**BAB II**

**TINJAUAN PUSTAKA**

**2.1 *Head Related Transfer Function***

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 (Hugeng, 2011).

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:

(2.1)

(2.2)

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 (Hugeng, 2011).

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) (Hugeng, 2011).

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 (Hugeng, 2011).

Sumber Suara

Suara yang diterima

telinga kanan

Suara yang diterima

telinga kiri

HRTF telinga kiri

HRTF telinga kanan

**Gambar 2.1** Ilustrasi HRTF pada Telinga Kanan dan Telinga Kiri

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 (Hugeng, 2011).

**2.2 *Binaural Cue***

Lord Rayleigh (1907) dalam teorinya, terdapat dua *binaural cue* utama yakni *interaural time difference* (ITD) dan *interaural level difference* (ILD)

**2.2.1 *Interaural Time Difference***

*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 telinga dengan elevasi 0o. Dengan demikian ITD merupakan fungsi dari azimuth pada bidang-bidang dengan elevasi tetap. Menurut Jeffress dalam penelitiannya (1948), ITD dapat juga dihitung sebagai waktu tunda yang bersesuaian terhadap posisi nilai maksimal dari hasil korelasi silang antara sepasang HRTF pada kanal frekuensi ke-, *time frame* ke-, dan *lag* ke-.

(2.3)

Dimana , adalah HRTF telinga kiri dan telinga kanan pada kanal frekuensi ke-, dan , adalah nilai rata-rata HRTF pada *integration window ke-.*

**2.2.1 *Interaural Level Difference***

Lord Rayleigh (1907) 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 (1907) mengatakan bahwa ILD dan ITD yang diambil bersama-sama memberikan informasi lokalisasi sepanjang jangkauan frekuensi yang terdengar.

**2.3 *Cochlear Filtering***

*Cochlear Filtering* dapat dimodelkan sebagai kumpulan dari *bandpass-filter (filterbank)*. *Filterbank* yang digunakan adalah *Patterson-Holdsworth's Filter* (Slaney, 1993).

**2.3.1 *Patterson-Holdsworth's ERB Filterbank***

Basilar membran dapat dibagi menjadi beberapa *critical-band*, dengan setiap *critical-band* dapat dimodelkan oleh *Equivalent Rectangular Bandwidth* (ERB). Glasberg dan Moore (1990) mengemukakan persamaan berikut untuk menghitung *bandwidth* pada frekuensi tertentu .

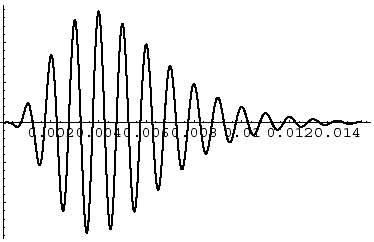
(2.4)

Model tersebut diatas menggunakan *gammatone-filter* yang didefinisikan sebagai berikut.

(2.5)

Dimana dan adalah orde dari filter dan *bandwidth* dari filter yang besarnya , adalah *center-frequency*,adalah fasa, adalah waktu.

Sebagai contoh, filter bersesuaian dengan *center-frequency* 1.000 Hz dengan dan . Respon impuls dari *gammatone-filter* ini adalah pada Gambar 2.2.

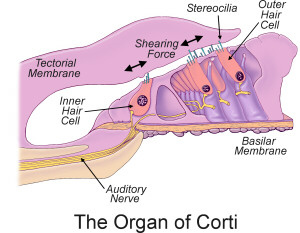


**Gambar 2.2** Respon impuls dari *gammatone-filter* orde-4 pada *center-frequency* 1000 Hz (Slaney, 1993).

Perhatikan Gambar 2.2**,** grafik tersebut disebut sebagai fungsi *gammatone* karena merupakan kombinasi antara distribusi *gamma* dan gelombang sinusoidal.

**2.3.2 *Hair Cell***

Didalam organ corti terdapat sel-sel yang berperan penting dalam pendengaran, yaitu *hair cell.* Ada 2 jenis *hair cell* yaitu *inner hair cell* dan *outer hair cell.* Fungsi dari *hair cell* adalah mengubah gelombang suara menjadi sinyal listrik. Sinyal ini kemudian diteruskan ke otak melalui a*uditory nerve* (Gambar 2.3)*.*

**

**Gambar 2.3** Gambar Organ Corti pada bagian dalam Koklea (Gong, 2005)

Meddis dalam penelitiannya (1986) memodelkan *hair cell* dalam diagram blok berikut:

**Gambar 2.4** Diagram blok dari model *hair cell* oleh Meddis (Meddis,1986)

Pada Gambar 2.4, saat sinyal suara datang, *transmitter* mulai bergerak dari *free transmitter pool* menuju *cleft. Hair cell* akan kehabisan *transmitter.* Dalam waktu yang singkat, sebagian *transmitter* akan kembali dari *cleft* menuju *reprocessing store* lalu ke *free transmitter pool*. Karena *hair cell* masih kekurangan *transmitter* maka *factory* akan membuat *transmitter.* Ketika jumlah *transmitter* yang hilang sama dengan jumlah *transmitter* yang dihasilkan *factory,* maka kesetimbangan tercapai dan jumlah *transmitter* pada *synaptic cleft* adalah *firing rate* yang merepresentasikan aktifitas pada *auditory-nerve.*

**2.4 Kualitas Suara**

Suara dalam hal ini percakapan manusia adalah salah satu metode komunikasi bahasa selain membaca. Terdapat dua proses yang terjadi dalam setiap percakapan manusia saat ini yaitu, berbicara dan mendengar yang keduanya jika digabungkan maka disebut percakapan. Suara dari pembicara diterima oleh telinga manusia 'suara' tadi diolah alat pendengaran manusia. Hasil pendengaran bergantung pada keadaan emosi seseorang, presepsi seseorang dan tambahan informasi lainnya tentang konteks kalimat suara tersebut. Karena tidak adanya alat yang dapat merepresentasikan apa yang diterima oleh manusia itu sendiri, maka subjek manusia adalah satu-satunya alat pengukuran yang paling tepat.

