**BAB IV**

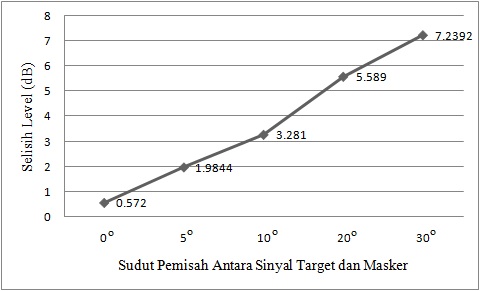
**ANALISA DATA DAN PEMBAHASAN**

Pada bagian ini akan dibahas hasil dari implementasi masing-masing tahap yang telah dijabarkan dalam BAB III.

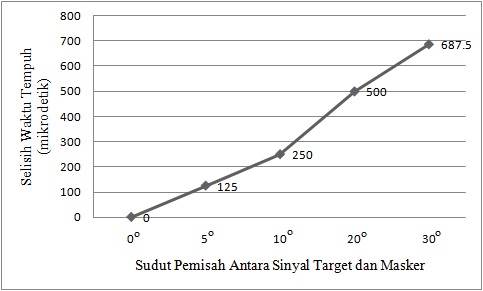
* 1. **Proses Data Training**
     1. Hasil Simulasi *Spatial Hearing*

Sinyal suara monaural yang digunakan merupakan data suara hasil rekaman pada ruang *anechoic* yang saat menggunakan fungsi *audioread* pada matlab, amplitudo dari suara tersebut terbatas pada rentang -1 hingga +1. Amplitudo tersebut lebih kecil dari amplitudo suara asli. Oleh karena itu, amplitudo akan ditingkatkan skalanya hingga rentang aslinya dengan mengalikan amplitudo tersebut sebesar 215 dalam domain waktu. Sinyal monaural yang telah dinormalisasi, kemudian dikonvolusi terhadap data HRTF untuk menghasilkan sinyal *binaural.*

Hasil dari proses simulasi sinyal *binaural* pada azimut 0o, 5o, 10o, 20o, dan 30o adalah terjadi perbedaan level loudnessdan perbedaan waktu tempuh antara sinyal suara yang diterima oleh telinga kiri dan telinga kanan. Hasil dapat dilihat pada grafik dibawah ini.



**Gambar 4.1** Selisih Level *Loudness* Antara Telinga Kiri dan Telinga Kanan



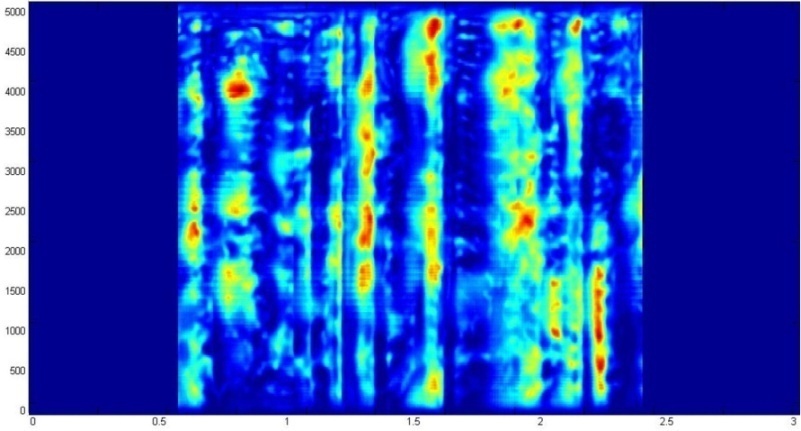
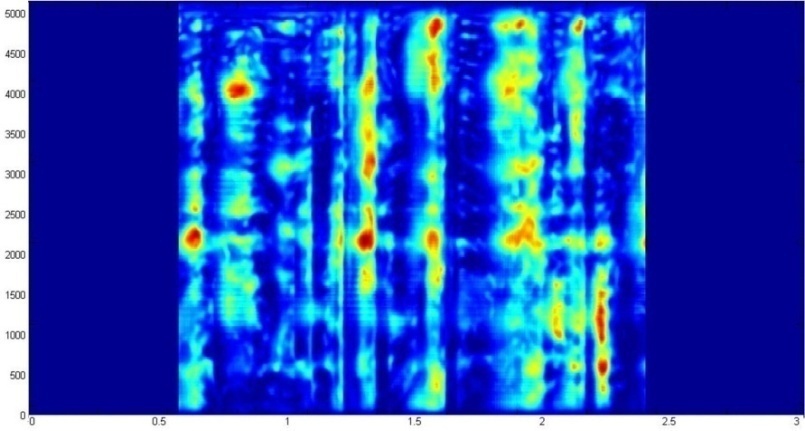
**Gambar 4.2** Selisih Waktu Tempuh Antara Telinga Kiri dan Telinga Kanan

Pada Gambar 4.1, nilai selisih level loudnesspada sudut 0o hingga sudut 30o meningkat. Begitu pula pada Gambar 4.2, nilai selisih waktu tempuh pada sudut 0o hingga sudut 30o juga meningkat. Hal ini terjadi karena saat sumber suara tidak berada pada bidang median (sudut 0o), yaitu pada sudut 5o, 10o, 20o, dan 30o, maka gelombang suara akan ditransmisikan dari sumber suara menuju telinga kiri akan mengalami hambatan dan terhalang oleh kepala. Selain itu, jarak antara sumber suara terhadap telinga kanan semakin kecil, dan sebaliknya untuk telinga kiri. Sehingga semakin jauh sumber suara dari bidang median maka semakin besar suara yang terhambat atau terhalang oleh kepala dan semakin jauh pula jarak sumber suara terhadap telinga kiri. Peristiwa ini disebut efek bayangan kepala (*headshadow effect*) yang menyebabkan adanya ILD (*Interaural Level Difference)* yaitu perbedaan level pada kedua telinga dan ITD (*Interaural Time Difference*) yaitu perbedaan waktu tempuh pada kedua telinga.

Selanjutnya, hasil dari simulasi *spatial hearing* dilakukan analisa *intensity, frequency contrast* dan *temporal contrast* dalam unit *time-frequency* yang disebut *auditory saliency map.* Hasil dari *auditory saliency map* ditampilkan pada Gambar 4.3, Gambar 4.4, Gambar 4.5 dan Gambar 4.6.

Gambar 4.3 merupakan gambar dari *saliency map* untuk sudut 5o. Pada gambar tersebut dengan garis kotak berwarna merah, terlihat perbedaan intensitas suara antara telinga kiri (Gambar 4.3.a) dan telinga kanan (Gambar 4.3.b) pada rentang frekuensi 2000 Hz hingga 2300 Hz. Selain itu pada Gambar 4.3 tersebut dengan garis kotak berwarna hitam, peningkatan intensitas juga terlihat namun tidak terlalu signifikan seperti pada Gambar 4.3 yang bergaris warna merah yaitu pada rentang frekuensi 3000 Hz hingga 3300 Hz. Dengan meningkatnya sudut antara sumber suara terhadap bidang median yaitu 5o (Gambar 4.4), 10o (Gambar 4.5), 20o (Gambar 4.6), 30o (Gambar 4.7) maka perubahan intensitas suara pada rentang frekuensi tersebut semakin meningkat. Namun untuk sudut 30o yaitu pada Gambar 4.7 dengan garis kotak berwarna merah putus-putus, perubahan intensitas suara juga terlihat pada rentang frekuensi 3300 Hz hingga 3500 Hz, sehingga pada sudut tersebut perubahan intensitas terjadi pada tiga rentang frekuensi yang berbeda yaitu 2000 Hz hingga 2300 Hz, 2500 Hz hingga 2700 Hz dan 3000 hingga 3300 Hz.

Selain analisa *saliency map* dalam domain *time-frequency* juga dilakukan analisa *waveform* dalam domain waktu. Analisa yang dilakukan adalah *saliency response* dalam domain waktu dan perseptual onset yang diperoleh dari *saliency response* tersebut. *Saliency response* diperoleh berdasarkan nilai *skewness* dari distribusi energi dalam *overlapping* dan *sliding time windows-*nya. Hasil dari analisa *waveform* dapat dilihat pada Gambar 4.7 hingga Gambar 4.10. Pada Gambar 4.7 hingga Gambar 4.10, *waveform* ditunjukkan dalam garis berwarna biru, *saliency response* ditunjukkan dalam garis berwarna hitam dan perseptual onset merupakan puncak dari *saliency response* yang ditunjukkan dengan titik berwarna merah. Pada Gambar 4.7, yaitu sumber suara pada sudut 5o dari bidang median, terlihat nilai *saliency response* dan perseptual onset pada telinga kanan (Gambar 4.7.b) lebih tinggi daripada telinga kiri (Gambar 4.7.a). Dengan meningkatnya sudut antara sumber suara terhadap bidang median yaitu 10o pada Gambar 4.8, 20o pada Gambar 4.9 dan 30o pada Gambar 4.10, terlihat perbedaan perseptual onset dan *saliency response* antara telinga kanan dan telinga kiri semakin meningkat begitu pula untuk amplitud dari sinyal suaranya. Sehingga dapat disimpulkan perseptual onset dan *saliency response* meningkat selaras dengan peningkatan amplitude dari sinyal suara. Dan semakin besar sudut antara sumber suara terhadap bidang median maka perbedaan nilai perseptual onset dan *saliency response* juga meningkat.



