37

Source Summation in the Vertical Plane

by

John A. Crouch

Submitted to the Department of Electrical Engineering and Computer Science

in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Electrical Engineering and Computer Science

and Master of Engineering in Electrical Engineering and Computer Science

at the Massachusetts Institute of Technology

December 20, 1996

Author____

Department of Electrical Engineering and Computer Science

December 20, 1996

Certified by____

Nat Durlach

Thesis Supervisor

Accepted by____

F. R. Morgenthaler

Chairman, Department Committee on Graduate Theses

Source Summation in the Vertical Plane

by John A. Crouch

Submitted to the

Department of Electrical Engineering and Computer Science

December 20, 1996

in Partial Fulfillment of the Requirements for the Degrees of

Bachelor of Science in Electrical Engineering and Computer Science

and Master of Engineering in Electrical Engineering and Computer Science

ABSTRACT

Sound localization in the vertical plane is thought to depend on spectral cues induced by the head-related transfer function (HRTF). The relative importance of different possible cues to vertical localization was investigated in this study. In one experiment, subjects were presented with simultaneous sounds from two speakers located at 0° and 60°. By controlling the relative amplitudes of these sounds, they matched the "summed sound" to the position of a target sound, located at 30°. In a second experiment, the two "summed" speakers were set to equal levels and subjects reported the apparent vertical locations of the sound. Head-Related Transfer Functions (HRTFs) were then measured in the median plane for each subject. HRTFs were analyzed to determine what cues dominate vertical localization. Generally, interaural phase cues did not predict results well; however, for 4 of 5 subjects, results were approximated by the interaural difference cue. For all subjects, reasonable predictions were made by examining the location of the first large spectral notch. These results imply that the spectral notch is the most consistent cue for median plane localizations, but that interaural difference cues may be used by some subjects.

Thesis Supervisor: Nat Durlach

Title: Research Scientist, MIT Research Laboratory of Electronics

Acknowledgements

I would like to acknowledge the guidance and support of Barbara Shinn-Cunningham and Abhijit Kulkarni. Both of them have made great efforts to improve my experience with this thesis and my education.

Table of Contents

1. Background

Research in sound localization has surged with the recent interest in producing virtual environment systems which immerse the user in a symbolic environment by using graphics and sound. These environments can represent real landscapes through which a user might navigate a robot or they can represent imaginary scenes through which a pilot might fly a simulated airplane. The need for three dimensional sound to accurately simulate the acoustics of these environments has spurred further investigation into the mechanisms of sound perception.

Past research has established three cues for localizing sound: interaural time delay, interaural intensity differences, and frequency (or spectral) cues. People's ability to localize sounds in the horizontal plane have been attributed primarily to interaural disparities. A sound source located next to the left ear will reach the left ear before the right ear thus resulting in an interaural time delay. That same sound wave will be composed of low frequencies that will diffract around the head and high frequencies that will be partially reflected. As a result, there will be an interaural intensity difference in high frequencies. Interaural time and intensity differences are not present with sources in the median plane; instead, spectral cues induced by the pinnae are thought to be the main cue in determining perceived elevation (Blauert (1983)).

The specific mechanism of median plane localization is thought to depend on the peaks and troughs in the perceived sound spectrum that result from interactions of the direct sound with the folds and structures of the pinnae. Previous research has quantified the temporal and spectral effects of the pinnae and has illustrated that frequency characteristics of the sound received at the ear drum change with the position of a sound source in the vertical plane (Watkins (1978)). People use the frequencies of peaks and troughs in the spectra that change with vertical source position to determine sound elevation.

Watkins suggested a model of the vertical plane localization process in which people localize sounds by their spectra. To test this model, the response of the external ear was synthesized with a computer program in which the direct sound source was delayed, scaled, and summed. This process produces notches and peaks in the received spectra similar to those produced by the pinnae. Processed sounds were then played to subjects through tubes inserted into the ear canal, thus bypassing the real spectral effects of the pinnae. Watkins discovered that the perceived elevation could be predictably altered by varying the delay and scaling factor used in the creation of the stimuli.

Watkins then showed that the perceived location of a source sound could be predicted by correlating the source spectrum with the spectra of sounds from different vertical locations. The source position whose received spectrum was most highly correlated with the synthesized source was a good predictor of perceived location.

Several other experiments have illustrated the significance of frequency "notches" in vertical plane localization but do not conclude that the cue solely determines perceived elevation. Butler *et. al* found that spectral cues determined perceived vertical location in monaural conditions. However, perception of spectral notches did not explain the better performance of subjects in binaural as opposed to monaural localization in the vertical plane 15° and higher. The study suggests that small spectral differences affect localization in that region of the vertical plane. These results indicate that studies performed in the vertical plane using sounds from below and above 15° should include binaural cues to ensure an accurate simulation.

Another related experiment illustrates the universality of spectral cues in the vertical plane. Using nonindividualized head-related transfer functions (HRTFs), Wenzel *et al.* were able to accurately control perceived vertical locations in 12 of 16 subjects. Since the same set of HRTFs were used for all subjects and still produced changes in vertical localization, spectral changes must be somewhat universal for subjects.

Although researchers have found predictable relationships between spectral cues and perceived location for single-source sounds in the median plane, it is not known what occurs perceptually when two sources are presented simultaneously. Matching the notches in the received sound spectrum with notches in HRTF's can predict perceived elevation for a single source (Watkins). The current study examines the predictability of perceived elevation for simultaneous sources based on spectral cues by comparing the total received spectrum for two simultaneous sources and the received spectrum of a target source. In addition, binaural level and phase difference spectra were examined.

2. Experimental Methods

Two experiments were performed on five subjects. Each experiment involved a "target" speaker located in between two "summed" speakers. The target speaker, located at 30°, played a constant sound pressure level when switched on. The summed speakers, located at 0° and 60°, each played the same noise source with weighted amplitudes that were either under subject control (in Experiment 1) or equal (in Experiment 2); however, the total sound energy emitted by both of these speakers remained constant. The amplitudes were controlled by the following formula:

$$(A_0)^2 + (A_{60})^2 = 1000^2 \tag{1}$$

where A_0 is the gain of the speaker at 0° and A_{60} is the gain of the speaker at 60°. Note that the total range of gains ran from 0 to 1000.

The two experiments were conducted in an anechoic chamber. Each listener sat in a chair surrounded by a ring located in the median plane (See Figure 1 below). Speakers were attached to this ring so that all of the speakers were equidistant from the listeners head. Speaker positions were identical for both experiments. The target speaker (used in Experiment 1) was positioned at 30° while the speakers used for the summed sound were located at 0^0 and 60°.

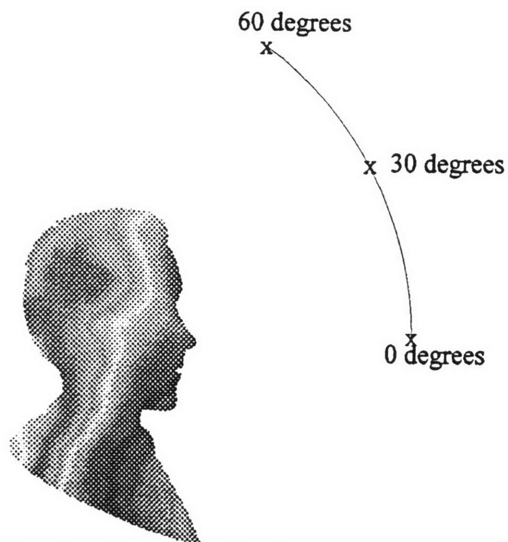


Figure 1: Speaker locations in the median plane

Sound bursts in the experiments originated from a white noise source. The noise was produced by a General Radio 1382 Random Noise Generator. Noise was amplified by a Crown D150 Series II amplifier. White noise was used to ensure that the high frequency characteristics that cue median plane localization were present. Source outputs were under computer control. The output was switched on and off digitally by the computer to produce 200 ms noise bursts. Source gains were set digitally before being presented through the speakers. All of the speakers played from the same source.