****

**Gambar 2.5** Skematis yang merepresentasikan tes subjektif (Raake, 2006)

Skematis pada Gambar 2.5 menjelaskan bahwa, masukan dari tes subjektif disebut *sound event*, diproses oleh *listener* (pendengar), kemudian luarannya adalah presepsi. Pada proses pendengar, terdapat *auditory event* yaitu suara diterima telinga dan akan dikirim ke otak. Dan presepsi yang ada pada luaran adalah hasil otak memproses *auditory event* tadi.

Dalam sebuah percakapan, terdapat tahap-tahap yang menjelaskan bagaimana pesan dari pengirim pesan dapat tersampaikan kepada si pendengar. Tahapan-tahapan tersebut adalah '*Comprehensibility*' yaitu bagaimana isi pesan bisa dime-ngerti oleh pendengar, '*Intelligibility*' yaitu bagaimana kejelasan suara yang diucapkan dalam penyampaian pesan, '*Communi-cability*' yaitu bagaimana pesan yang diterima dapat diteruskan dengan baik ke orang lain. Dari ketercapaian penyampaian informasi itulah kualitas suara dibagi dalam 3 kategori, *Comprehensibility, Intelligibility* dan *Communicability*.

****

**Gambar 2.6** Terminologi kualitas suara (Raake, 2006)

Semua tahapan tersebut memiliki syarat-syarat untuk mencapainya dan urutan kualitasnya berurutan dari '*Comprehensibility*', '*Intelligibility*' dan '*Communicability*' yang kemudian dikenal istilah '*Comprehension*' yaitu hasil dari presepsi si pendengar tentang pemahaman yang komunikatif dan memang sebelumnya si pendengar memang berkeinginan mendengar dan mengerti kondisi dari percakapan yang dilakukan (lihat Gambar 2.6). Sesuai dengan *'Listener Factor*' atau Faktor yang ada pada pendengar, jika yang ingin diukur adalah *Comprehensibility* maka dalam ujian diberikan kata-kata yang saling tidak berhubungan dan bukan dari bidang si pendengar, '*Intelligibility*' maka konteks kata-kata yang akan diberikan sebelumnya memang sudah familiar dengan pendengar, '*Communicability*' maka situasi pembicaraan dan konteksnya telah dipahami betul oleh pendengar. Dan dalam penelitian ini yang digunakan adalah pengukuran '*Intelligibility*' (Raake, 2006).

Selain *Intelligibility*, faktor lain adalah kenyamanan. Jika informasi yang diterima baik, maka selanjutnya kualitas suara yang diuji adalah tingkat kenyamanannya. Tingkat kenyamanan bergantung dari persepsi manusia dan nilainya berbentuk kualitatif, sedangkan dalam pengukuran data yang dibutuhkan adalah data kuantitatif. Sehingga dibutuhkan cara atau metode untuk mengkuantifikasikan data dari kualitatif menjadi kuantitatif atau terukur.

Pengukuran Auditory atau subjektif yang paling akurat adalah saat pengguna menilainya di lingkungan umum. Secara teori seperti itu, namun faktanya hal tersebut sangat sulit diimplementasikan maka penelitian dilakukan disebuah tempat yang mana desainnya direncakan dan kondisinya dapat dikontrol. Dijelaskan oleh ITU-T syarat dan kondisi apa yang dapat di pakai dalam metode subjektif tes yaitu sebagai berikut (ITU-T, 1996):

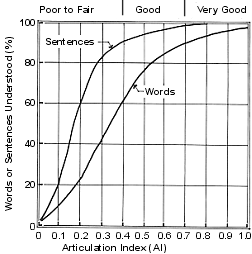
* Ruang yang digunakan harus memiliki waktu dengung kurang dari 500ms dan derau latar dibawah 30dBA
* Kalimat yang digunakan berupa kalimat sehari-hari yang dapat dipahami subjek
* Subjek tidak pernah mengikuti percobaan seperti ini selama minimal 6 bulan
* Subjek belum pernah mendengar kalimat yang sama dengan yang diuji sebelumnya
* Sebelum diberikan tes, subjek dilatih dengan kalimat yang berbeda

Jenis penilaian kualitas suara yang diusulkan dalam penelitian ini adalah *Percent Correct Words* untuk pengukuran subjektif dan *Signal to Noise Ratio* (SNR) untuk pengukuran Objektif.

**2.4.1 *Percent Correct Words***

Penilaian *Percent Correct Words* didasari pada seberapa banyak (dalam persentase) 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. 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. Perolehan nilai kemudian di cocokan dengan grafik hubungan *Percent Correct Words* dengan Articulation Index sesuai dengan ANSI S3.5 1969 untuk mengetahui tingkat Speech Intelligibility (ANSI, 1969).



**Gambar 2.7** Grafik Hubungan Percent Correct Word dengan Articulation Index (ANSI, 1969)

**2.4.2 *Signal to Noise Ratio* (SNR)**

*Signal to noise ratio* (SNR) merupakan salah satu metode pengukuran *speech intelligibility* secara objektif. Metode ini menghitung ratio antara sinyal yang diinginkan terhadap *noise* yang diterima. Semakin besar nilai SNR maka semakin baik kualitas suatu suara. Nilai SNR dapat dihitung menggunakan rumus berikut.

(2.6)

Dimana adalah sinyal target sebelum tercampur dan adalah sinyal target hasil pemisahan.