A PERCEPTION BASED APPROACH FOR ACOUSTIC EVENTS MODELING IN INTERACTIVE SOUND FIELD NETWORK

Manabu fukushima[1], *and Hirofumi yanagawa*[2]

[1] Fukuoka Institute of Technology, 3-30-1 Wajiro-Higashi, Higashi-ku, Fukuoka, 811-0295, JAPAN
fukusima@fit.ac.jp
[2] Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba, 275-0016,JAPAN
yanagawa@net.it-chiba.ac.jp

ABSTRACT

This paper describes the relation between physical / acoustic parameters and psychological scale for the sound fields in order to create an artificial impulse response of the room based on the perception. First, 19 specific words were chosen that expressing subjective impressions of the sound field from a Japanese language dictionary with 42,000 vocabularies. To classify the 19 words, speech sounds are compared in the way of dichotic listening. The speech sounds are convolution of an anechoic speech and impulse responses of rooms measured by using a dummy head microphone. The words are clustered into 4 categories, 1) high tone timbre, 2) low tone timbre, 3) spaciousness and 4) naturalness or clearness. Then, the 'spatial impression' was selected among 19words and a scale of it was obtained by way of Thurstone's case V since it is one of the important factors in the sound field design. Second, to create an impulse response corresponding to the 'spatial impression', we investigate the relation between the 'spatial impression' and physical/acoustic parameters. As a result, we found that the initial part of impulse response is an important part for controlling 'spatial impression'. The result is confirmed by listening test using artificial impulse responses. Finally, we propose a psychological approach for AEML(Acoustic Modeling Language) for Interactive Sound Field Network.

1. INTRODUCTION

The purpose of our study is to realize a system that provides a virtual sound space sharing in a network environment. We call the system as ISFN (Interactive Sound Field Network). The concept of ISFN and its modules are figured out as figure 1. An interactive use of the sound space sharing such as a virtual conference room[1] needs real time reproduction of room impulse responses.

When we share the sound space by using a network with limited bandwidth, the amount of the data to reproduce the sound space should be reduced. As the data reduction technique, MPEG-4 proposes structured sound description[2]. However, it is not fully sufficient to describe acoustic events, since its major subject is sound signal description. It means MPEG-4 deals less with the sound field than with sound signal. Therefore we propose a concept of AEML (Acoustic Events Modelling Language)[3], which is the language for

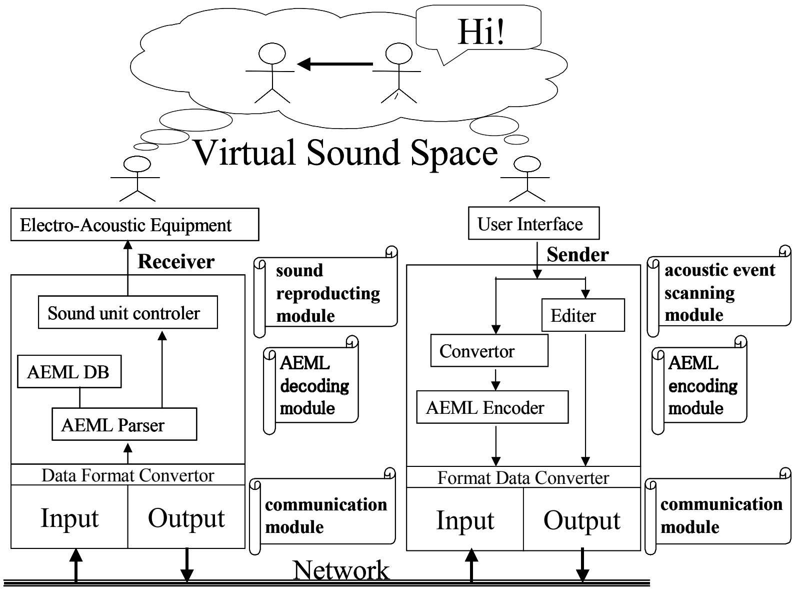


Fig.1 Concept of Interactive Sound Field Network

describing acoustic events including description of the sound field and reproduces impulse responses of the sound space. The characteristic of a human perception for the sound field can be applied to reduce the amount of the data by describing the sound filed using AEML.