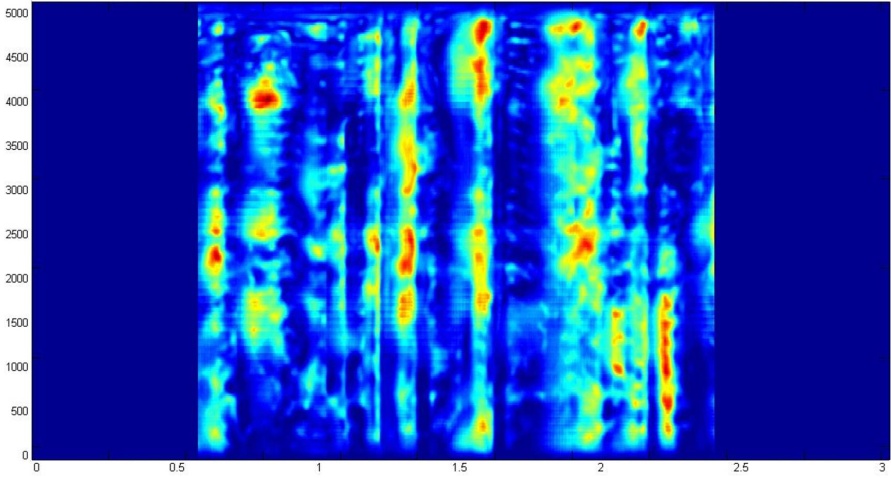
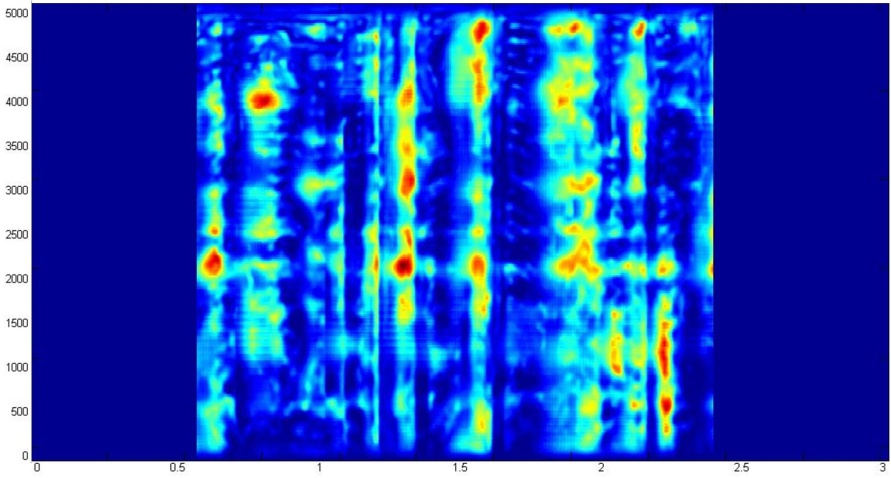
**Gambar 4.3** *Saliency Map* pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut 5o

Waktu (detik)

Waktu (detik)

Frekuensi (Hz)

Frekuensi (Hz)



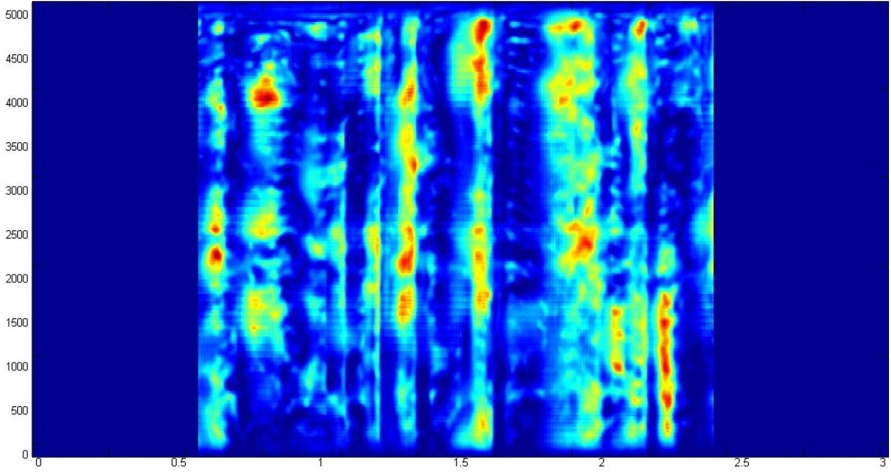
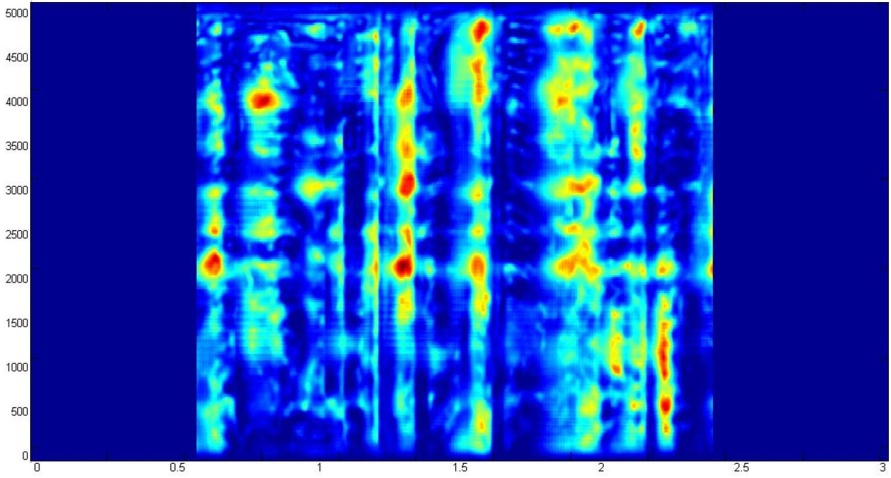
**Gambar 4.4** *Saliency Map* pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut 10o

**Waktu (detik)**

**Waktu (detik)**

**Frekuensi (Hz)**

**Frekuensi (Hz)**



**Waktu (detik)**

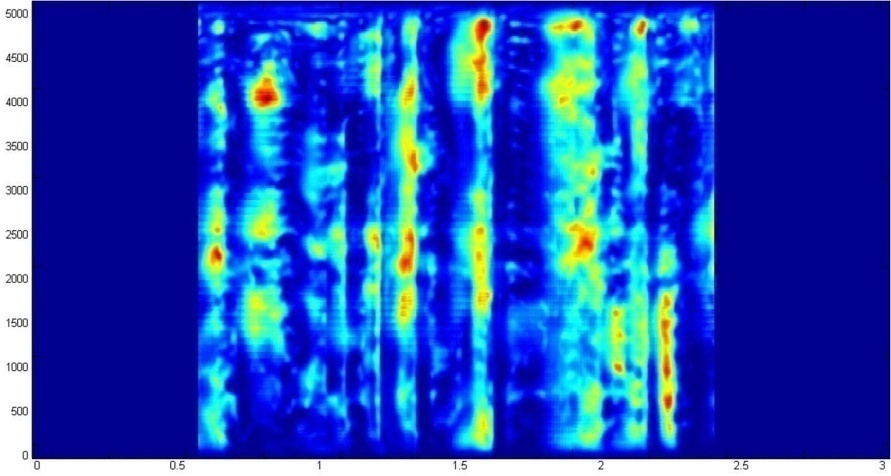
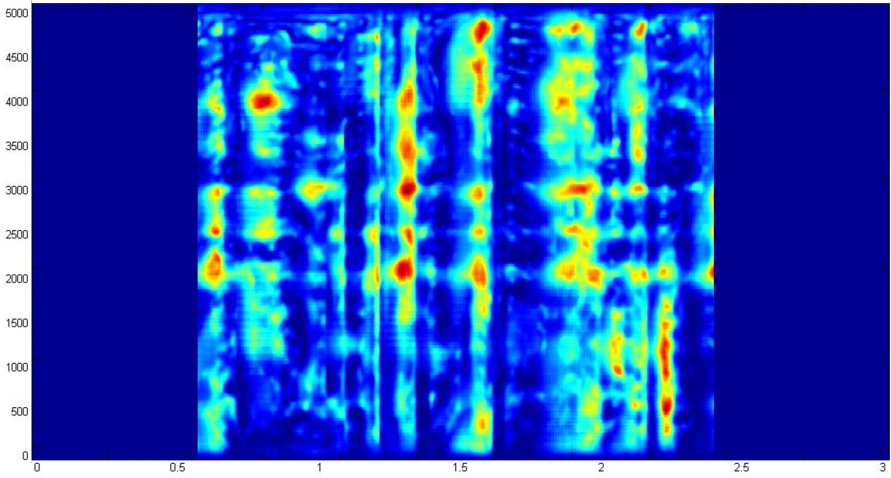
**Frekuensi (Hz)**

