Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

Fast Head-Related Transfer Function Measurement in Complex Environments

Qinghua Ye, Qiujie Dong, Yan Zhang and Xiaodong Li

Key Laboratory of Noise and Vibration Research, Institute of Acoustics,
Chinese Academy of Sciences, Beijing 100190, China

PACS: 43.38.-p

ABSTRACT

Head-related transfer function (HRTF) describes a transfer function from the sound source to the listener's ears and plays a central role in binaural spatial and virtual hearing studies. Measuring HRTF requires rigorous experimental conditions and specially designed equipments, and the procedure becomes very time consuming and tiring for the participants. In this paper a fast HRTF measurement method is presented. By multi-point simultaneous measurement using a loudspeaker array, rigorous acoustical conditions and special equipments are not required and the needed HRTFs of a subject are rapidly measured as well as its head and position information. Quality of the measured HRTF is also evaluated. Experiments in an ordinary room demonstrated its effectiveness.

INTRODUCTION

Head-related transfer function (HRTF) describes a transfer function from the sound source to the listener's ears in the frequency domain and plays a central role in spatial and virtual hearing studies [1]. It relates closely to listener's physiological structures and sizes thus obviously differs to different ones. In virtual auditory display, the result is best with individual HRTFs built for specified listeners while results using other HRTFs depend on how similar they are to the listener's HRTF. To measure HRTF is not only time consuming but also rigorous to the conditions and equipments, e.g., needs anechoic chamber, head position tracking machine, swiveled chair, and body-fixed equipment. Some groups have made such measurements and built up databases for different locations in the space, including MIT KEMAR database [2] and CIPIC HRTF database [3]. In 2005, Xie's group built up a database including measurements of HRTFs and physiological structure sizes to 52 Chinese people [4]. Our cooperation with Peking University finished measurements and set up a near-field HRTF database using a KEMAR dummy head in 2007 [5].

The most common method in individualized HRTF research is to measure anthropometric parameters of the subjects, then estimate and build up the corresponding HRTF by approximation. This method further diverges to be individualized HRTFs by anthropometric parameters match [6], by scaling in frequency [7], by linear multiple regression approximation [8], by structural model of HRTF [9] and by subjective assessment [10] etc. Drawbacks of these methods include their lack of precision and need for HRTF database with large number of subjects and anthropometric parameters, which is difficult to obtain in research and applications.

An interesting question is, whether we can obtain our individual HRTFs "online" in an ordinary room with our own audio equipment? We did some research on this issue. Different from the general HRTF measuring techniques which pursue "clean" environment in order to get very high signal-to-noise ratio (SNR), this paper concentrates on the measurement done with high level of noise as well as disturbance and tries to find out subjects with high tolerance to the environment. This paper aims at the fast HRTF measurement in complex environments and listener's head parameter and its movement parameter as well. Rigorous acoustical conditions and special equipments are not required and the needed HRTFs of a subject are rapidly measured and quality evaluated. This method alleviates the conflict mentioned above, and provides a good supplement to the existing individualized HRTF methods.

HRTF MEASUREMENT USING LOUDSPEAKER ARRAY

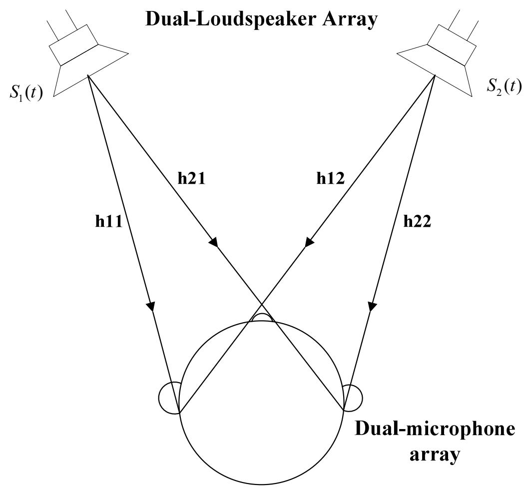

Figure 1. HRTF measurement using a dual-loudspeaker array

The loudspeaker array contains two or more loudspeakers. The uncorrelated test signals are magnified by the power amplifier and then emitted from the loudspeakers simultaneously. To illustrate the measurement process, we take a dual-loudspeaker array for example.