In formulating the experiment, incoherent noise was tested as a source from the summed speakers. The resulting summed sound did not have a distinct location. The sound was described as "diffuse" as if the sound were produced by a source that was a foot in diameter. In contrast, summed sounds using coherent noise had an identifiable location that spanned a few degrees at most. Although the perceived location of the summed sound could be matched to the target the summed sound differed subjectively from the target. This differentiation allowed the subjects to distinguish between the target and the summed sound.

2.1 Experiment 1

In the first experiment, listeners matched the perceived location of a target sound by controlling A_0 and A_{60} (the gains of the 0° and 60° speakers). In the second experiment, each listener identified the perceived location of a summed sound when the two summed sources were presented at equal amplitudes. The first experiment consisted of twenty trials. In each trial, sound was alternately played through the target and summed speakers. Initially, the 60° speaker amplitude was 1000 (the maximum) and the 0° speaker amplitude was 0.

Sound was first played through the target speaker and then the summed speakers. The listener then pressed buttons on a hand-held device to indicate whether the perceived sound was located above or below the target. This process continued until both sound locations matched. The subject then indicated satisfaction by pressing an additional button. At the end of each trial, the weights assigned to the summed sound speakers by the subjects were recorded.

Each subject matched the apparent target location twenty times in one experimental session. There were eleven possible combinations of amplitudes. The amplitude of the 0° speaker was determined by the formula:

$$A_o = (1000)\,(x/10)^{0.5} \tag{2}$$

where x ranged from 0 to 10. Equation (2) ensures that the amplitudes are weighted to the extremes (1000) at positions 0 and 10 and are equally weighted (707) at position 5. Equations (1) and (2) constrained the amplitudes A_0 and A_{60} as follows:

A_{60}	1000	948	894	836	774	707	632	547	447	316	0
A_0	0	316	447	547	632	707	774	836	894	948	1000

2.2 Experiment 2

At the end of the first experimental session, each listener was presented a summed sound in which the two summed speakers amplitudes were equal ($A_0 = A_{60} = 707$). The listener then verbally reported the perceived angular elevation of the sound in degrees.

2.3 HRTF Measurements

In a second experimental session, Head Related Transfer Functions (HRTFs) were recorded for each subject every five degrees from 0° (directly in front of the head) to 180° (directly behind the head).

3. Results

3.1 Experiment 1

The results of experiment 1 are given in Table 1 and illustrated in Figure 2 (note that the maximum possible gain is 1000). The standard deviations show that results were consistent for each subject; however, differences in the mean gain suggest significant differences among subjects.

Subject	SS	JC	EP	EW	CM
Mean gain (A_0)	702	633	733	355	365
Standard deviation	111	67	87	62	82

Table 1: Results of experiment 1. Mean gain of 0° speaker and standard deviation when perceived location equaled target speaker location. Maximum speaker gain is 1000.

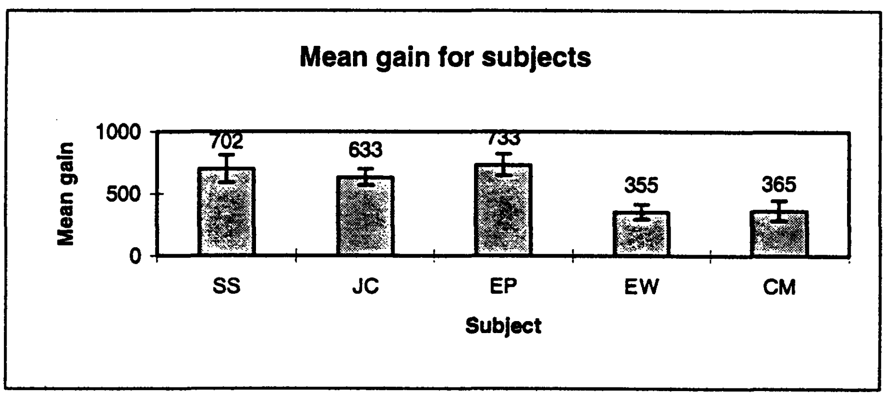


Figure 2: Mean gain and standard deviation for each subject.

3.2 Experiment 2

The results for experiment 2 are illustrated in Figure 3. In general, subjects perceived the source from a location in between the locations of the summed speakers. However, subject EP experienced front-back confusion, hearing the summed source as coming from behind the head.

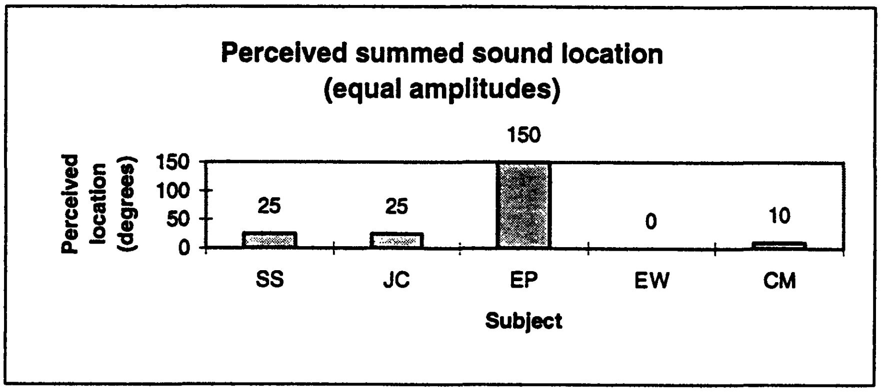


Figure 3: Perceived summed sound location with $A_0 = A_{60} = 707$.

3.3 HRTF Measurements

HRTFs recorded at 0°, 30°,and 60° for the five subjects are shown in appendices 1 and 2.

4. Data Analysis

The effective spectrum is a calculation of the total spectrum that reaches the ear canal in the summed speaker conditions. The following formula determined the effective spectrum (ES):

$$ES = (A_0/1000)^{0.5} (HRTF_0) + (A_{60}/1000)^{0.5} (HRTF_{60}) \qquad (3)$$

where A_0 and A_{60} are the weights assigned to the speakers at 0° and 60° respectively and $HRTF_0$ and $HRTF_{60}$ are the HRTFs from 0° and 60° respectively. The effective interaural difference spectra were calculated by the formula:

$$L/R\ ES = ES\ (left)\ /\ ES\ (right) \qquad (4)$$

4.1 Experiment 1

Using four methods, effective spectra were compared to recorded HRTFs. Each analysis method produced expectations of the gain A_0. All of the methods compared a large number of computed effective spectra to one HRTF at 30°. In the first method, 1000 effective spectra were computed for gain term A_0 from 0 to 1000. The magnitude of each of these spectra was then correlated to the recorded 30° monaural HRTF (for this purpose, only the HRTF spectrum reaching the left ear was considered). The gain that produced the maximum correlation was labeled the predicted gain (A). This method found the best fit for the overall monaural spectral magnitude shape.

The second method used the same process as the first method except that 1000 effective interaural level difference spectra were computed and compared to the recorded interaural level difference HRTF at 30°. The gain A_0 that corresponded to the maximum correlation was labeled the predicted gain (B). This method matched the overall interaural difference spectral magnitude shape.

In the third method, the same analysis was performed on the phase of the interaural difference spectra. One thousand effective interaural phase difference spectra were computed and correlated to the recorded interaural phase difference HRTF at 30°. The maximum correlation corresponded to the predicted gain (C). The resulting effective interaural phase difference spectrum was the best match to the overall shape of the interaural phase.

The final method involved matching the spectral notches present in the magnitude of the HRTFs. Notches were visually apparent in the magnitude plots of monaural HRTFs (see appendix 2). Fitting the overall shape of the HRTF magnitude as in the first method does not match the notches in the spectra. Matching the notches of the effective spectra and the recorded 30° HRTFs first required visually determining the frequency of the notch, Fnotch, in the recorded 30° HRTF. Then, 1000 monaural effective spectra were computed with A_0 varying from 0 to 1000. The gain that produced the largest notch in the vicinity of Fnotch was the predicted gain (D).