**Frekuensi (Hz)**

**Gambar 4.5** *Saliency Map* pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut 20o

**Waktu (detik)**

**Frekuensi (Hz)**

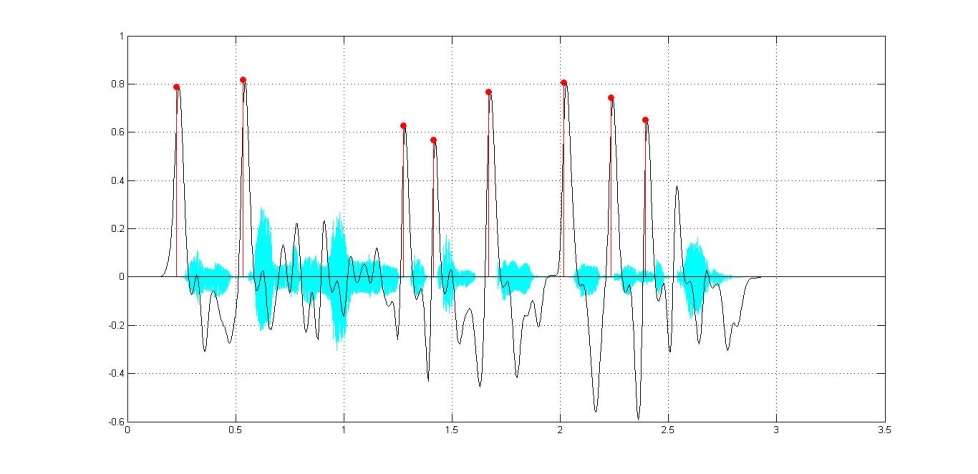
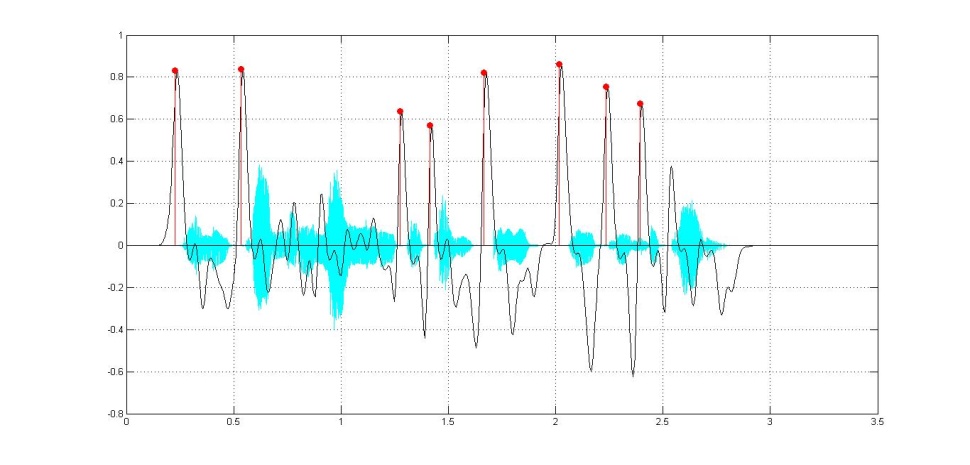


**Gambar 4.6** *Saliency Map* pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut 30o

**Waktu (detik)**

**Waktu (detik)**

**Frekuensi (Hz)**



**Saliency**

**Perseptual Onset**

**Saliency**

**Gambar 4.7** *Waveform, Saliency* dan Perseptual Onset pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut Pemisah 5o

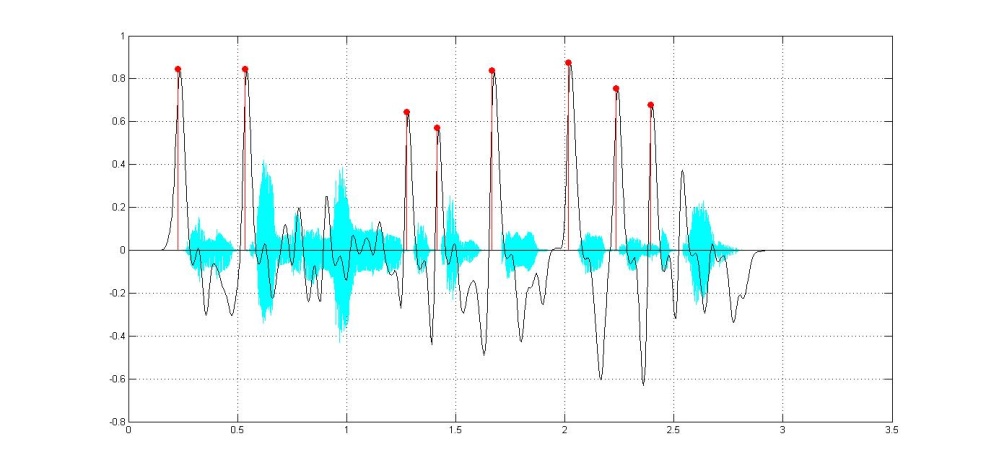
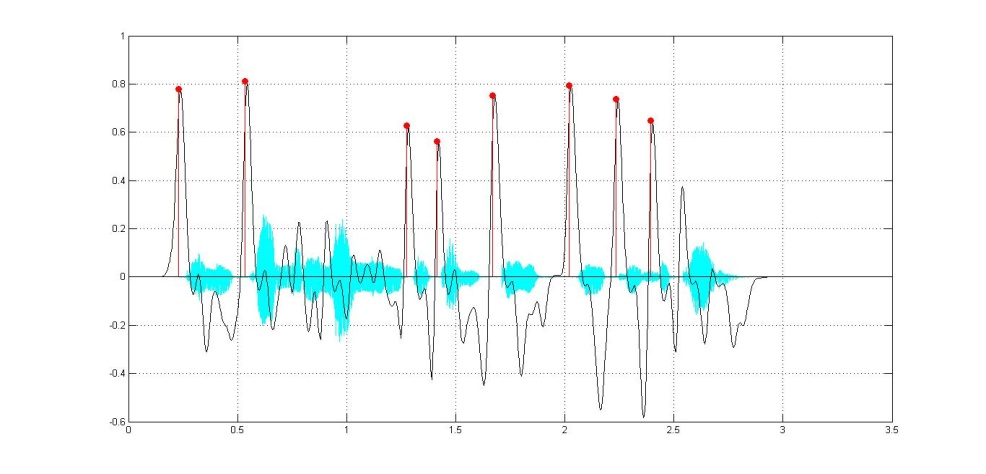
**Perseptual Onset**

**Waktu (detik)**

**Amplitude**

**Amplitude**

**Waktu (detik)**



**Perseptual Onset**

**Amplitude**

**Saliency**

**Waktu (detik)**

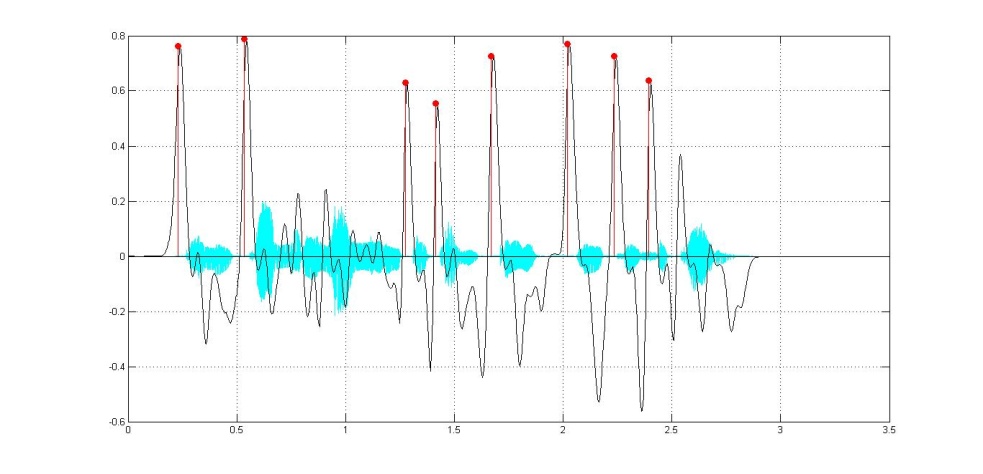
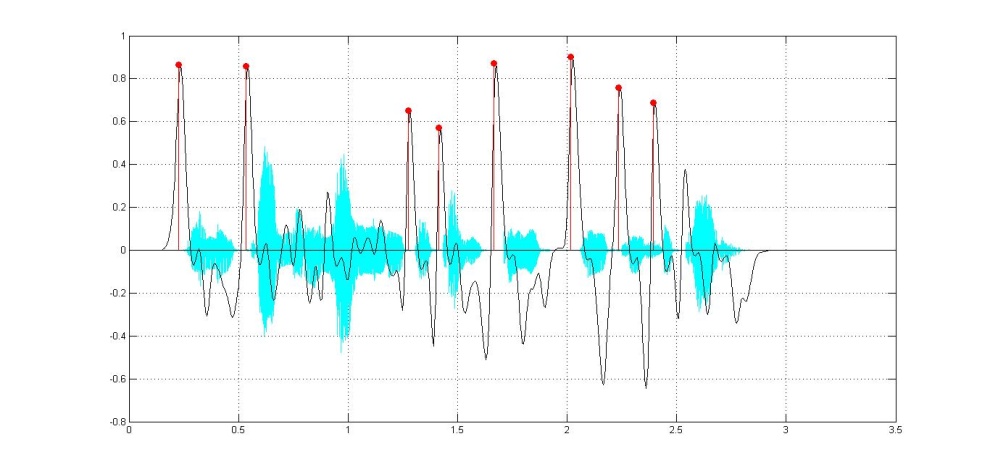
**Perseptual Onset**