It is necessary to scale the characteristic of a human perception (psychological scale) for the AEML. In this study, the relation between physical/acoustic parameters and psychological scale for the sound field is examined in order to create the room impulse response based on the perception of the sound field.

2. WORDS EXPRESSING THE SOUND FIELD

We chose words that expressing subjective impressions of the sound field from a Japanese language dictionary[4] which contains about 42,000 vocabularies. For the first step, native Japanese speakers chose candidate words from the dictionary. The words were judged from their usable ness when they were used as adjective for a sound space. Furthermore, the words were selected on the condition that a scene of the sound field could be pictured with the word. Finally, 19 words were picked out as shown in table 1. Table 1 shows the selected words that expressing subjective impression of the sound field.

The selected words were classified by hearing test in the way of dichotic listening with 16 subjects as shown in figure 2. The headphones used are Pioneer SE-900D. Test signals were generated as a convolution of impulse responses of 7 rooms which volume ranges from 0.53[m^3] to 6041[m^3] as shown in table 2 / 3 and an echoic female speech signal[5] . The impulse responses were measured by using a dummy head microphone (OSS HATS). Figure 3 shows an example of a measured impulse response and transfer function.

Table 1 Selected words that expressing subjective impression of the sound field

	words		words		words
1	spacious	8	clear	15	dry
2	reverberate	9	rough	16	flutter echoic
3	echoic	10	hard	17	natural
4	deep	11	brilliant	18	beautiful
5	hollow	12	bright	19	pleasant
6	clean	13	warm		
7	booming	14	heavy		