Fig. 1 illustrates the measurement system containing two independent sound sources and two independent sound receivers. It is assumed that the system is linear time-invariant (LTI) within the test time.

As shown in Fig. 1, a pair of test sequence $s_1(t)$ and $s_2(t)$ is used as the left and right output signal respectively, and the microphone pair placed at the entrance to the ear canal receives the signals. We use h_{ij}, $i, j = 1, 2$ to indicate the head-related impulse response (HRIR) between the *i*-th microphone input and the *j*-th loudspeaker output. According to the multi-input system theory, the *i*-th channel signal received can be expressed as:

$$y_i(t) = \sum_{j=1}^{2} s_j(t) * h_{ij}(t) \tag{1}$$

In Eq. (1), symbol $*$ indicates linear convolution. Thus, the cross-correlation function between the *j*-th sound source and the *i*-th received signal can be expressed as:

$$s_j(t) \otimes y_i(t) = \sum_{k=1}^{2} \left[s_j(t) \otimes s_k(t) \right] * h_{ik}(t) \tag{2}$$

In Eq. (2), symbol $\otimes$ indicates the linear correlation operator. If $R_{jk}(t)$ indicates the cross-correlation function of the *j*-th and the *k*-th sound source signal, and $R_{jk}(t)$ almost satisfies the relationship showed below:

$$R_{jk}(t) = s_j(t) \otimes s_k(t) = \begin{cases} \delta(t) & j = k \\ 0 & j \neq k \end{cases} \tag{3}$$

take Eq. (3) into Eq. (2), thus we get Eq. (4) below:

$$h_{ij}(t) = s_j(t) \otimes y_i(t) \tag{4}$$

For example, if $i = 1, j = 2$, then

$$h_{12}(t) = s_2(t) \otimes y_1(t) \tag{5}$$

In this way the HRTFs of the loudspeakers to each ear can be obtained simultaneously.

Different kinds of testing signals meet the needs of HRTF measurement, and the maximum-length-sequence (MLS) is used in this paper. MLS is easily implemented and its autocorrelation function is a very close approximation to a delta function. The response of repeated MLS can be computed by the synchronous averaging technique. As a result of the randomness of background noise, the noise energy can be reduced and the SNR be raised by using multi synchronous averaging. When MLS repeats a double time, the SNR could be raised by 3dB, which is benefit for high-noise environment measurement.

Some proper fast algorithms can be employed to reduce the computational burden of Eq. (4). For instance, a fast two channel algorithm using reciprocal MLS pairs [11] could be used in the above dual-loudspeaker HRTF measurement case.

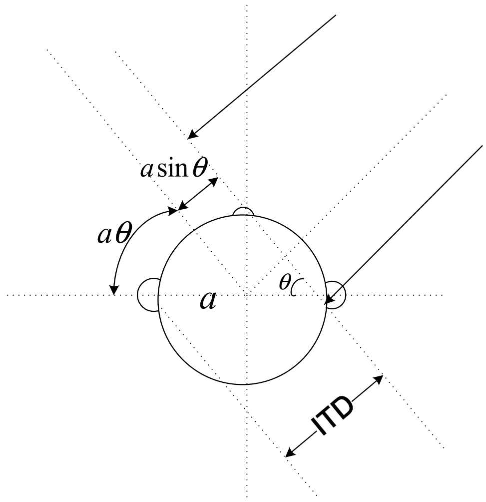

Figure 2. ITD calculation on horizontal plane using a sphere head model

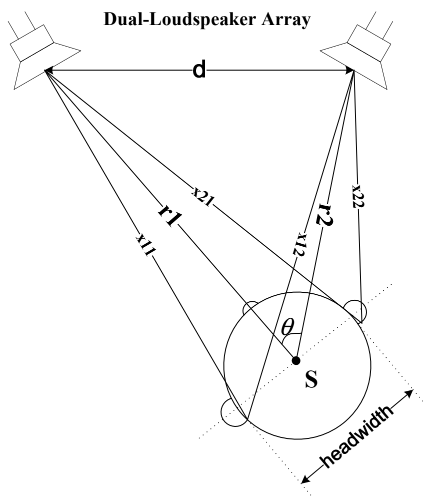

Figure 3. The diagram of head and dual-loudspeaker array geometry

HEAD SIZE AND MOVEMENT ESTIMATION

Fig. 2 is a sketch drawing of head model and incidence sound wave. Considering the effect of head, we make an approximation that head is a sphere with radius a while ears are two points on the surface of the sphere. The sound wave with elevation angle 0° and azimuth angle θ transmits on the bending surface of head, and the interaural time difference (ITD) could be [12]:

$$ITD(\theta) = \frac{a}{c}(\sin\theta + \theta), \quad \text{when} \quad 0 \le \theta \le \frac{\pi}{2}. \tag{6}$$

where *c* indicates sound velocity. If $ITD > 0$, then right ear is positional ahead. Otherwise, left ear is positional ahead.