**Amplitude**

**Saliency**

**Waktu (detik)**

**Gambar 4.8** *Waveform, Saliency* dan Perseptual Onset pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut Pemisah 10o



**Gambar 4.9** *Waveform, Saliency* dan Perseptual Onset pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut Pemisah 20o

**Amplitude**

**Saliency**

**Waktu (detik)**

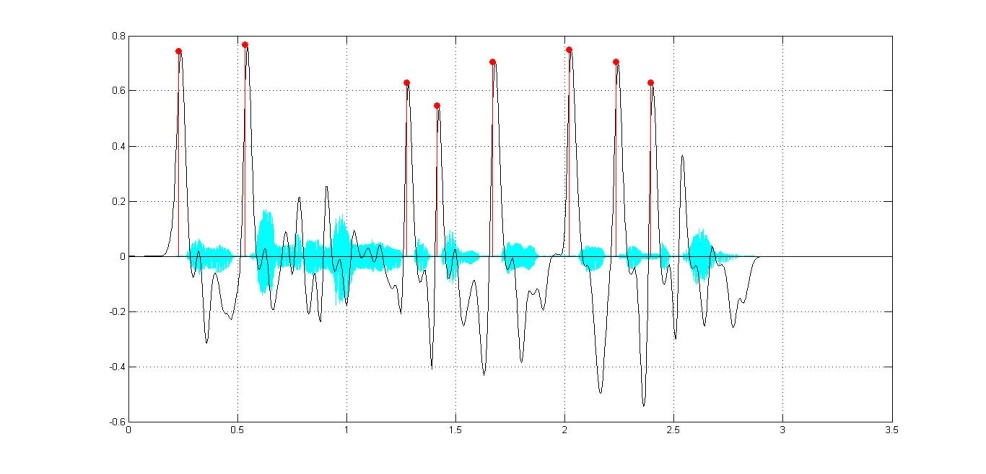
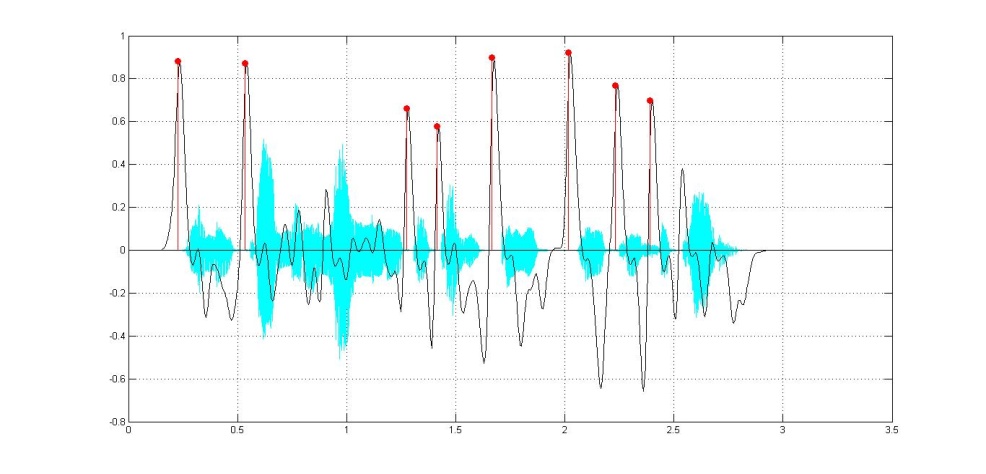
**Waktu (detik)**

**Perseptual Onset**

**Saliency**

**Amplitude**

**Perseptual Onset**



**Waktu (detik)**

**Amplitude**

**Saliency**

**Perseptual Onset**

**Waktu (detik)**

**Amplitude**

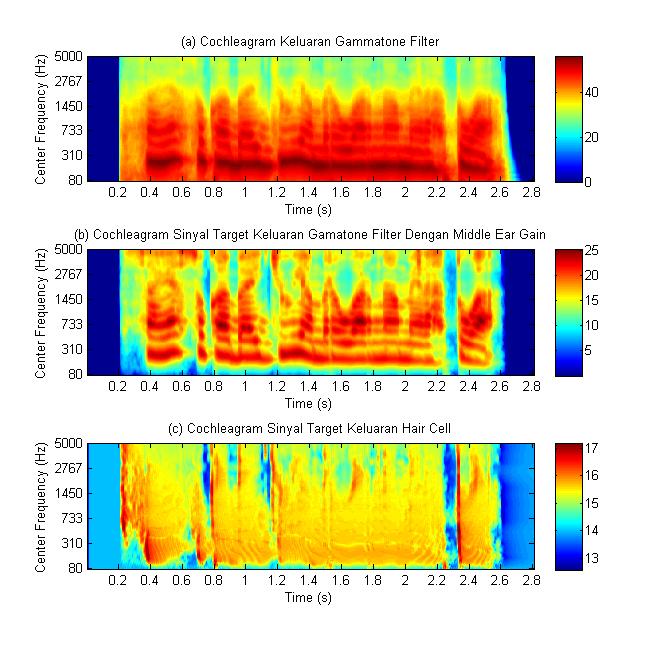
**Saliency**

**Perseptual Onset**

**Gambar 4.10** *Waveform, Saliency* dan Perseptual Onset pada (a) Telinga Kiri dan (b) Telinga Kanan dengan Sudut Pemisah 30o

* + 1. ***Auditory Periphery***

Hasil dari proses simulasi *spatial hearing,* suara yang diterima pada setiap telinga hanya mensimulasikan proses transmisi suara pada telinga bagian luar dan tengah sehingga perlu dilakukan simulasi telinga bagian dalam. Dalam penelitian ini, proses transmisi suara pada telinga bagian dalam disimulasikan menggunakan *auditory periphery* yaitu proses perubahan sinyal suara menjadi sinyal listrik yang akan dikirim ke otak. Hasil dari proses *auditory periphery* ditunjukkan pada Gambar 4.11 berikut.



**Gambar 4.11** Cochleagram Sinyal Target

Pada Gambar 4.11 (a) yang dilingkari, level pada rentang frekuensi 2000 Hz hingga 5000 Hz menunjukkan nilai yang lebih kecil daripada level pada rentang frekuensi dibawah 2000 Hz. Seharusnya pada pendengaran manusia, untuk rentang frekuensi 2000 Hz hingga 5000 Hz memiliki sensitifitas yang lebih tinggi, maka diperlukan pemberian *gain* untuk memberikan efek *middle ear.* Hasil *gammatone filter* dengan penambahan *gain* pada *cochleagram* pada Gambar 4.11 (b) yang dilingkari dengan garis hitam putus-putus mengalami kenaikan level pada rentang 2000 Hz hingga 5000 Hz dan pada Gambar 4.11 (b) yang dilingkari dengan garis merah putus-putus mengalami penurunan level. Pada Gambar 4.11 (c) merupakan hasil dari proses *hair cell* merupakan nilai *firing rate* dari aktivitas *hair cell* pada *basilar membrane*. Pada Gambar 4.11 (c) yang dilingkari garis hitam terjadi *initial peak* dari nilai *firing rate* yang tinggi dalam waktu yang singkat dan semakin lama nilainya semakin turun dan *steady* hingga pada lingkaran garis merah *initial peak* terjadi lagi. Ini merepresentasikan bahwa saat suara tepat terdengar yaitu pada saat *initial peak* terjadi, tepat saat itu *hair cell* melepaskan *transmitter* kepada *cleft* dan dalam waktu yang singkat jumlah *transmitter* yang dilepaskan oleh *hair cell* memiliki jumlah yang sama dengan *transmitter* yang tersisa dalam *hair cell.*