Table 3 below shows the predicted gains using the four analysis methods. Gain (A) denotes the gain produced by matching monaural level spectra, gain (B) interaural level spectra, gain (C) interaural phase difference spectra, and gain (D) from the notch method. Notch matching (D) was the only method that consistently approximated the gain term chosen by the subjects. Correlating monaural magnitude spectra (A) approximated the mean gains for subjects SS and EP. Interaural level difference spectra (B) predicted the gains for EP and CM. Although interaural phase difference spectra (C) accurately predicted the gain for SS, there was significant error for the other subjects.

Figure 4 below illustrates the predictions of each method. The effective spectra that result from the predicted gains are shown in Appendix 1.

Subject	SS	JC	EP	EW	CM
A_0 chosen in Exp #1	702	633	733	355	365
Predicted gain (A)	636	1000	510	1	769
Predicted gain (B)	1000	1000	826	1	461
Predicted gain (C)	702	1	466	94	1000
Predicted gain (D)	640	500	610	0	100

Table 2: Gains, A_0, predicted by various methods of analysis.

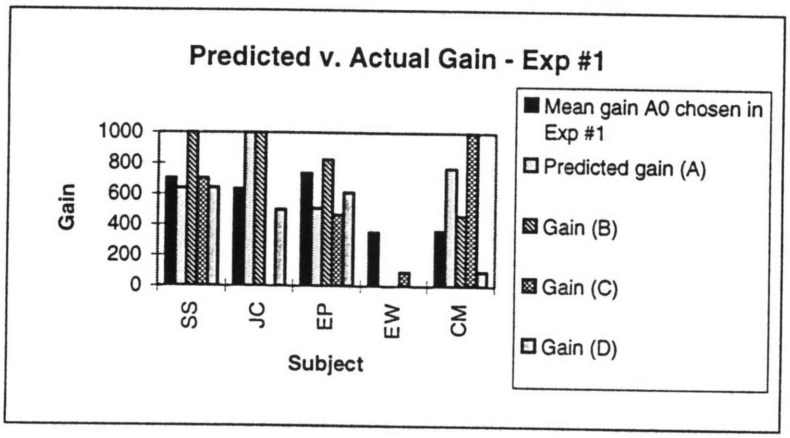


Figure 4: Predicted v. Actual Gain A_0 for the four analysis methods.

4.2 Experiment 2

Analysis of the second experiment used similar methods as the analysis in experiment 1. Analyzing the first experiment involved comparing a number of effective spectra to one recorded HRTF to produce a predicted A_0. In the second experiment analysis, one effective spectrum was compared to all the recorded HRTFs to predict a position. A_0 was set equal to A_{60} in computing the effective spectrum for each analysis method.

The first method correlated the magnitude of the monaural effective spectrum ($A_0 = A_{60}$) with the monaural recorded HRTFs. The position corresponding to the highest correlation was labeled predicted position (A). The overall shape of the magnitude of the monaural spectra were matched in this method.

In the second method, the effective interaural level difference spectrum ($A_0 = A_{60}$) was correlated to the recorded interaural difference HRTFs. The highest correlation corresponded to the predicted position (B). This process matched the overall shape of the interaural level difference spectra.

The third method correlated the effective interaural phase difference spectrum to the phase of the recorded interaural difference HRTFs. This method predicted position (C) and matched the overall shape of the interaural phase difference spectra.

Finally, spectral notches of the effective monaural spectrum and the HRTFs were matched visually. The frequency of the notch, Fnotch, in the effective monaural spectrum was determined. The HRTFs that had a notch within 1 kHz of Fnotch were then compared. The predicted position (D) had a notch whose magnitude closely corresponded to the magnitude of the notch in the effective spectrum. Since the first three methods match overall shape, they do not match the spectral notches.

Table 4 below shows the positions at which each subject perceived the summed sound and the positions predicted by the four analysis methods. Notch matching (D) provides the most consistent approximations for all subjects. The method that correlated the magnitudes of the monaural spectra (B) generally approximated the actual result with 20°. Only this method's prediction for subject EP was behind the head. Correlating the interaural difference spectra, (C) and (D), did not produce approximations that were consistently reasonably close to the actual results. Figure 5 below illustrates the differences in predicted positions for the four analysis methods. The spectra from the predicted positions are shown in Appendix 2.

Subject	SS	JC	EP	EW	CM
Pos. perceived in Exp #2	25°	25°	150°	0°	10°
Predicted position (A)	60°	0°	60°	55°	20°
Predicted position (B)	45°	5°	145°	110°	25°
Predicted position (C)	130°	85°	90°	0°	80°
Predicted position (D)	30°	50°	20°	35°	25°

Table 3: Predicted positions using the four analysis methods.

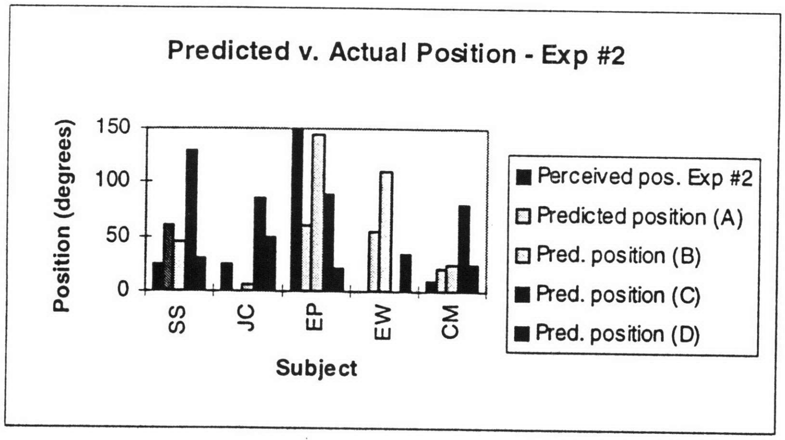


Figure 5: Predicted v. actual position using the four analysis methods.

5. Conclusions

All subjects consistently matched the summed sound position to the target as evidenced by the relatively small standard deviation associated with the gains assigned to the summed speakers in experiment 1. The difference in mean gain among the subjects indicates that the cues controlling perceived vertical location are idiosyncratic across subjects. This work examined whether localization can be explained by monaural spectral cues, interaural level cues, interaural phase cues, or the position of the first large spectral notch.

Predictions of sound location using overall spectral shape of the magnitude of the monaural level spectra and of interaural phase differences were not accurate for subjects in experiments 1 and 2. However, correlations of interaural level difference spectra produced reasonable predictions for 2 of 5 subjects in experiment 1 and for 4 of 5 subjects in experiment 2.

Based on the predictability of mean gain, notch matching produced the best overall predictions across all subjects for both experiments. This data indicates that subjects attempt to match spectral notches in the summed sound to the notches in their HRTFs when locating summed sound sources in the vertical plane. However, since correlations of interaural level difference spectra provided accurate predictions for some subjects, it is possible that in addition to matching notches, some listeners use interaural level differences to localize sound in the vertical plane.

Further research should examine what occurs perceptually when the target-speaker separation is varied and also when the summed speakers are moved below 0° and above 60°. Also, increased experimentation with the difference between left and right HRTFs may reveal a cue used to localize sounds in the vertical plane.

References

Blauert, J. (1983). Spatial Hearing: the phychophysics of human sound localization. Cambridge, MA: MIT Press. 44.

Butler, Robert A. *et al.* (1990). "Binaural and monaural localization of sound in two-dimensional space." 254.

Watkins, Anthony J. (1978). "Psychoacoustical aspects of synthesized vertical locale cues." Journal of the Acoustical Society of America. **63**(4). 1152-64.