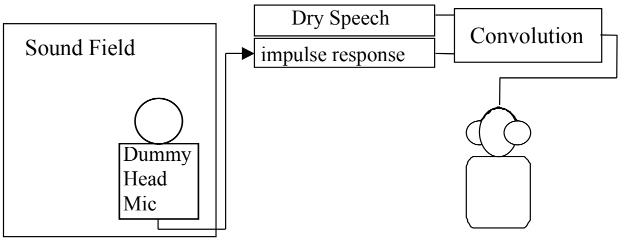


Fig.2 The configuration of healing test

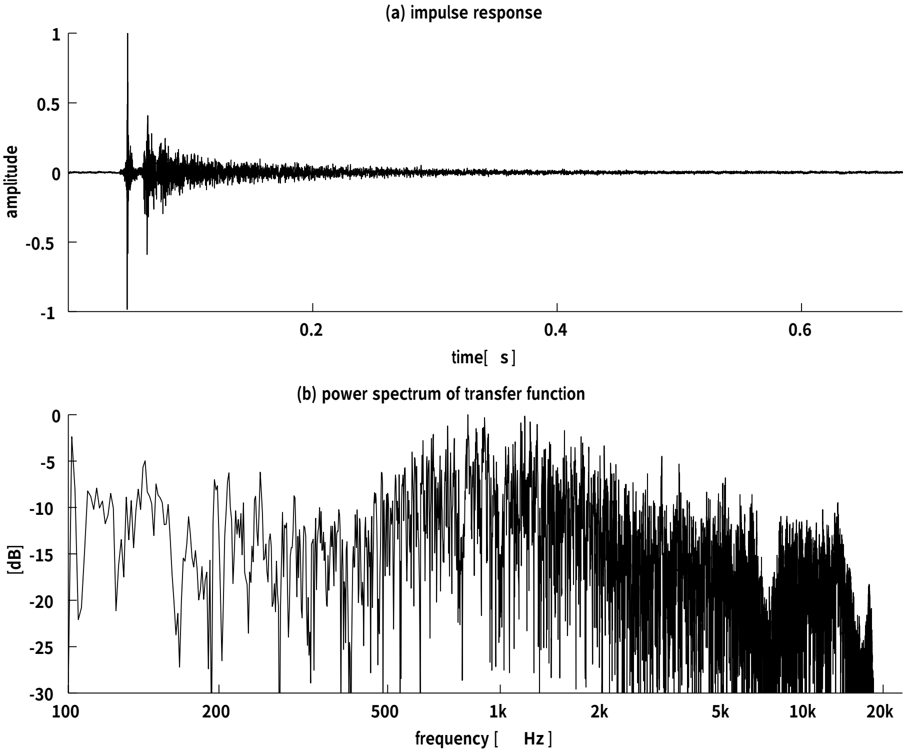


Fig.3 An example of a measured impulse response and transfer function

Table 2 Volume of rooms

room name	Volume [m^3]
lab1	379（L22.9×W7.2×H2.3）
lab2	63.1（L3.65×W7.2×H2.4）
class room 1101	1911.2
class room 5101	1000
path way	500
small room	4.56（L1.95×W1.2×H1.95）
one side hard wall	0.53（L0.93×W0.77×H0.74）

Table 3 Physical characteristic of rooms

room name	D_{50} [%]	C_{80} [dB]	R [dB]	STI
lab1	92	14	-0.35	0.94
lab2	89	12	-0.52	0.93
class room 1101	71	7	-1.46	0.85
class room 5101	82	10	-0.88	0.89
path way	48	3	-3.2	0.72
small room	74	8	-1.32	0.88
one side hard wall	99	21	-0.05	0.98
	EDT[ms]	Ts[s]	RT_{500Hz}[s]	IACC
lab1	25	0.0159	0.489	0.588
lab2	16	0.0209	0.516	0.530
class room 1101	3	0.0575	1.050	0.302
class room 5101	4	0.0297	0.678	0.446
path way	458	0.0795	1.304	0.112
small room	208	0.0412	0.567	0.060
one side hard wall	12	0.0104	0.082	0.092

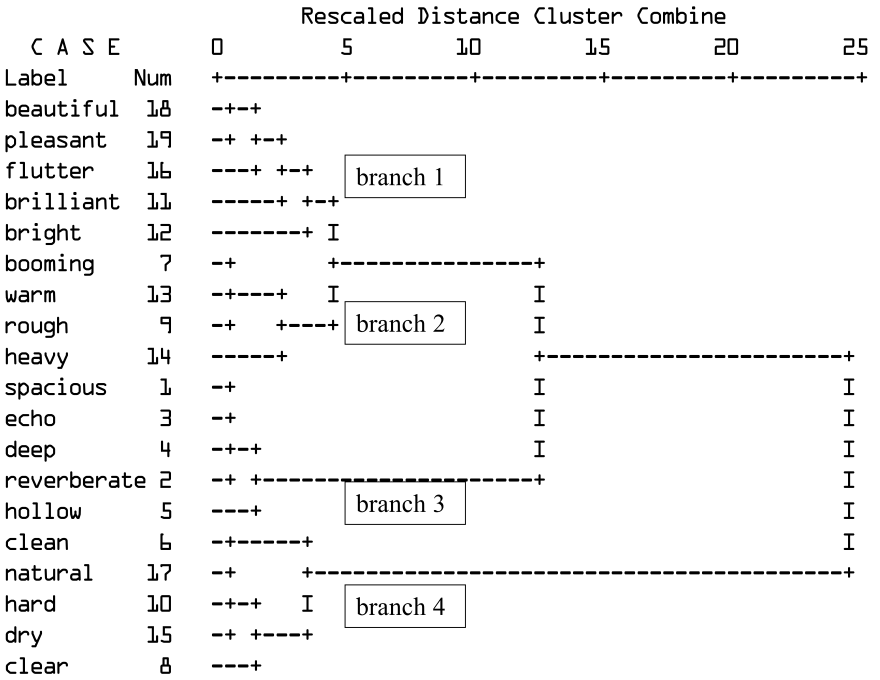
The words are clustered into 4 categories by using a statistical analysis software SPSS as shown in figure 4, which are, branch 1: high tone timbre, branch 2: low tone timbre, branch 3: spaciousness and branch 4: naturalness or clearness.

