,5-5 detik[7]. Kemudian sebelum diperdengarkan kalimat uji, pendengar dilatih agar dapat membiasakan diri dengan suara dan teknik ini. Nantinya setiap selesai 1 kalimat pendengar diberikan waktu untuk menulis kembali kalimat yang ia dengar dengan media yang nyaman bagi mereka, bisa berupa Laptop (ketik) ataupun kertas (tulis). Pendengar hanya diperkenankan mendengar 1 kali setiap stimuli yang diberikan. Dan pendengar dapat meminta waktu beristirahat jika ia merasa perlu.



**Gambar 3.** Grafik hubungan %Correct Word dengan Articulation Index[8]

**Mean Opinion Score (MOS)**

Pada aplikasinya sebuah uji subjektif tetap membutuhkan hasil yang kuantitatif, untuk itu dikenal istilah MOS (Mean Opinion Score). MOS cocok untuk mengolah data dari perindividu menjadi hasil kualitas untuk beberapa kondisi. Berbagai macam MOS dikenal tergantung dari[9]:

* Tes Modelitas: Mendengar/Listening (MOSLQ), Bicara/Talking (MOSTQ), atau Percakapan/Conversation (MOSCQ)
* Metode pengukuran: Instrumental (MOS­LQO untuk model berbasis sinyal dan MOSCQE untuk model berbasis parameter) dan Auditory (MOSLQS­)
* Konteks dari pengukuran: Narrow-Band (MOSLQON), WideBand (MOSLQOW) dan mix bandwidths(MOSLQOM)

**Tabel 1.** Parameter penilaian uji MOS[8]

|  |  |  |
| --- | --- | --- |
| Parameter Kualitas | *Score* | Keterangan |
| Sangat baik | 5 | Tidak membutuhkan usaha sama sekali untuk mengerti |
| Baik | 4 | Sedikit memperhatikan sudah bisa mengerti apa yang diucapkan |
| Cukup baik | 3 | Membutuhkan usaha untuk mengerti apa yang diucapkan |
| Buruk | 2 | Membutuhkan usaha yang keras untuk mengerti apa yang diucapkan |
| Sangat buruk | 1 | Tidak ada yang bisa dimengerti, bagaimanapun berusaha untuk mengerti |

Metode asli dari MOS adalah tentang bagaimana sekelompok pendengar (16 atau lebih) memberi rating serangkaian file audio dengan skala 1 sampai 5. Kemudian, rata-rata dari setiap MOS pada setiap file audio di kalkulasi. Ada dua versi dalam pengkategorian MOS, yang pertama berdasarkan ITU-T dan yang kedua berdasarkan ASCOM. Untuk di penelitian ini karena banyak standar yang dipakai dari ITU-T maka yang dipakai adalah versi ITU-T. Dalam ITU-T setiap angka memiliki kelompok yaitu, 1 untuk yang terburuk, 2 untuk Buruk, 3 Cukup Baik, 4 Baik dan 5 Sangat Baik.

1. **Metode Penelitian**

Penelitian pada tugas akhir ini dirancang pada beberapa tahapan, yang digambarkan pada diagram alir seperti pada gambar . Tahapan tersebut meliputi beberapa hal berikut:



**Gambar 4.** Diagram alir penelitian

1. Studi literatur

Tahap awal meliputi pemahaman tentang *Head Relative Transfer Function, Interaural Time Difference, Interaural Level Difference, Binary Masking*.

1. Pengumpulan Data Suara Tercampur

Pencampuran sinyal suara dilakukan dengan melakukan konvolusi antara HRIR (*head related impulse respons*) dari *CIPIC database* dengan sumber sumber suara yang digunakan.

1. Perancangan Algoritma Pemisahan Suara.

Tahap berikutnya adalah perancangan *binary mask* berdasarkan nilai ITD dan ILD terhadap *Relative Strength*.

1. Evaluasi Hasil Pemisahan Dengan PESQ, SNR, % *Correct Word* dan MOS.
2. Penyusunan Laporan.
3. **Jadwal Kegiatan**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Tabel 2**. Jadwal pelaksanaan program | | | | | | | | | | | | | | | | | |
| No | Jadwal Kegiatan Progam | Bulan  September | | | | Bulan Oktober | | | | Bulan November | | | | Bulan Desember | | | |
| 1 | Studi literatur |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 2 | Pengumpulan Data HRTF |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | Perancangan Program dan Simulasi *Estimated Binary Masking* |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 4 | Evaluasi Hasil Pemisahan dengan Metode PESQ, SNR, %*Correct Word* dan MOS |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 5 | Penyusunan Laporan |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

1. **Daftar Pustaka**

[1]Tris Atmaja, Bagus. 2012. “Pemisahan Sumber-Sumber Suara Tercampur Dari Input Binaural” , Tugas Akhir, Jurusan Teknik Fisika Institut Teknologi Sepuluh Nopember Surabaya.

[2] D. Wang and G. J. Brown. 2006. Computational Auditory Scene Analysis: Principles, Algotithms, and Application. Wiley-IEEE Press.

[3] Roman. Nicoleta, Wang. Deliang, and J. Brown. Guy.,"Speech Segretion Based On Sound Localization". Acoustical Society of America, vol.114, no.4, Oct 2003.

[4] Hugeng. 2011. “Individualisasi Model Parametrik *Head Related Transfer Functions*” , Disertasi, Jurusan Teknik Elektro Universitas Indonesia.

[5] Rayleigh, L., “On our perception of sound direction”, Philosophical magazine, vol. 3, pp. 214 – 232, 1907

[6] <http://earlab.bu.edu/databases/collections/cipic/Default.aspx> (Diunduh 25/10/2013)

[7] Telecommunication Standardization Sector Of ITU, 1994, " *P.85: A Method For Subjective Performance Assessment Of The Quality Of Speech Voice Output Devices*", ITU-T

[8] <http://www.mcsquared.com/classrooms2.htm> (Diunduh 11/11/2013)

[9] Vlahodimitropoulos K., Katsaros E., 2007 **"***Monitoring The End User Perceived Speech Quality Using The Derivative Mean Opinion Score (MOS) Key Performance Indicator***"**, IEEE International Symposium on Personal PIMRC'07