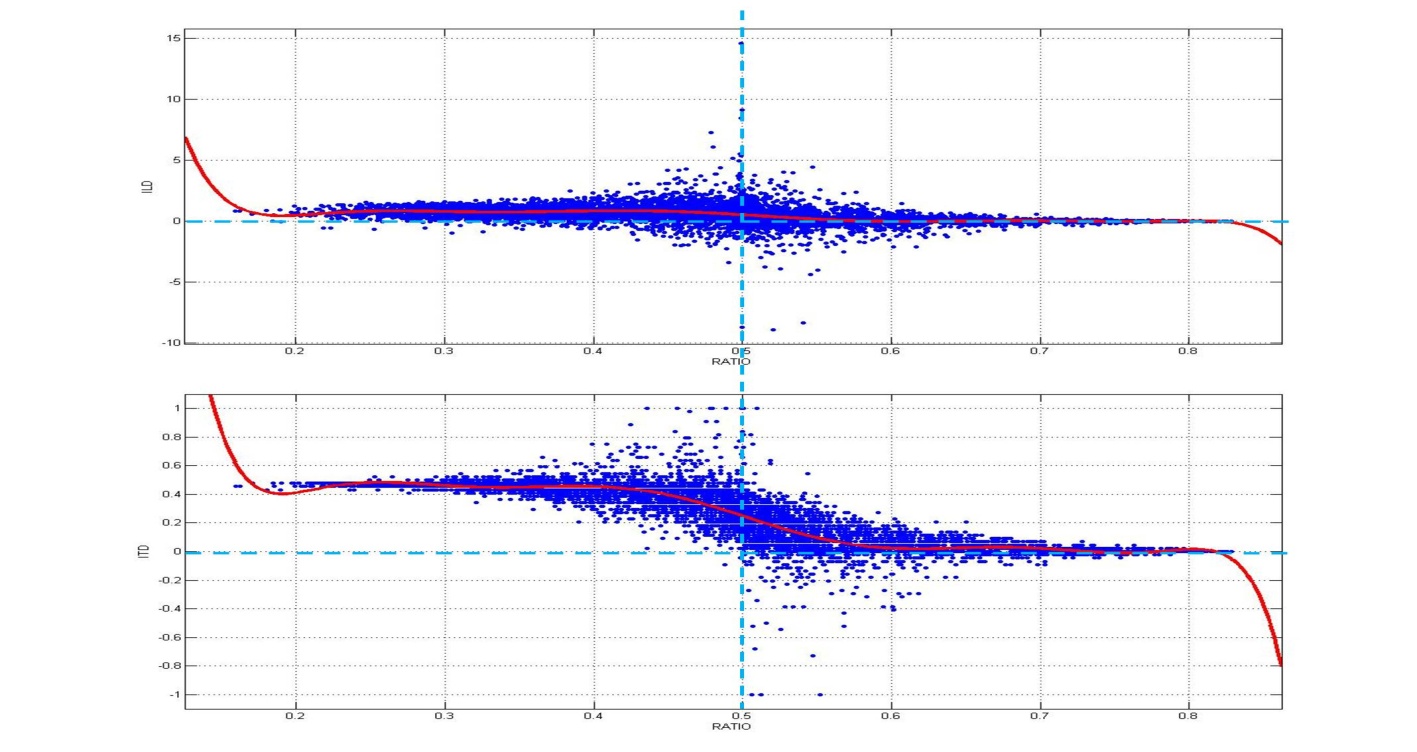
* + 1. **Perhitungan *Binaural Cue* dan *Relative Strength***

Hasil perhitungan *binaural cue* dan *relative strength* yang didapat pada seluruh percobaan, dilakukan *scatterplot* pada setiap kanal frekuensinya. *Scatterplot* yang dilakukan adalah antara nilai *interaural level difference* (ILD) terhadap *relative strength* pada setiap *center frequency*-nya dan nilai *interaural time difference* (ITD) terhadap nilai *relative strength-*nya. Dari *scatterplot* tersebut diperoleh hubungan antara *binaural cue* terhadap *relative strength* yaitu fungsi peralihan spasial dari telinga kiri ke telinga kanan akibat perubahan posisi relatif kiri (diam) terhadap kanan (bergerak). Fungsi peralihan yang diperoleh berupa persamaan polinomial pada proses *fitting* dari *scatterplot.* Fungsi peralihan yang diperoleh adalah sebagai berikut:

(4.1)

Gambar 4.12 adalah grafik *scatterplot* pada *center frequency* 80 Hz. Pada gambar tersebut, garis berwarna merah merupakan grafik dari fungsi peralihan yang diperoleh dari proses *fiiting* dari *scatterplot*.Garis putus berwarna hijau horizontal adalah representasi dari bidang median dari kepala. Daerah yang berada diatas garis tersebut merepresentasikan daerah disebelah kanan kepala dan daerah disebelah bawah garis tersebut merepresentasikan daerah disebelah kiri kepala. Sedangkan garis putus-putus vertikal berwarna hijau merepresentasikan nilai tengah dari *relative strength* sehingga daerah yang berada disebelah kiri merupakan daerah dimana sinyal masker lebih dominan dari pada sinyal target dan sebaliknya untuk daerah kanan garsi vertikal tersebut.

Pada Gambar 4.12 (a) terlihat perubahan posisi relatif kiri (diam) terhadap kanan bergerak tidak mempengaruhi fungsi peralihan (garis berwarna merah) untuk nilai ILD terhadap *relative strength*. Hal ini ditunjukkan dengan fungsi peralihan yang diperoleh yang linier pada nilai *relative strength* sebesar 0.1 hingga 0.9. Dibandingkan dengan fungsi peralihan untuk nilai ITD terhadap *Relative Strength* (Gambar 4.12b)*,* fungsi peralihan yang diperoleh menunjukkan pola tertentu yaitu pada nilai *relative strength* 0.2hingga 0.4 linier pada ITD 0.4 ms, menurun hingga 0 ms pada relative strength 0.4 hingga 0.6 dan tetap pada 0 ms hingga *relative strength* 0.9. Namun pada Gambar 4.13 (a), yaitu scatterplot pada *center frequency* 1000 Hz, terlihat fungsi peralihan memiliki pola yang sama dengan Gambar 4.13 (b). Semakin tinggi *center frequency* maka fungsi peralihan untuk nilai ILD terhadap *relative strength* yang semula linier perlahan menunjukkan pola yang sama seperti fungsi peralihan untuk nilai ITD terhadap *Relative Strength* (Gambar 4.11)*.*