3. PSYCHOLOGICAL SCALE FOR 'SPACIOUS'

We focus on a word 'spacious'[6] that belongs to branch 3: spaciousness, since it is one of the important factors in the sound field design. It is necessary to scale the psychological result for creating impulse responses. For scaling the 'spacious', we add other 8 rooms (totally 15 rooms) in the hearing test. The selection is done by the evaluation of the variance of physical characteristics of 40 impulse responses as shown in figure 5. Thurstone's case V[7] was applied to scale.

In order to find the relation between the psychological scale and the physical parameters, we apply 9 physical parameters, 1) room volume, 2) D: deutlichkeit, 3) C: clarity, 4) R: hallmass, 5) STI: speech transmission index, 6) EDT: early decay time, 7) Ts: center time, 8) RT: reverberation time and 9) IACC: inter aural cross correlation. Figure 6 shows the relation between the result of the scaling 'spacious' and physical parameters. In fiure 6, correlation coefficients between each physical/acoustic parameters and psychological scale are also shown.

The correlation coefficients shows that the 'spacious' correlates with D, C, R, RT, EDT, and Ts. Therefore



```
                         Rescaled Distance Cluster Combine
  C A S E         0         5        10        15        20        25
Label        Num  +---------+---------+---------+---------+---------+
beautiful     18  -+-+
pleasant      19  -+ +-+
flutter       16  ---+ +-+   branch 1
brilliant     11  -----+ +-+
bright        12  -------+ I
booming        7  -+       +---------------+
warm          13  -+---+   I               I
rough          9  -+   +---+ branch 2      I
heavy         14  -----+                   +-----------------------+
spacious       1  -+                       I                       I
echo           3  -+                       I                       I
deep           4  -+-+                     I                       I
reverberate    2  -+ +---------------------+                       I
hollow         5  ---+       branch 3                              I
clean          6  -+-----+                                         I
natural       17  -+     +-----------------------------------------+
hard          10  -+-+   I   branch 4
dry           15  -+ +---+
clear          8  ---+
```