The mutual position of listener's head and the dual-loudspeaker array when measuring listener's head radius is shown in Fig. 3. A couple of loudspeakers with measured distance *d* are fixed in front of the listener's head on the same horizontal plane. We define the distance between the left loudspeaker to the left and right ear as x_{11} and x_{21} respectively, meanwhile the right loudspeaker to the two ears as x_{12} and x_{22}. The distance between the two loudspeakers and head center is r_1 and r_2 respectively. Considering that head radius is much smaller than the distance between loudspeakers to the head, the following approximation can be made:

$$r_1 \approx \frac{1}{2}(x_{11} + x_{21}) \qquad (7)$$
$$r_2 \approx \frac{1}{2}(x_{12} + x_{22})$$

The $x_{11}, x_{21}, x_{12}, x_{22}$ is computed using the time difference Δt between the signal transmitting and receiving, e.g., $x_{11} = c\Delta t_{11}$.

Consider the situation that the subject is facing one of the loudspeakers, e.g., the left one in Fig. 3. Since the median plane of listener's head coincides with the sound wave direction from the left loudspeaker, the measurement of ITD_1 should give zero. A real-time measurement system can notify the listener to adjust his/her position when ITD_1 is nonzero. The angle between the median plane of listener's head and right loudspeaker is defined as θ. In the triangular of head center S and both loudspeakers, a geometry relationship could be:

$$\theta = \arccos\left(\frac{r_1^2 + r_2^2 - d^2}{2r_1r_2}\right) \qquad (8)$$

Together with the measurement of ITD_2, the head radius can be expressed as:

$$a = \frac{c \times ITD_2}{\sin(\theta) + \theta} \qquad (9)$$

With the head radius, listener's head movement parameter, in another word the azimuth angle α between loudspeaker and face median plane can be derived from $ITD(\alpha) = \frac{a}{c}(\sin\alpha + \alpha)$ as

$$\alpha = g^{-1}\left(\frac{c}{a} \times ITD(\alpha)\right) \qquad (10)$$

where g^{-1} is the inverse function of $g(\alpha) = \sin\alpha + \alpha$. Using Chebyshev sequences to expand g^{-1} to be:

$$\alpha = \tilde{g}^{-1}(x) = \frac{x}{2} + \frac{x^3}{96} + \frac{x^5}{1280} \qquad (11)$$

where $x = \frac{c}{a} \times ITD(\alpha)$. This approximation can significantly speed up the calculation and enhance the efficiency.

The ITD and Δt can be estimated by the measured HRIRs when the head position is fixed. If its position is changing slowly, real-time tracking can be done by application of short-time fragments of MLS sequences under the assumption that loudspeakers continuously broadcast repeated signals.

WEAK SIGNAL MEASUREMENT

In the situation of room with multi reflecting surfaces, HRTF could be obtained through applying a suitable time window function to the measured impulse response, therefore this section focuses on the HRTF measurement with low SNR. The measurement is done with low SNR if the environment noise is strong or the signal is very weak. Pseudo-random sequence such as MLS can improve SNR significantly after several times of repeating. Meanwhile a convenient and instant evaluation of the quality of measured HRTF is necessary.