**

**Telinga Kiri**

**Telinga**

**Kanan**

**f(x)**

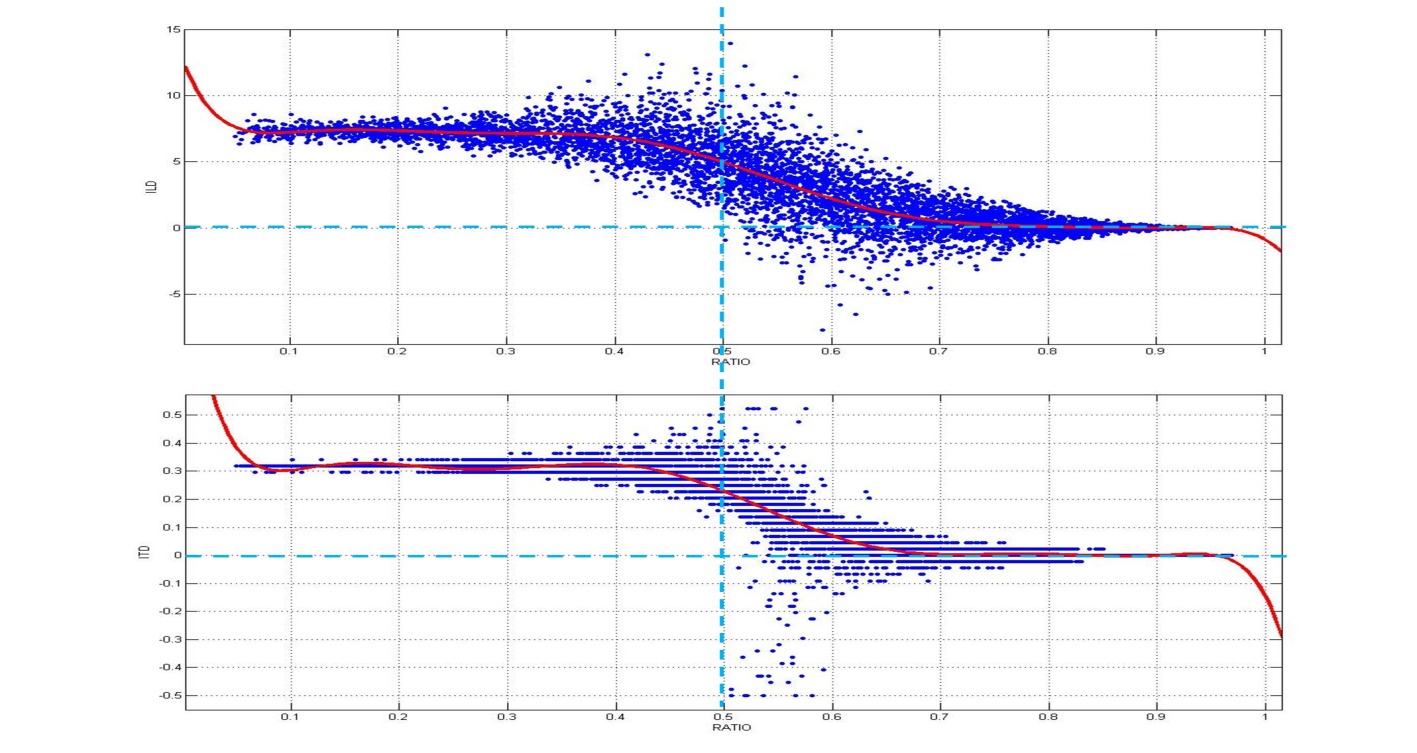
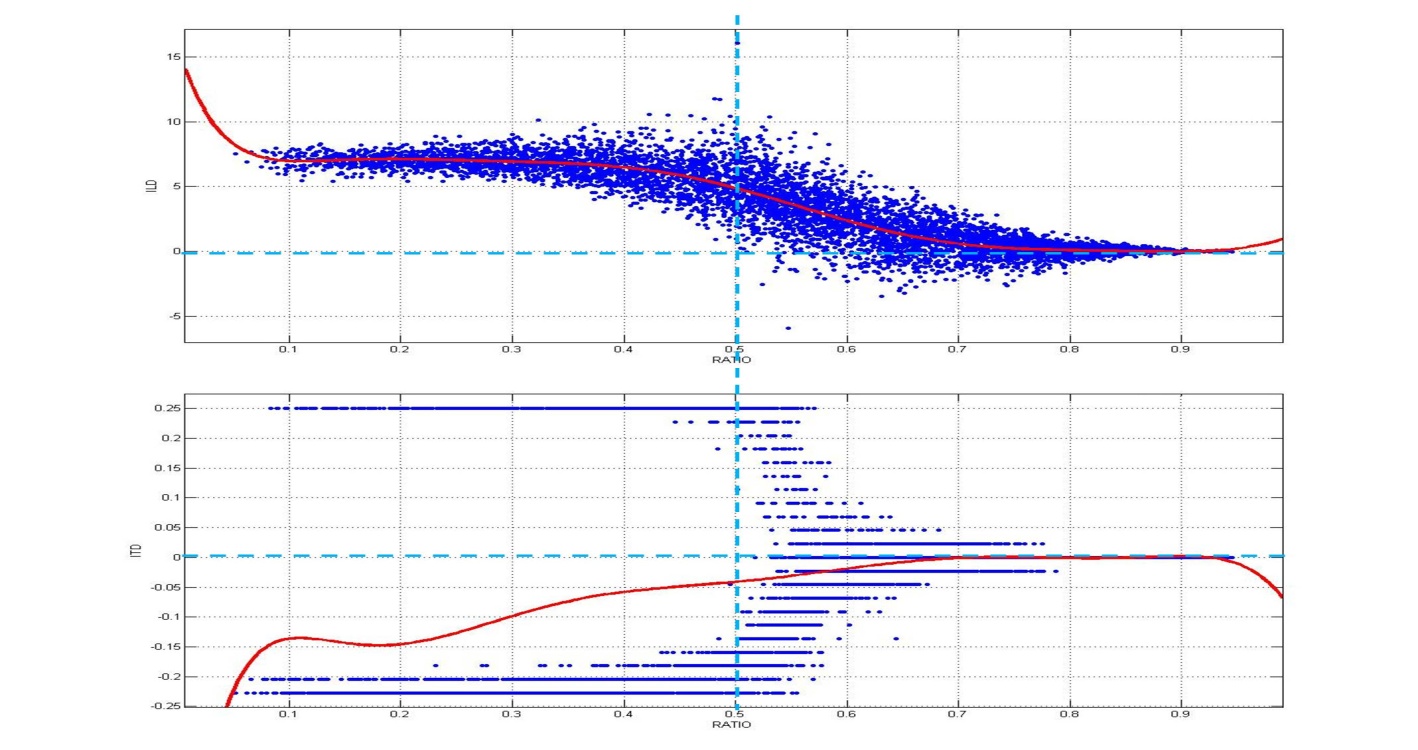
**f(x)**

**Telinga Kiri**

**Telinga**

**Kanan**

**Gambar 4.12** *Scatterplot (a) Interaural Level Difference* (ILD)terhadap *Relative Strength* (RATIO) dan (b) Interaural Time Difference (ITD) terhadap Relative Strength (RATIO) pada *Center Frequency* 80 Hz

* *

**f(x)**

**f(x)**

**Telinga**

**Kanan**

**Telinga Kiri**

**Telinga Kiri**

**Telinga**

**Kanan**

**Gambar 4.13** Scatterplot (a) Interaural Level Difference (ILD) terhadap Relative Strength (RATIO) dan (b) Interaural Time Difference (ITD) terhadap Relative Strength (RATIO) pada Center Frequency 1000 Hz

**Gambar 4.14** Scatterplot (a) Interaural Level Difference (ILD) terhadap Relative Strength (RATIO) dan (b) Interaural Time Difference (ITD) terhadap Relative Strength (RATIO) pada Center Frequency 2000 Hz

**f(x)**

**f(x)**

**Telinga Kiri**

**Telinga**

**Kanan**

**Telinga**

**Kanan**

**Telinga Kiri**

**Telinga Kiri**

**Telinga Kiri**

**Telinga**

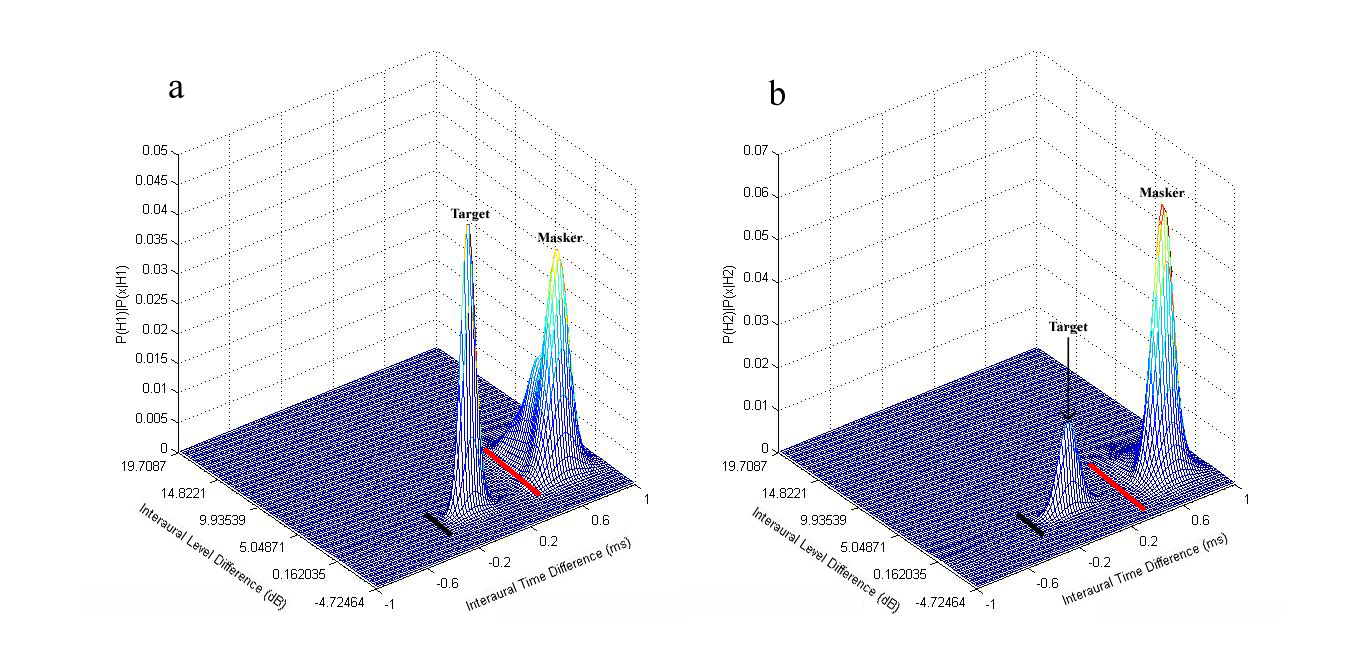
**Kanan**

Namun pada *center frequency* 2000 Hz, fungsi peralihan untuk nilai ITD terhadap *Relative Strength* tidak memiliki pola yang sama seperti pada center frequency sebelumnya (Gambar 4.12). Hal ini karena pada metode *cross correlation* yang digunakan dalam perhitungan ITD pada *center frequency* tersebut memiliki puncak yang lebih dari satu. Banyaknya puncak yang terjadi karena resonansi yang disebabkan oleh kanal telinga yang diibaratkan sebagai fungsi tabung terbuka dimana resonansi terjadi pada empat kali panjang kanal yang senilai dengan frekuensi 2000 Hz.