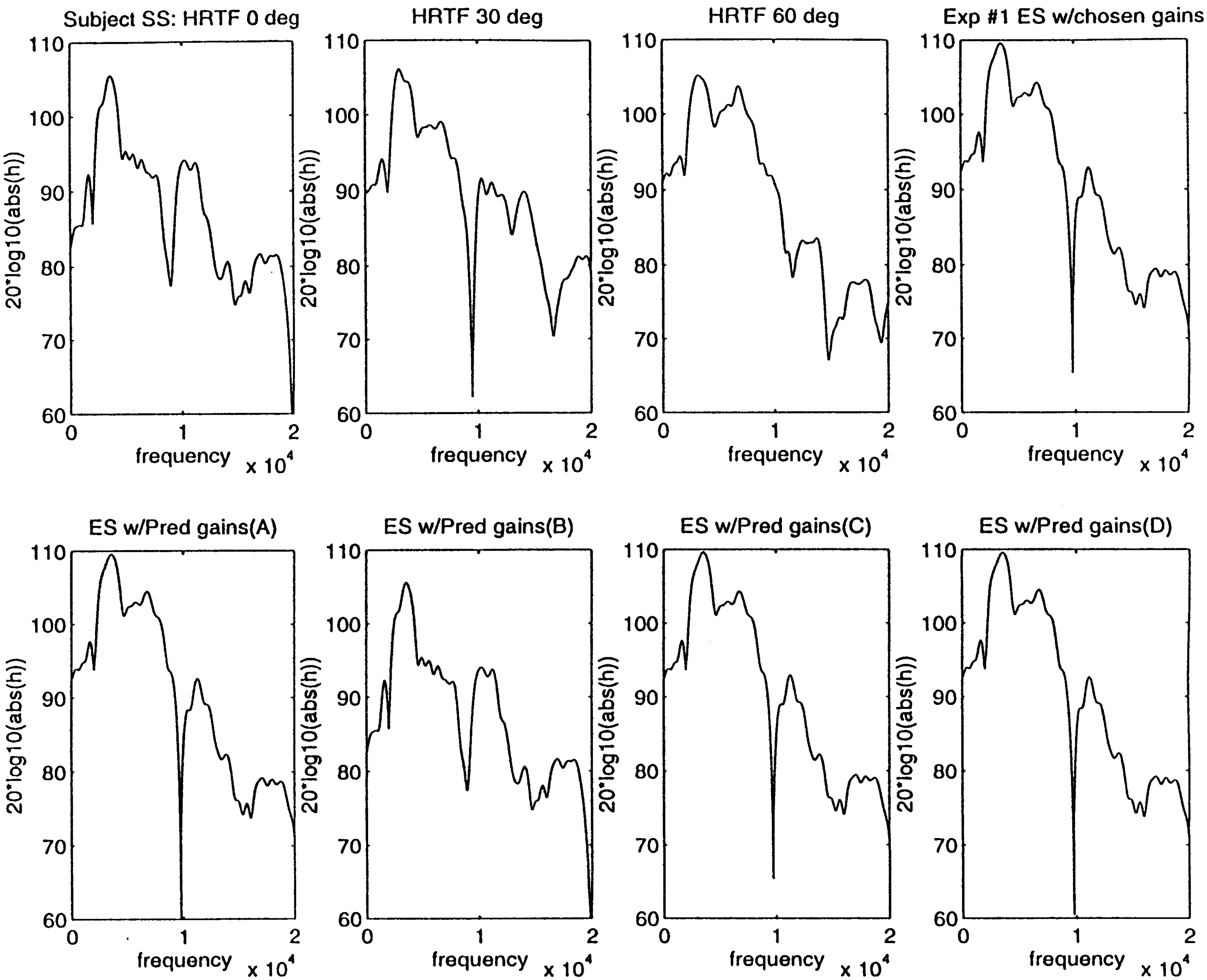
Wenzel, Elizabeth M. *et al.* (1993). "Localization using nonindividualized head-related transfer functions." Journal of the Acoustical Society of America. **94**(1). 111-22.