Fig. 4 The result of clustering the specific words

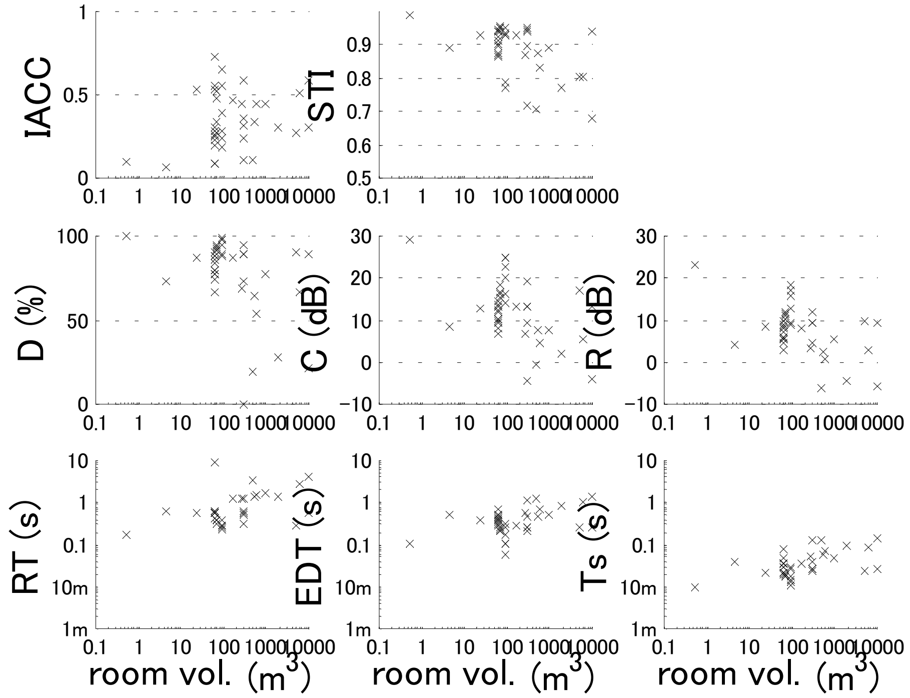


Fig.5 The variance of physical characteristic of 40 impulse

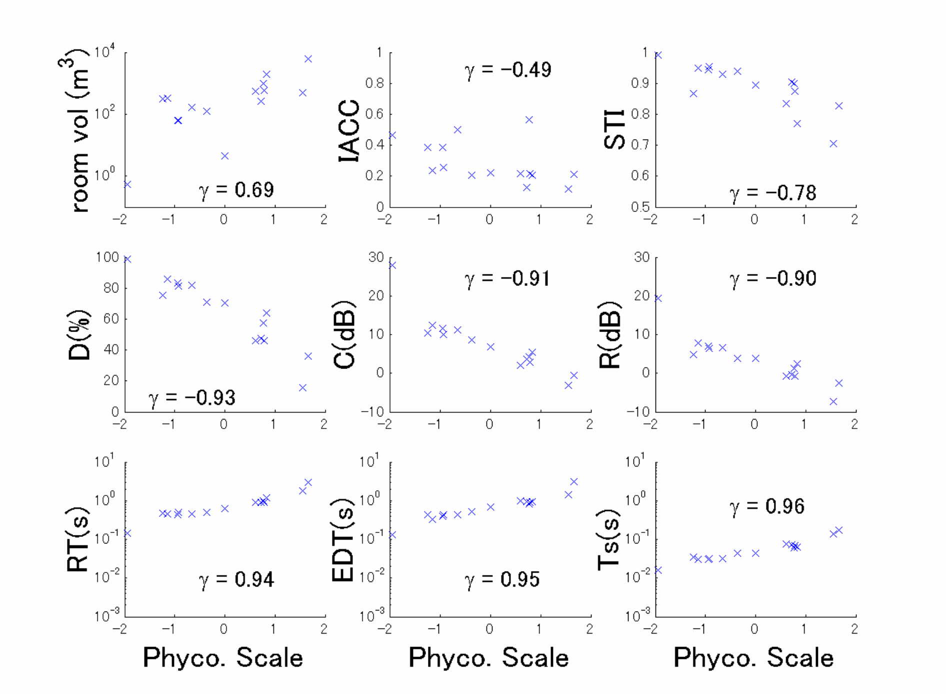


Fig. 6 The relation between physical parameters of rooms and psychological scale of 'spacious'

the 'spacious' seems to be controlled by changing these parameters. To confirm it, the hearing test was done by using simulated impulse responses instead of room impulse responses. The impulse response was calculated by an image method for a rectangular room. Then, it was changed in two ways. (i) The ratio between the direct sound and the reverberation was altered. This causes change of Ts, D, R and so on. (ii) The decay rate was altered between an initial part of the impulse response and the other part of it. This means change of EDT. The result of the hearing test by diotic listening is shown at an upper half of figure 7 where Ts is a value changed in the way of (i) and also EDT in the way of (ii). Lower two figures in figure 7 are extracts from figure 6. Figure 7 shows that the 'spacious' is controlled by changing Ts and EDT. In figure 7, correlation coefficients are also shown.

4. PSYCHOLOGICAL APPROACH FOR ACOUSTIC EVENTS MODELING LANGUAGE (AEML)

We are working out for the AEML approaching physical and psychological. In this paper mainly psychological approach is described.

It is necessary to define an axis of the world of sound space. We call the world as *World Space* and the axis as *World Space Map*. The definition of the World Space Map is;

World Space Map = {axis_length,
{{room_definition, location} ......},
{{object, location} ....}}
axis_length = {width, height, length}
location = {origin_point, rotation}

The axis_length is a definition for limit the sound space. The room_definition is a set of boundaries in the space. The object is a reflector or a sound source in the world. The rooms and objects are placed in the World Space Map. The condition of how place them are described by location.

room_definition = {room_ID, volume, define_type,
{width, room_position},.............},
{length, room_position}, ...},
{height, room_position}, ...},
{{boundary, room_position}, ...},
{{psychological_parameter},.....}}

psychological_parameter =
{{'specfic word', value} , ...}

The room_position is described using a local axis in the room. The origin point of the local axis is the front left bottom. By using the World Space Map and room_definition, basic space definition can be written. The definition includes both parameters, one is physical value and the other is psychological parameter described in above. These parameters make confliction between physical definition and psychological definition. Therefore the definition includes define_type which indicates the definition is based on the physical parameter or psychological parameter.