We use the coherence function of the input and output signal to evaluate the HRTF quality in different frequency ranges. The coherence function is defined in terms of power spectral densities and the cross-spectral density by [13]

$$C_{xy}(\omega) \triangleq \frac{|R_{xy}(\omega)|^2}{R_x(\omega)R_y(\omega)} \qquad (12)$$

In practice, an estimate of the coherence, the sample coherence function $\widehat{C}_{xy}(\omega_k)$, may be estimated by time-averaging $\overline{X}(\omega_k)Y(\omega_k)$, $|X(\omega_k)|^2$ and $|Y(\omega_k)|^2$ over successive signal blocks [13]:

$$\widehat{C}_{xy}(\omega_k) \triangleq \frac{\left|\left\{\overline{X}_m(\omega_k)Y_m(\omega_k)\right\}_m\right|^2}{\left\{\left|X_m(\omega_k)\right|^2\right\}_m \cdot \left\{\left|Y_m(\omega_k)\right|^2\right\}_m} \qquad (13)$$

where $\{\cdot\}_m$ denotes time averaging across frames.

If y is produced from x via a LTI filtering operation, the coherence becomes 1. On the other hand, when x and y are uncorrelated (*e.g.*, y is a noise process not derived from x), the sample coherence converges to zero at all frequencies.

To fully benefit from the improvement of SNR by applying MLS and to minimize the size of data to deal with, a new output sequence is reorganized from the serial of measured HRIRs:

$$y'(n) = \left[HRIR_1, HRIR_2, \cdots, HRIR_q, \cdots, HRIR_Q\right] \qquad (14)$$

where $q = 1, 2, \cdots, Q$, Q is the repeated times of MLS in a single HRTF measurement, and $HRIR_q$ indicates the windowed HRIR result accordance to the q-th MLS. Accordingly, the input sequence is the alignment with Q delta functions:

$$x'(n) = \left[\delta_1, \delta_2, \cdots, \delta_q, \cdots, \delta_Q\right] \qquad (15)$$

Ideally, the coherence should be 1 at all frequencies. However, if the microphones record mostly noise at some frequency range, then it is indicated in the measured coherence by a significant dip below 1 at that frequency. If this dip is corresponding to the one in HRTF, the measured result within this frequency range is still credible. On the other hand, the coherence function could also be close to 1 within some frequency range if there is an intense noise source in this frequency range. Then if the noise source is not nearby to loudspeaker's direction through observing the measured HRTF phase spectrum, the results within this frequency range have to be abandoned.

EXPERIMENTAL RESULTS

The measurement experiment was taken in an ordinary room. A dual-loudspeakers array was used as the sound source and a B&K dummy head Type 4100 with dual-microphones was placed any site on the same horizontal plane. Two MLS sequences, one of which is in reverse arrangement (time rever-

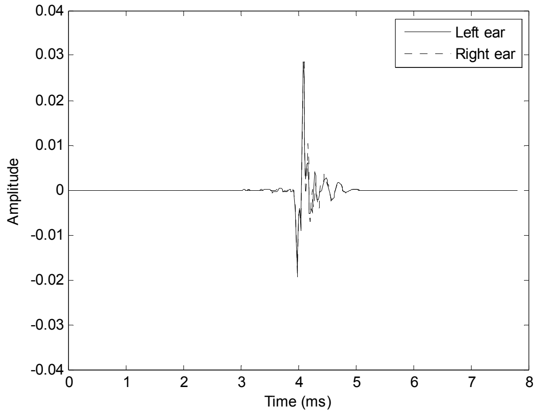

Figure 4. HRIR of left ear (solid line) and right ear (dotted line) for the right loudspeaker

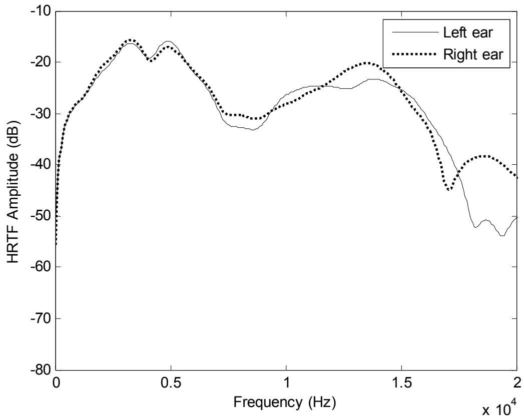

Figure 5. HRTF amplitude of left ear (solid line) and right ear (dotted line) for the right loudspeaker