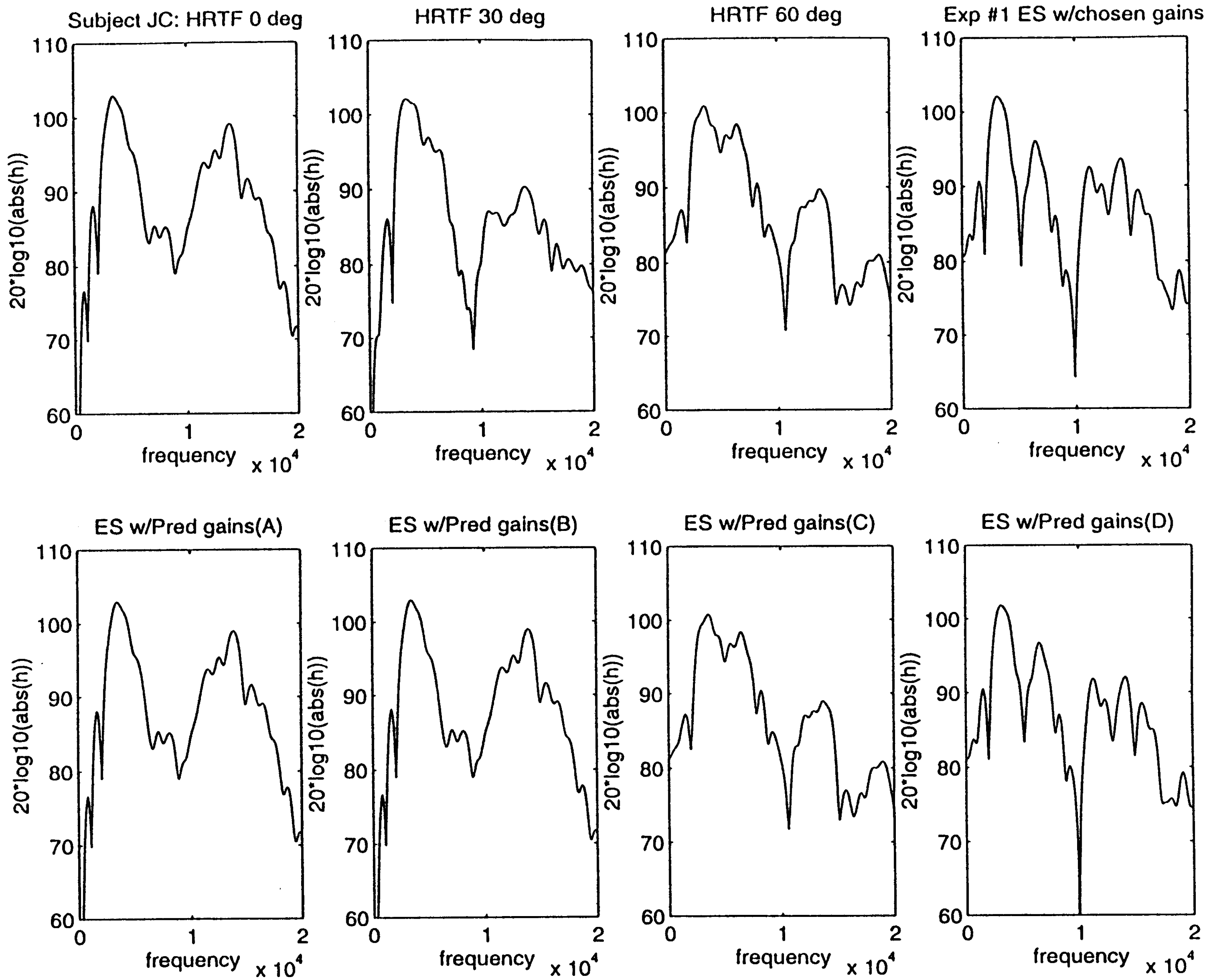
Appendix 1

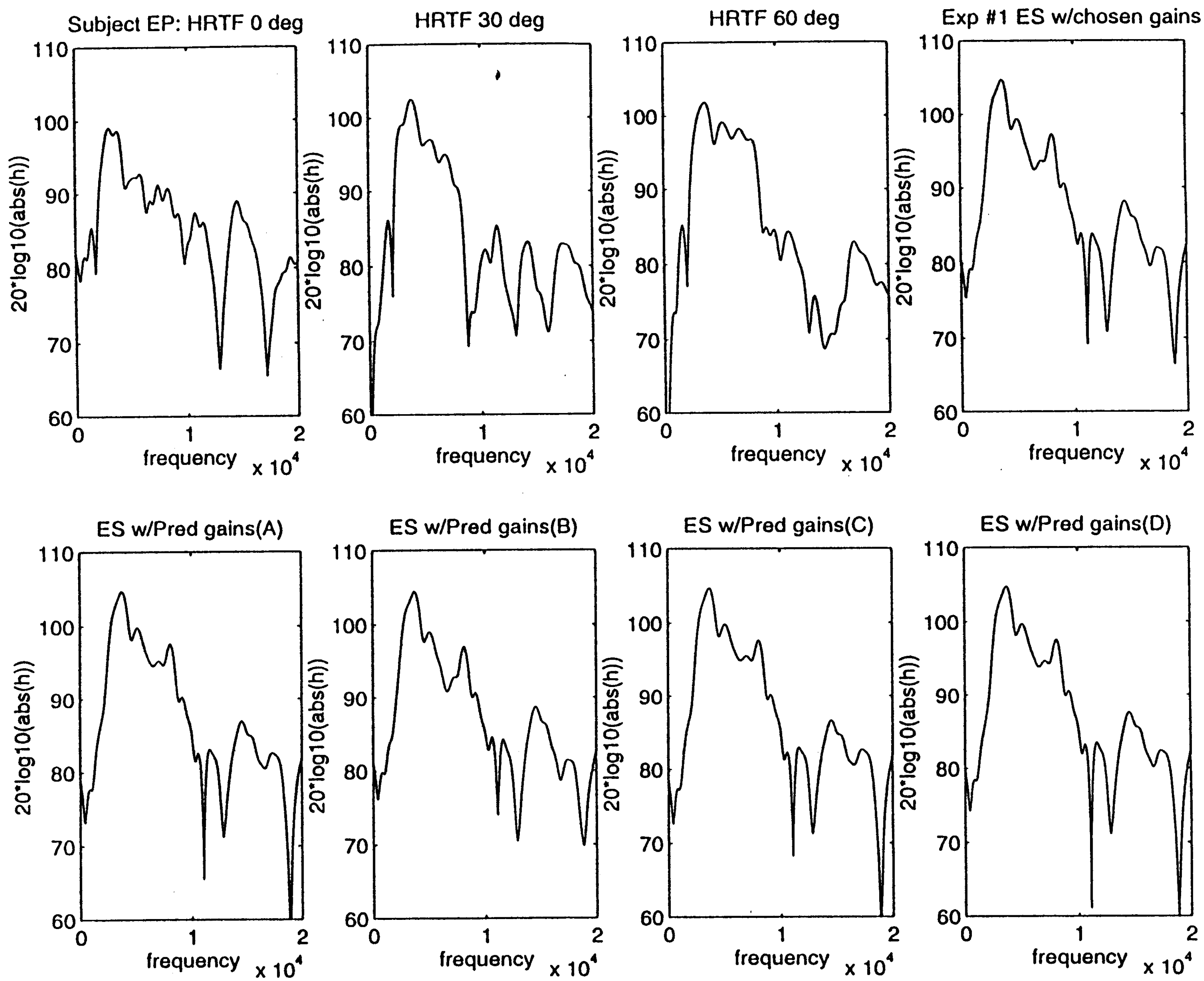
Effective spectra using predicted

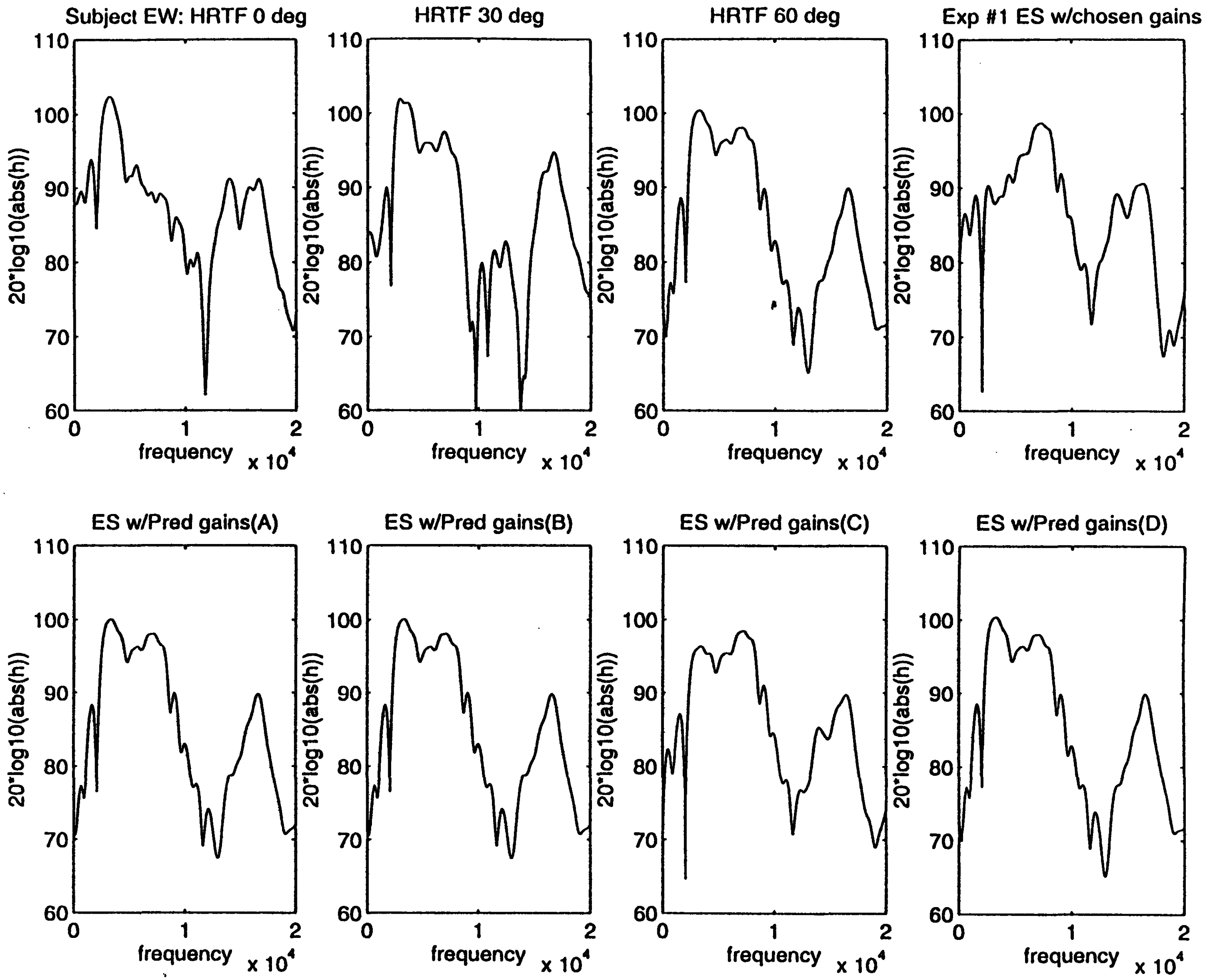
gains from the four analysis

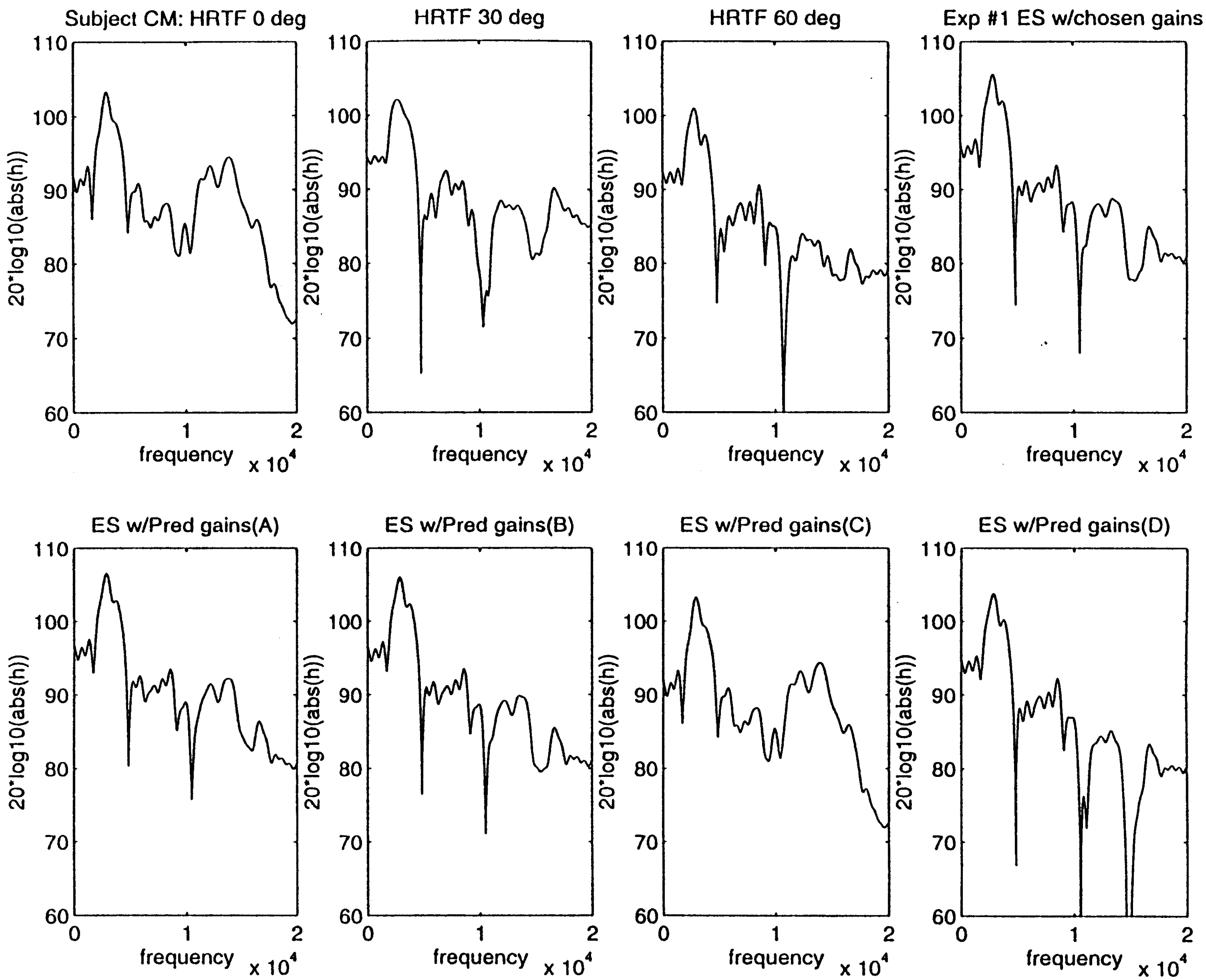
methods relating to experiment 1.


Subject SS: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #1 ES w/chosen gains
ES w/Pred gains(A)
ES w/Pred gains(B)
ES w/Pred gains(C)
ES w/Pred gains(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2


Subject JC: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #1 ES w/chosen gains
ES w/Pred gains(A)
ES w/Pred gains(B)
ES w/Pred gains(C)
ES w/Pred gains(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2