Primitive room style can be defined with the room_definition. Thus the rectangle room or fan fold are defined and used as a primitive polygon.

The sound source like a speaker and reflector such as desk, chairs and so on are necessary for a sound field. This means the substance placed in the World Space Map can be categorized in two types. One is sound source, and the other is reflector. A speaker can be categorized a combination of them.

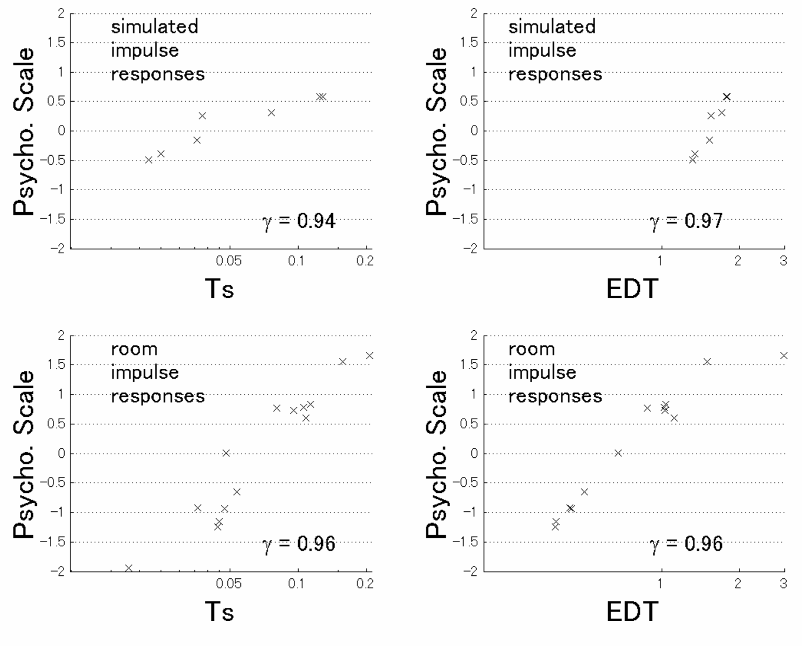


Fig. 7 The change of psychological scale of 'spacious' by Ts and EDT

substance = {substance_ID,
{{substance, substance_position},.............} |
{{sound_source, room_position}, ...} |
{{sound_reflector, room_position}, ...}}
sound_source = {sound_data, volume, directivity}

The substance_position is a local position like the room_position. One root object must be defined. The branch substances are defined within the root substance or a substance which related to the root substance. The sound_data which includes in sound_source can be defined as an online streaming data or stored data.
With the definition *static* sound field can be described.

5. MOTION DESCRIPTION

The motion is defined for the objects and the boundaries. We defined here the motion as;

motion = {{substance_ID | room_ID}, direction,
speed_parameter,
start_time, end_time, duration}

The direction is described using the local axis of its room object. The direction is described with a function or a vector form. The speed_parameter is described with a function or a constant value.