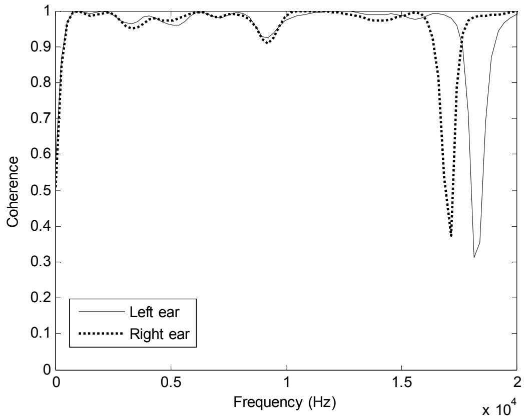

Figure 6. Coherence function curve of left ear HRIR (solid line) and right ear HRIR (dotted line) for the right loudspeaker

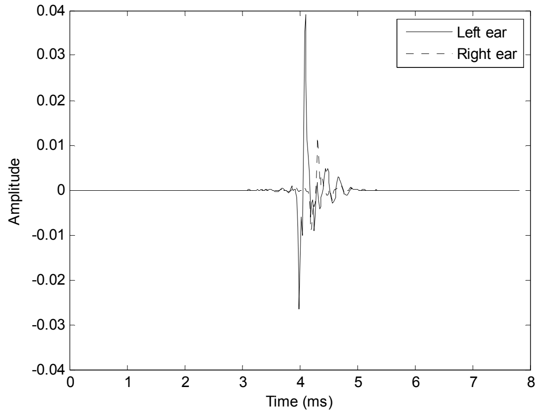

Figure 7. HRIR of left ear (solid line) and right ear (dotted line) for the left loudspeaker

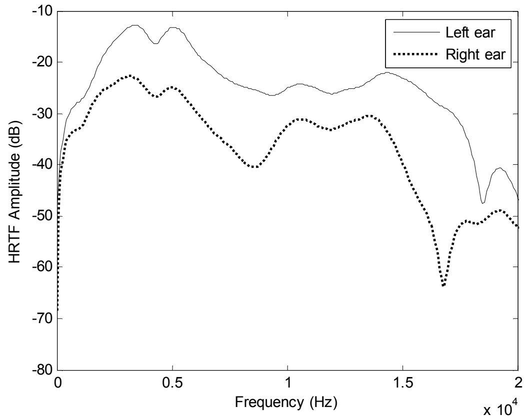

Figure 8. HRTF amplitude of left ear (solid line) and right ear (dotted line) for the left loudspeaker

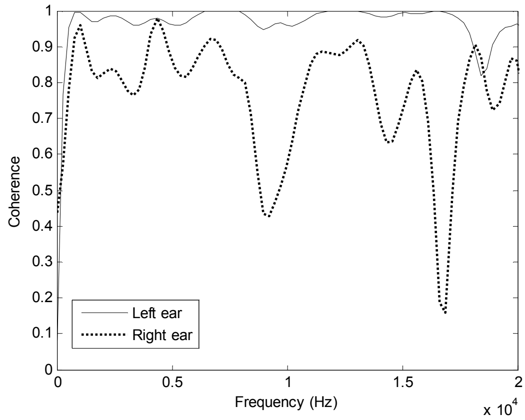

Figure 9. Coherence function curve of left ear HRIR (solid line) and right ear HRIR (dotted line) for the left loudspeaker

sal) of the other, were used as the loudspeakers signal. Figs. 4-6 above is an example illustrating the HRIR, HRTF and the coherence function of the right loudspeaker and each ear, while Figs. 7-9 illustrating the features of left loudspeaker. The HRIR or HRTF here is not compensated, which is to say, it contains the electro-acoustic system response. According to this measurement, the head is facing the right loudspeaker, as a result, the head radius could be easily obtained which is 7.95cm.

As seen in Fig. 9, on the coherence function illustration there are two significant dips below 1 at frequency range 800-1k and 1.6k-1.8k. However, the two dips totally correspond to the HRTF dips at the same frequency range, that's to say, the data is credible.