Subject EP: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #1 ES w/chosen gains
ES w/Pred gains(A)
ES w/Pred gains(B)
ES w/Pred gains(C)
ES w/Pred gains(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2


Subject EW: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #1 ES w/chosen gains
ES w/Pred gains(A)
ES w/Pred gains(B)
ES w/Pred gains(C)
ES w/Pred gains(D)
20*log10(abs(h))
110
100
90
80
70
60
0
1
2
frequency
x 10^4

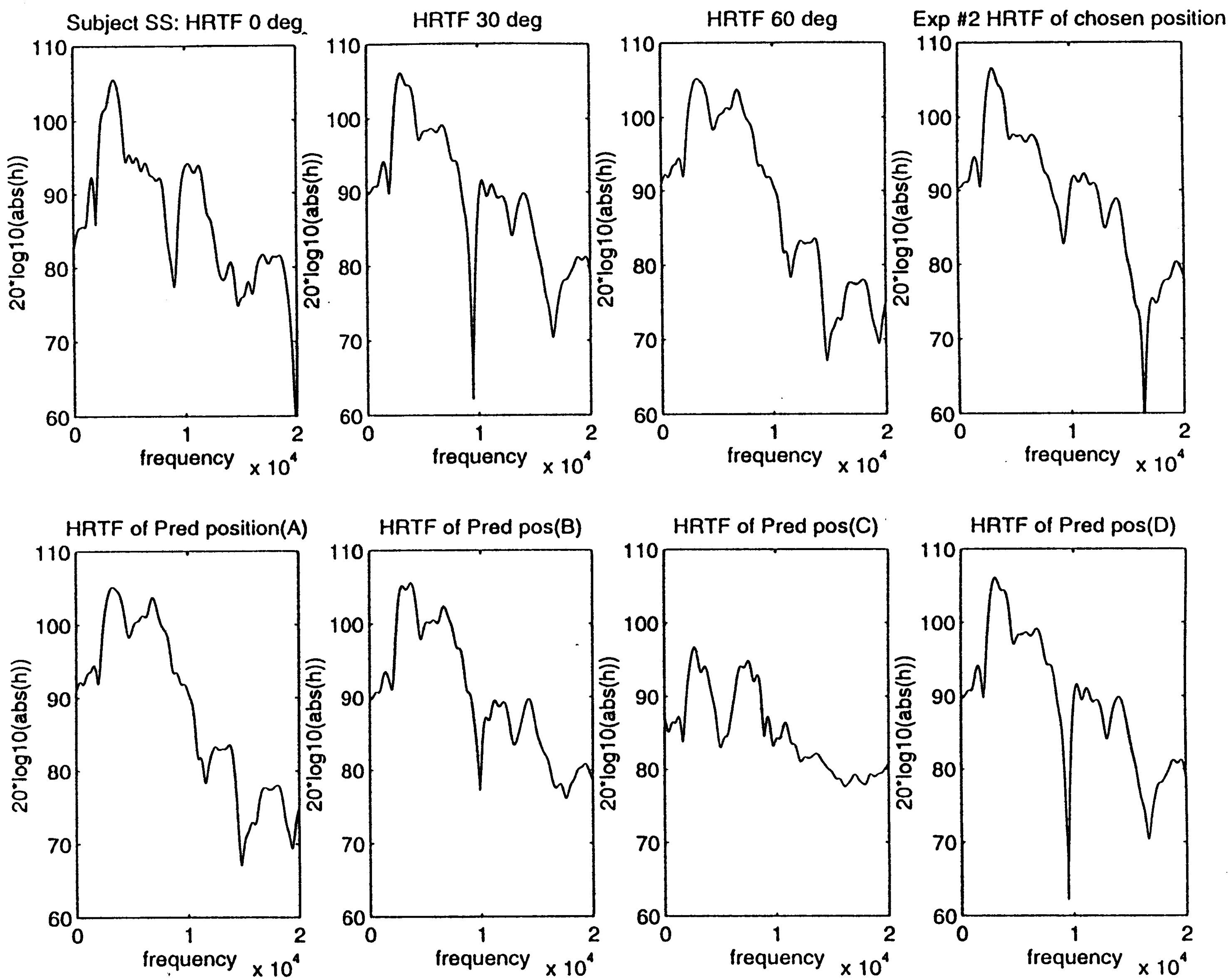

Subject CM: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #1 ES w/chosen gains
ES w/Pred gains(A)
ES w/Pred gains(B)
ES w/Pred gains(C)
ES w/Pred gains(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2