* + 1. **Estimasi *Probability Density***

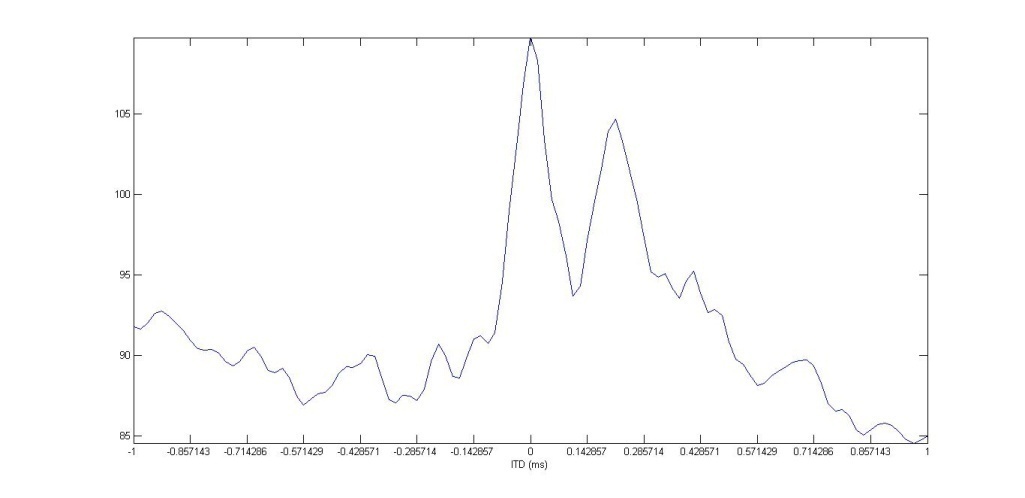
Hasil perhitungan nilai *probability density estimation* untuk kedua jenis *relative strength* ditampilkan dalam Gambar 4.15. Pada Gambar 4.15 (a), puncak tertinggi terjadi pada nilai ITD = 0 ms yang artinya nilai relative strength lebih besar dari 0,5 lebih banyak terdistribusi pada tersebut, sedangkan pada Gambar 4.15 (b) puncak tertinggi terjadi pada ITD=0.8 ms yang artinya nilai relative strength kurang dari/sama dengan 0,5 lebih banyak terdistribusi pada ITD tersebut. Garis berwarna hitam pada Gambar 4.15 (a) dan (b), yang merepresentasikan distribusi dari nilai ILD pada Target, lebih pendek dari pada ILD pada masker (garis berwarna merah). Hasil tersebut bersesuaian berkurangnya nilai *relative strength* jika posisi sumber suara menjauhi bidang median. Dari sini dapat disimpulkan, untuk estimasi *binary mask* yang membutuhkan nilai *relative strength* dapat diestimasi dengan hanya diketahui nilai ITD dan ILD saja pada sumber suara tercampur.

**Gambar 4.15** *Probability Density Estimation* dari ITD dan ILD (a) pada *relative strength* (R) >0,5 dan (b) *relative strength* (R)≤0.5 dengan C*enter Frequency* 500 Hz

******

* 1. **Proses Pemisahan**
     1. **Perhitungan Azimut**

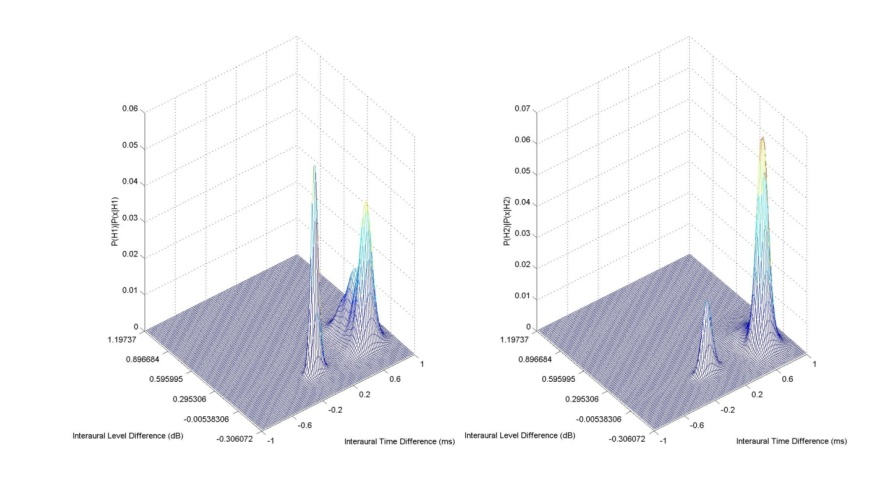
Azimut dari sinyal target dan masker dapat diestimasi dari menjumlahkan hasil *cross correlation* antara sinyal tercampur pada telinga kiri dan telinga kanan dalam domain waktu pada semua *center frequency* dari 80 Hz hingga 5000 Hz*.* Nilai puncak dari hasil penjumlahan *cross correlation* dalam fungsi ITD dapat direpresentasikan sebagai nilai azimut seperti pada Gambar 4.16. Pada gambar tersebut, puncak pertama berada pada *lag* 0 ms diartikan sebagai azimut target dan puncak kedua pada 0.21 ms sebagai azimut masker. Kedua nilai ini sebagai parameter pengambilan database *probability density estimation* dari hasil proses *data training.*



**Gambar 4.16** Hasil *Cross Correlation* Sinyal Suara Tercampur pada Telinga Kiri dan Telinga Kanan

* + 1. **Estimasi *Binary Masking***

Nilai *binary masking* diestimasi pada setiap unit *time-frequency* berdasarkan persamaan 3.4. *Binary Mask (BM)* bernilai 1 jika nilai dari *probability density estimation* dengan *relative strength* (R)*>*0.5(PDE R>0.5) pada Gambar 4.17 (a) lebih besar daripada *probability density estimation* dengan *relative strength* (R) *≤* 0.5 (PDE R≤0.5) pada Gambar 4.17 (b) dan akan bernilai 0 jika nilai dari *probability density estimation* dengan nilai *relative strength*(*R)>*0.5(PDE R>0.5) pada Gambar 4.17 (a) lebih kecil atau sama dengan PDE R≤0.5 pada Gambar 4.17 (b). Nilai dari masing-masing PDE diperoleh dari ITD dan ILD masker dalam setiap unit *time-frequency* dari sinyal suara tercampur. Pada Gambar 4.17 adalah grafik dari *probability density estimation* untuk target pada azimut 0o dan masker pada azimut 30o pada *center frequency* 500 Hz. ITD dan ILD masker yang diperoleh pada *center frequency* 500, dihitung nilai *binary mask-*nya berdasarkan nilai *probability density estimation* pada Gambar 4.17.



ITD

ILD

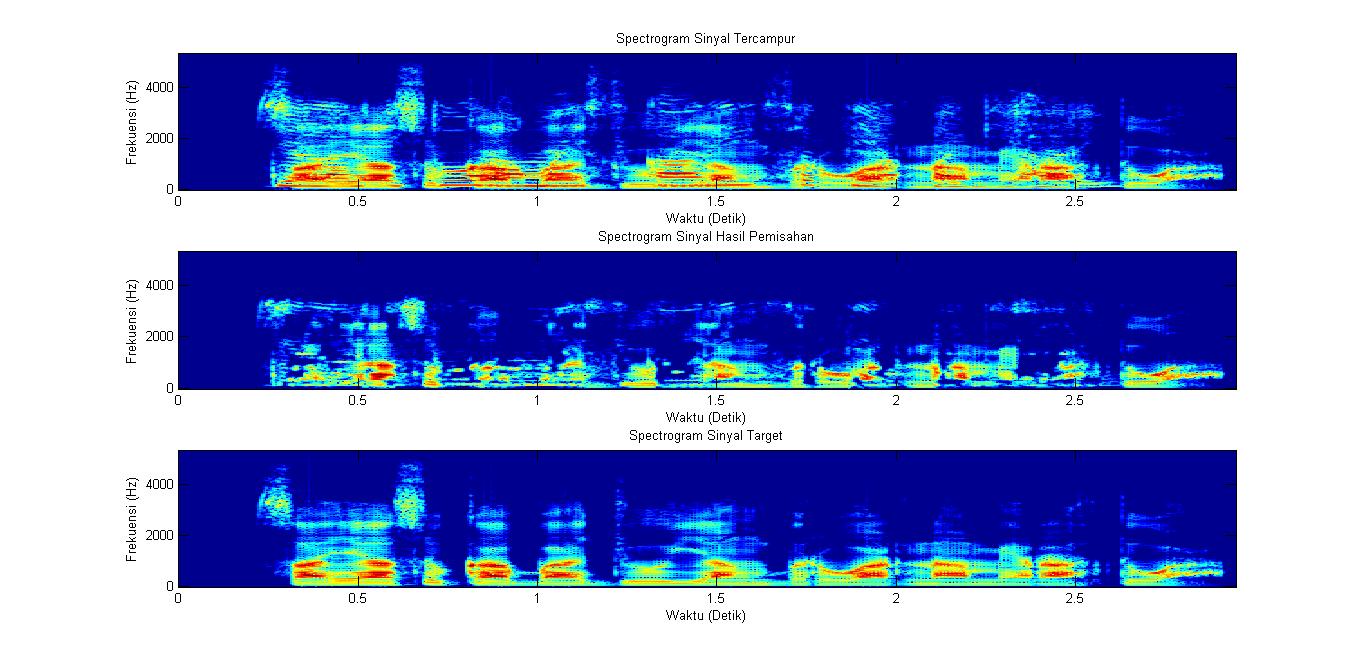
ILD

ITD

**Gambar 4.17** Pengambilan Nilai *Binary Mask* Berdasarkan Nilai ITD dan ILD *Masker* pada *Center Frequency* 500 pada (a) PDE *relative strength* (R)>0.5 dan (b) PDE *relative strength* (R)≤0.5