In next measurements, the dummy head was placed at five different sites and angles. It is defined that the distance between each loudspeaker and the head center is r_1 and r_2,

Table 1. Head localization results at different positions

		r1 (m)	α_1 (deg)	r2 (m)	α_2 (deg)
Positon 1	Tracker	0.9994	28.7648	1.0274	1.74
	Our Result	1.0091	25.202	1.0168	0
Position 2	Tracker	0.9935	13.2579	1.0085	13.2579
	Our Result	0.9811	11.9814	1.004	12.9107
Position 3	Tracker	1.0445	10.0732	1.0392	16.0601
	Our Result	1.0428	14.0929	1.0428	15.0431
Position 4	Tracker	0.9377	15.99	0.9962	12.2502
	Our Result	0.9442	14.0929	1.0013	13.1446
Position 5	Tracker	1.2757	9.0918	1.1507	12.0589
	Our Result	1.2737	6.5506	1.1569	14.0929

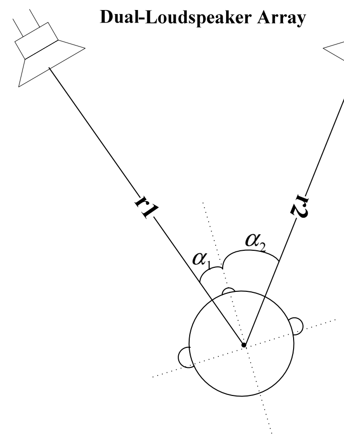

Figure 10. Diagram of HRTF measurement and head localization

meanwhile, the angle between the head median plane and each loudspeaker is α_1 and α_2, which is shown in Fig. 10. The distance and angle results are computed by the acoustics method discussed in Sec. 3. To test the accuracy, we equip three 6DOF electromagnetic motion trackers (Polhemus FASTRAK system) on the head and loudspeakers to record the head moving posture. Data is shown in Table 1. The electro-acoustic delay has been compensated in the results.

As shown in Table 1, the distance error is within 1cm while the angle error is almost within 2°, which could nearly satisfy the ordinary measurement. The localization accuracy could be enhanced if more precise head model is used in Eq. (6).

CONCLUSIONS

This paper concentrates on the measurement done with high level of noise as well as disturbance and tries to find out subjects with high tolerance to the environment. A method using loudspeaker array to measure the HRTF and obtain head width and its position parameters is presented. This study is benefit for the building of large sample HRTF databases, meanwhile, it offers preliminary theoretical and experimental results to the "online" HRTF measurement. Our future work aims at the active signal design to explore the combined processing with short and long sequences, to obtain real time HRTF simultaneously with precise estimation of head and position parameters. The prediction or compensation method for the unreliable data is our next work as well.

REFERENCES

1 J. Blauert, *Spatial Hearing* (Revised edition, MIT Press, Cambridge, MA, England, 1997)
2 http://sound.media.mit.edu/resources/KEMAR.html
3 http://interface.cipic.ucdavis.edu/CIL_html/CIL_HRTF_database.htm
4 B. Xie, X. Zhong, D. Rao, et al, "Head-related transfer function database and analyses" *Science in China Series G, Physics, Mechanics & Astronomy*. **50**(3), 267-280 (2007)
5 T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li and X. Wu, "Distance-Dependent Head-Related Transfer Functions Measured With High Spatial Resolution Using a Spark Gap" *IEEE Trans. On Audio, Speech, and Language Processing*. **17**(6), 1124-1132 (2009)
6 D.N. Zotkin, R. Duraiswami and L.S. Davis, "Rendering localized spatial audio in a virtual auditory space" *IEEE Trans. On Multimedia*. **6**(4), 553-564 (2004)
7 J.C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency" *J. Acoust. Soc. Am.* **106**(3), 1493-1510 (1999)
8 T. Nishino, N. Inoue, K. Takeda, et al, "Estimation of HRTFs on the horizontal plane using physical features" *Applied Acoustics*. **68**(8), 897-908 (2007)
9 C.P. Brown and R.O. Duda, "A structural model for binaural sound synthesis" *IEEE Trans. On Speech and Audio Processing*. **6**(5), 476-488 (1998)
10 U.S. Bernhard and F. Hugo, "Subjective selection of non-individual head-related transfer functions" *Proceedings of the 2003 International Conference on Auditory Display*, Boston, MA, USA. 259-262 (2003)
11 N. Xiang and M.R. Schroeder, "Reciprocal maximum-length sequence pairs for acoustical dual source measurements" *J. Acoust. Soc. Am.* **113**(5), 2754-2761 (2003)
12 R.S. Woodworth and H. Schlosberg, *Experimental Physchology* (Holt H. (Ed.), New York, USA, 1954)
13 Julius O. Smith III. *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications* (2nd Ed., W3K Publishing, 2007)