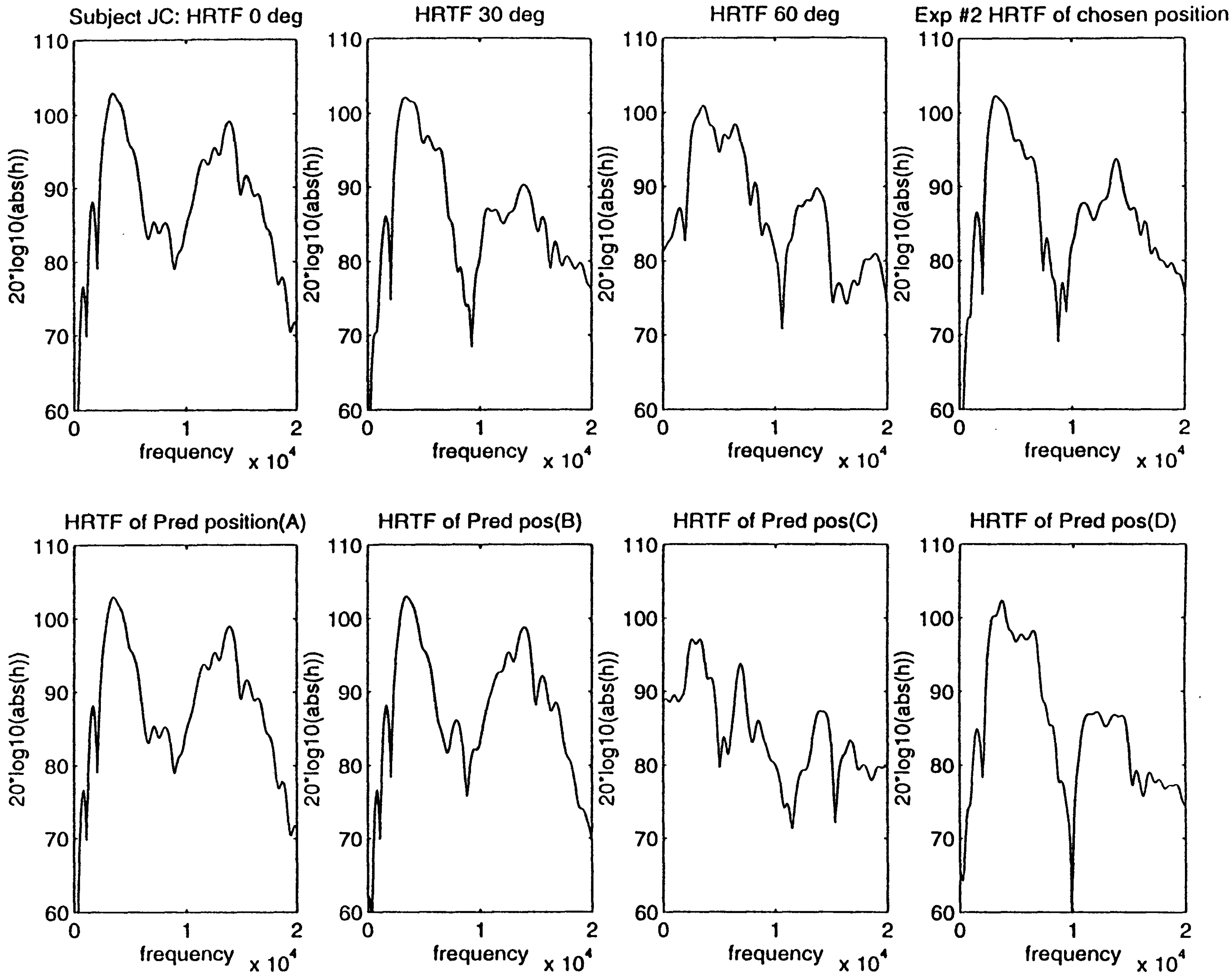
Appendix 2

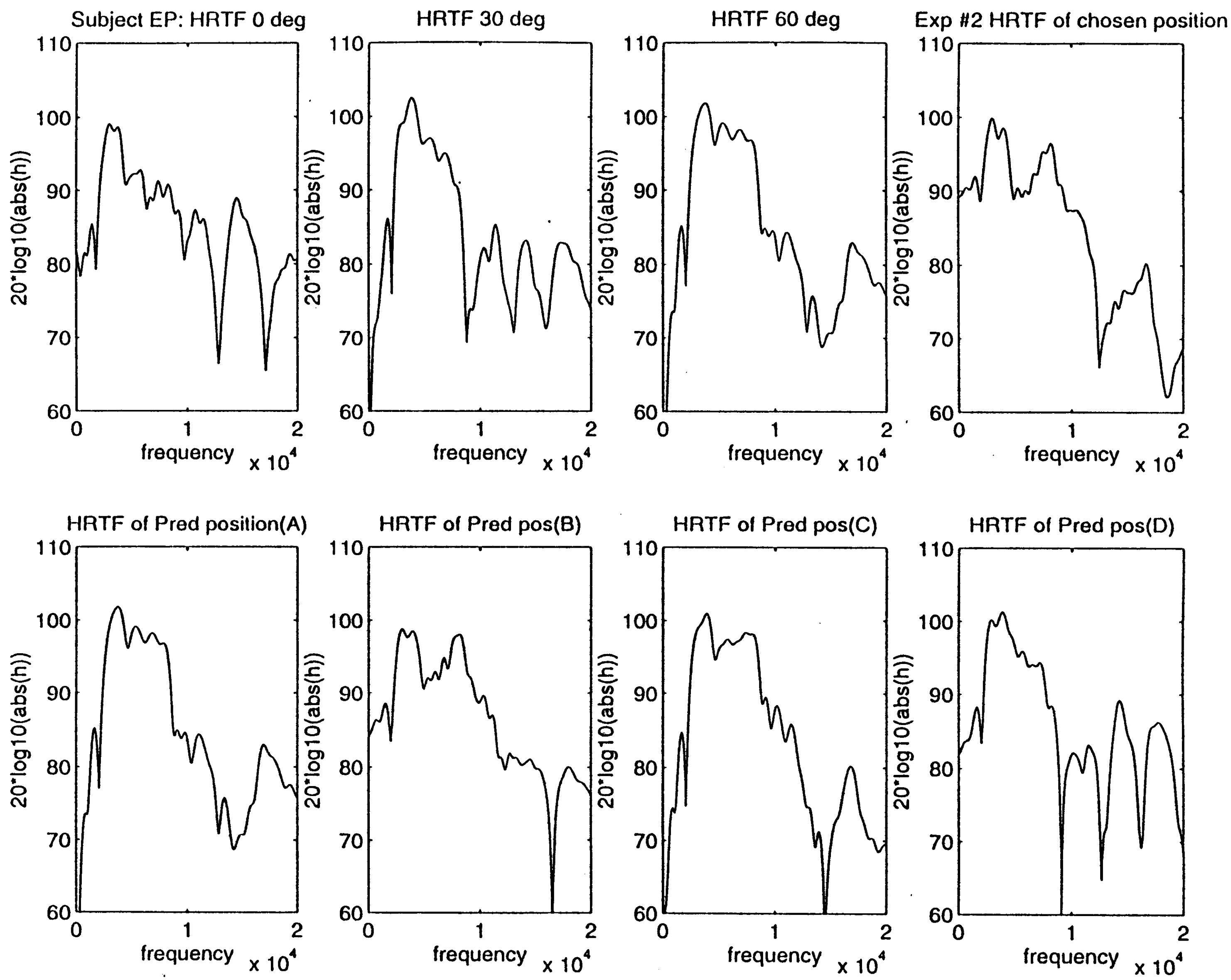
HRTFs for the positions predicted

by the four analysis method

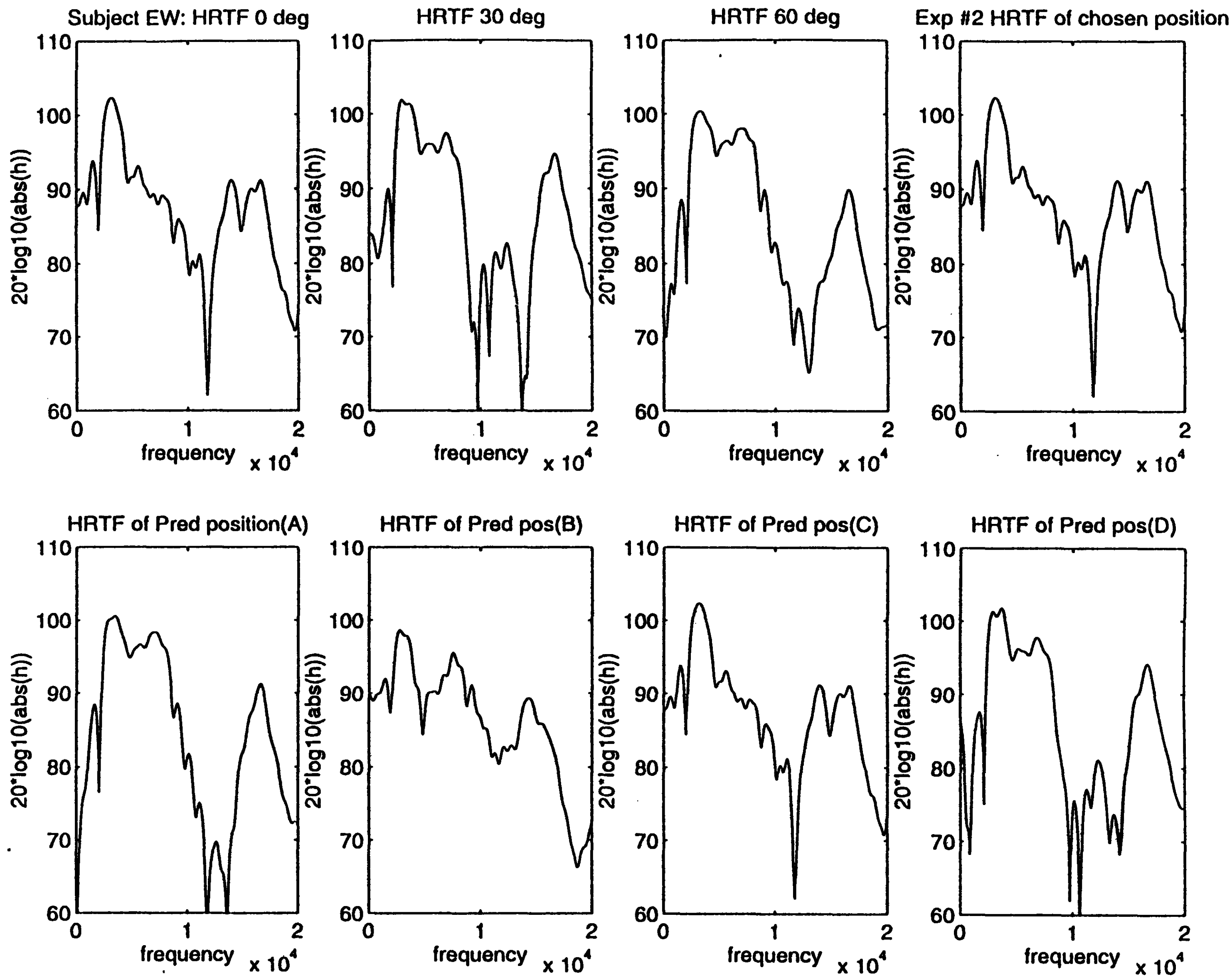
related to experiment 2.

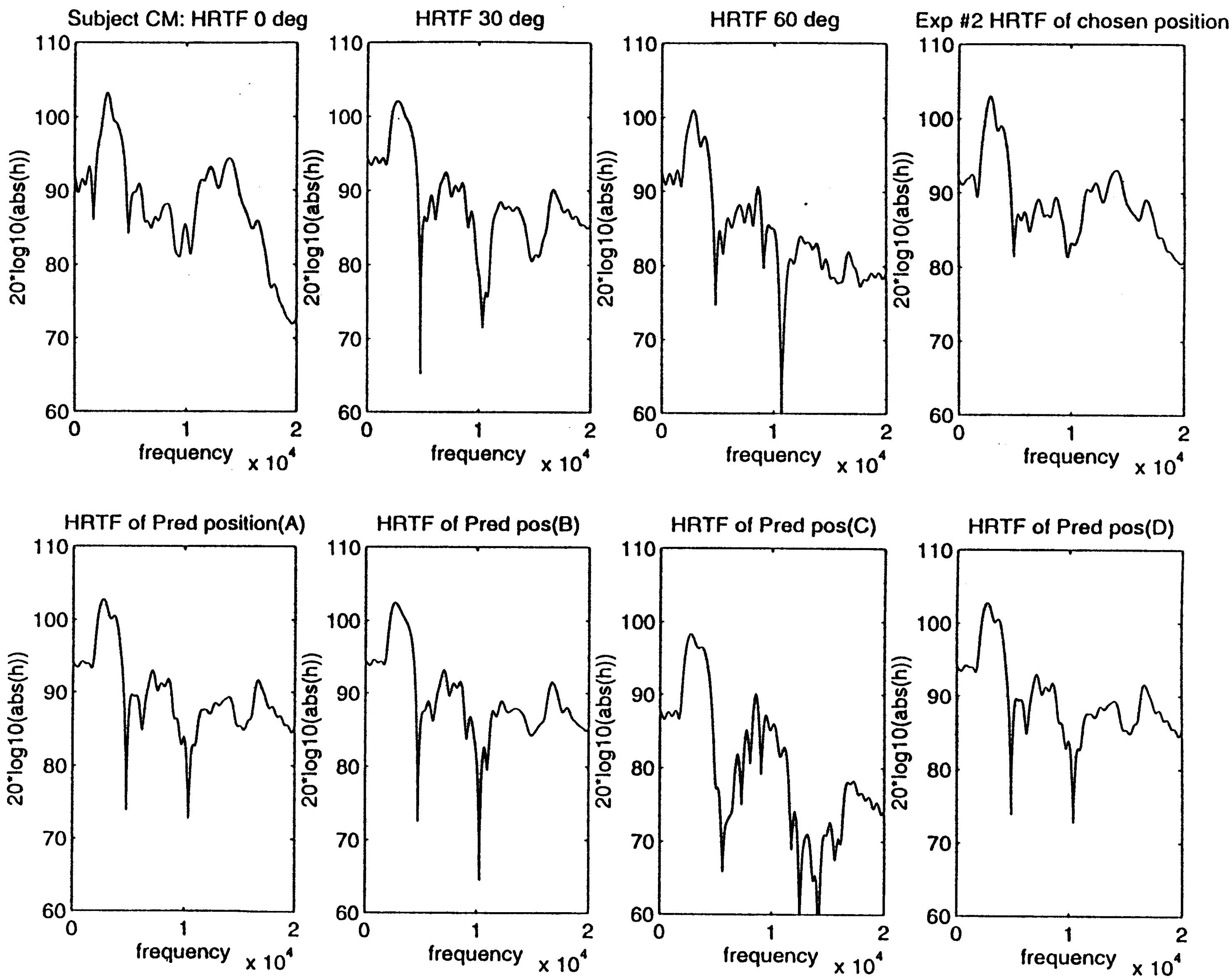

Subject SS: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #2 HRTF of chosen position
HRTF of Pred position(A)
HRTF of Pred pos(B)
HRTF of Pred pos(C)
HRTF of Pred pos(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2


Subject JC: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #2 HRTF of chosen position
HRTF of Pred position(A)
HRTF of Pred pos(B)
HRTF of Pred pos(C)
HRTF of Pred pos(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2



Subject EP: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #2 HRTF of chosen position
HRTF of Pred position(A)
HRTF of Pred pos(B)
HRTF of Pred pos(C)
HRTF of Pred pos(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2


Subject EW: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #2 HRTF of chosen position
HRTF of Pred position(A)
HRTF of Pred pos(B)
HRTF of Pred pos(C)
HRTF of Pred pos(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2


Subject CM: HRTF 0 deg
HRTF 30 deg
HRTF 60 deg
Exp #2 HRTF of chosen position
HRTF of Pred position(A)
HRTF of Pred pos(B)
HRTF of Pred pos(C)
HRTF of Pred pos(D)
20*log10(abs(h))
frequency
x 10^4
110
100
90
80
70
60
0
1
2