* + 1. **Hasil Pemisahan**

Dari *binary mask* yang diperoleh, kemudian dilakukan *filtering/weighting* terhadap sinyal tercampur pada Gambar 4.18 (a) menghasilkan sinyal hasil pemisahan pada Gambar 4.18 (b). Pada Gambar 4.18 yang dilingkari adalah daerah dari sinyal suara masker yang di *filter.* Sinyal suara target hasil pemisahan yang diperoleh terdapat *musical noise,* hal ini terjadi karena hasil pemisahan yang diperoleh terdapat informasi suara yang ikut ter-*masking* oleh *binary masking* yaitu pada rentang frekuensi diatas 3700 Hz (lingkaran berwarna hitam)*.* Hilangnya informasi pada frekuensi diatas 3700 Hz karena nilai ITD yang diperoleh tidak linier seperti pada frekuensi dibawah 3700 Hz. Ketidaklinieran ini terjadi karena terdapat banyak *peak* yang diperoleh pada saat perhitungan ITD yang disebabkan oleh resonansi yang terjadi pada kanal telinga. Resonansi ini disebabkan karena kanal telinga, yang memiliki panjang sebesar 2,3 cm, dimodelkan seperti tabung tertutup. Sehingga tabung tertutup tersebut akan mengalami resonansi pada panjang gelombang 9,2 cm atau frekuensi 3700 Hz.

**

b

c

a

**Gambar 4.18** (a) Spectrogram Sinyal Tercampur (b) Sinyal Target Hasil Pemisahan dan (c) Sinyal Target Sebelum Tercampur

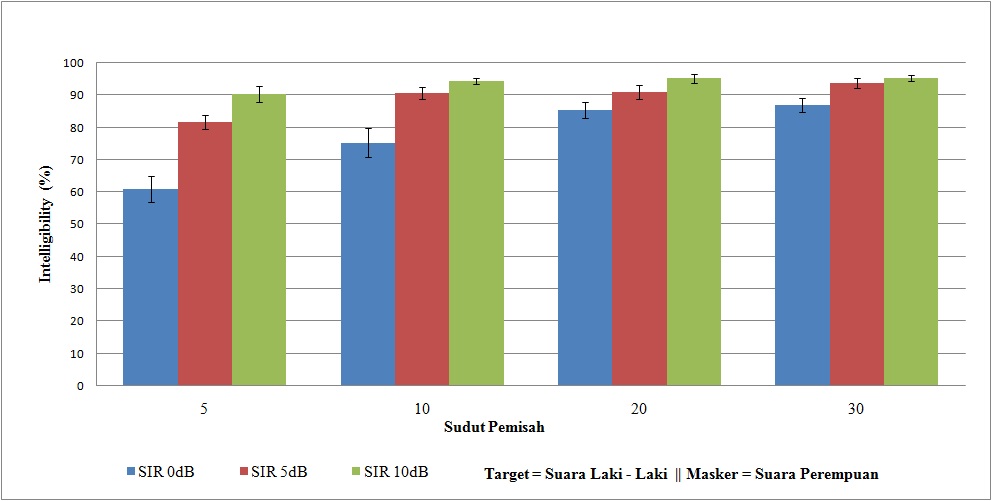
* + 1. **Hasil Evaluasi Subjektif dan Objektif**

Berdasarkan hasil evaluasi subjektif pada Gambar 4.19 dan Gambar 4.20, semakin besar sudut pemisah dan SIR antara sumber suara target dan masker maka nilai *intelligibility* dari suara hasil pemisahan semakin meningkat. Untuk gender yang berbeda antara sinyal suara target dan masker, memiliki nilai *intelligibility* yang lebih tinggi daripada sinyal suara target dan masker yang sama gender.

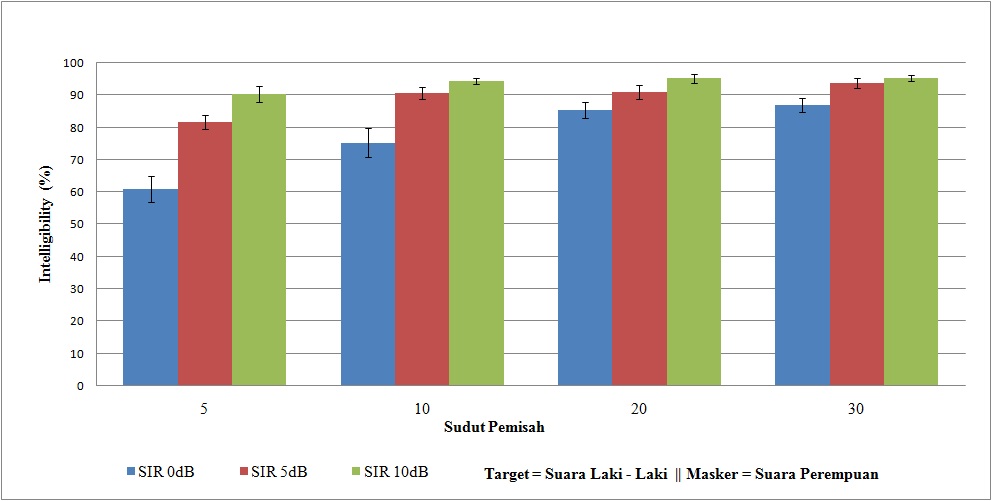
Sedangkan hasil evaluasi objektif pada Gambar 4.21, semakin besar sudut pemisah dan SIR antara sumber suara target dan masker maka nilai quality dari suara hasil pemisahan semakin meningkat. Untuk gender yang berbeda antara sinyal suara target dan masker, memiliki nilai quality yang lebih tinggi daripada sinyal suara target dan masker yang sama gender.

Pada Gambar 4.19, nilai peningkatan nilai *intelligibility* pada SIR 0 dB terlihat sangat signifikan yaitu 60% disudut pemisah 5o, 75% di sudut pemisah 10o, 85% di sudut pemisah 20o dan 86% di sudut pemisah 30o. Demikian pula pada SIR 5 dB yaitu 81,8% disudut pemisah 5o, 90.6% di sudut pemisah 10o, 91% di sudut pemisah 20o dan 93.7% di sudut pemisah 30o. Sedangkan untuk SIR 10 dB, peningkatan nilai *intelligibility* tidak terlalu signifikan yaitu 90,1% di sudut pemisah 5o, 94,4% di sudut pemisah 10o, 95,3% di sudut pemisah 20o dan 95,4% di sudut pemisah 30o. Padahal jika diperhatikan dengan selisih level yang terjadi antara telinga kiri dan telinga kanan pada Gambar 4.1 yang meningkat hingga dua kali lipat, nilai intelligibility pada sudut 30o seharusnya menunjukkan peningkatan yang signifikan juga.

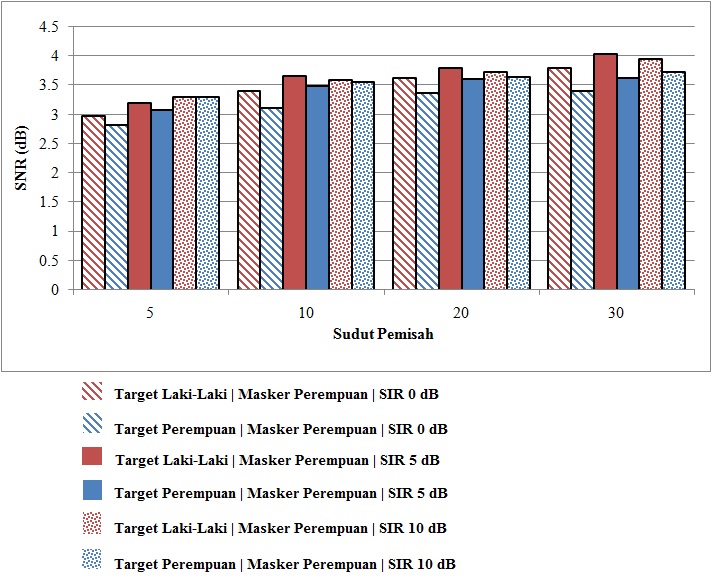
**Gambar 4.19** Hasil Evaluasi Subjektif *(percent Correct Word)* untuk Target Suara Laki-laki dan Masker Suara Perempuan



**Gambar 4.20** Hasil Evaluasi Subjektif *(percent Correct Word)* untuk Target dan Masker Suara Perempuan

**

**Gambar 4.21** Hasil Evaluasi Objektif (SNR)



***Halaman ini sengaja dikosongkan***