6. TIME CONTROL IN ISFN

The motion description includes a time sequence. For realize a sound field with the time sequence, the system must control the time in network environment.
Figure 8 shows the start up sequence when a new user is login to the system. When the new user login, it communicate with NTP(Network Time Protocol) server or SNTP(Simple Network Time Protocol) server to update current time. And after update current time, broadcast ISFN start-up packet. When a user receive the packet, the system update the user list with the network time delay

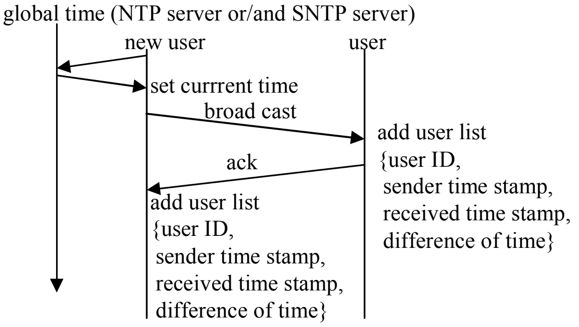


Fig.8 Login and current time setup sequence in ISFN

When an event is occurred the transmission will be follow as in figure 9. Case 1 shows the AEML packet can be received by receiver before limit time. Thus the receiver send back ack packet to the sender. Case 2 shows the AEML packet can not be received by the receiver before limit time. Receiver updates or waits for the next AEML packet and send back nack packet to the sender. When the motion is described with a function receiver skip and update to the time. In that case receiver sends back the ack packet. Case 3 shows the AEML packet can be received close to the limit time. Receiver send back nack packet when the rest time is less than network time delay and action as case 2.

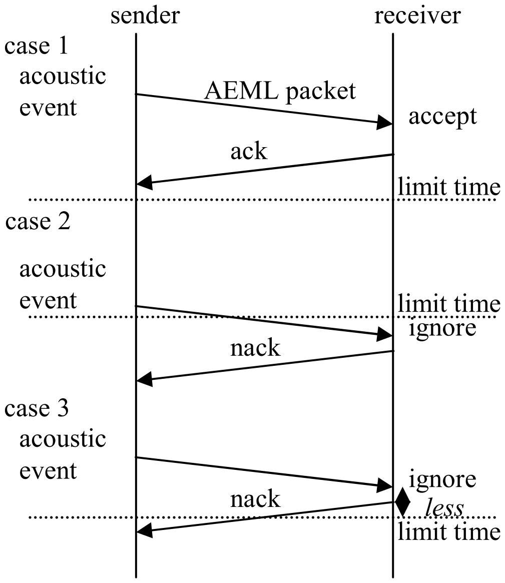


Fig.9 Control sequence in sending and receiving AEML in ISFN

7. CONCLUSIONS

This paper described the relationship between physical/acoustic parameters and psychological scale for the sound field to create the room impulse response that causes a specific subjective impression of the sound field. Firstly, we clustered 19 words expressing subjective impressions of the sound field into 4 categories, which are (1) high tone timbre, (2) low tone timbre, (3) spaciousness and (4) naturalness or clearness. Secondly, we focused on a word 'spacious' that belongs to (3) spaciousness, since it is one of the important factors in the sound field design. We scaled the 'spacious' by Thurstone's case V. As a result, 'spacious' correlates with most of the physical parameters of the sound field and can be controlled by them. Finally, we propose a psychological approach for AEML(Acoustic Modeling Language) for Interactive Sound Field Network.

REFERENCES

[1] M.Cohen and N.Koizumi,"Exocentric control of audio imaging in binaural telecommunication", *IEICE Trans.* E75-A, 164-170(1992)

[2] http://sound.media.mit.edu/mpeg4/

[3] M. Tohyama, M. Kazama, H. Yanagawa, "Acoustic events modeling for 3D sound rendering and perception", *Proceeding of the 4th World Multiconference on Systemics, Cybernetics and Informatics Co-organized by IEEE Computer Soc.,* IS48-3 (2000)

[4] IWANAMI Japanese Dictionary, *IWANAMI SHOTEN, PUBLISHERS* (1994) (in Japanese)

[5] DENON Professional Test CDs COCO-75084->86,DISK2-Track37

[6] M.Fukushima, H.Suzuki, I.Yako, H.Yanagawa, "A perceptual sound field design focusing auditory spaciousness", *Proc. of The 7th WESTPRAC*, Vol.1, pp.313-316 (2000)

[7] L. L. Thurstone,"The measurement of values", *Chicago Univ. Press* (1960)