Synthesis of footstep sounds of crowd from single step sound based on cognitive property of footstep sounds.

Takuro Kayahara* Hiroki Abe**

Department of Spatial Design and Information Systems,
School of Project Design, Miyagi University

ABSTRACT

The crowd sound effect ("Gaya" in Japanese technical word) plays important role to create and perceive the atmosphere of the crowded scene of a movie, but the technique for authoring "Gaya" sound has not been scientifically described so far.

In this study, footstep sounds of the crowd was focused and cognitive characteristics of footstep sounds was described through 3 psychophysical experiments, applying to the technique for authoring effective footstep sounds of crowd.

In the experiment 1, the effect of the varying tempo of recorded footstep sounds to the reality of "walking" or "running" cognition of footstep sound was examined, revealing that there was most effective tempo for the reality of cognition of "walking" and "running" footstep sounds.

In the experiment 2, the interaction effect between footstep duration and inter-step interval in the footstep sound to "walking" or "running" cognition of the footstep sound was examined, indicating the main contribution of inter-step interval to the discrimination between "walking" and "running" cognition.

In the experiment 3, the perceived number of "walkers" from the hearing of the footstep sounds as a function of physical number of those was examined, indicating the non-monotonic increase and the underestimation of perceived number of "walkers" of footstep sounds.

In conclusion, the footstep sound can be synthesized by merely lining recorded single step sound (element sound) alternating with appropriate duration of silence (inter-step interval) and the discrimination between "walking" cognition and "running" cognition was mainly depend on inter-step interval (silence duration).

However, the footstep sounds of the crowd could not be synthesized merely by stacking footstep sound for one person because of the underestimation of the perceived number of "walkers" from the hearing of the footstep sounds in the experiment 3, suggesting that there are some kind of direct representation for crowd footstep sounds itself such as appropriately filtered noise. In fact, sound effect engineer's empirical technique to synthesize "Gaya" sound (crowd sound effect) is known that effective crowd footstep sounds consist of direct crowd sound (such as filtered noise) as background and a few footstep sounds as foreground. Thus, this study could explain those empirical knowledge.

E-mail: {*kayahara|**p0952001}@myu.ac.jp

IEEE International Symposium on Virtual Reality Innovation 2011
19-20 March, Singapore
978-1-4577-0054-5/11/$26.00 ©2011 IEEE

KEYWORDS: Sound effect, footstep sound, synthesis from element sound, cognitive properties, psychophysics.

INDEX TERMS: J.4 [Social and Behavioral science]—Psychology

1 INTRODUCTION

The sound is important information to create and perceive the atmosphere of an environment. Especially for the visual expression like a movie, the crowd sound effect ("Gaya" in Japanese technical word) plays important role to represent the atmosphere of the crowded scene of a movie.

In many cases, those "Gaya" sound is not recorded simultaneously to the filming but authored with many actors/actresses after the filming and editing, however, authoring of "Gaya" sound for a crowded scene of a movie is based on the empirical technique of an sound engineer, whose fundamentals have not been scientifically clarified yet.

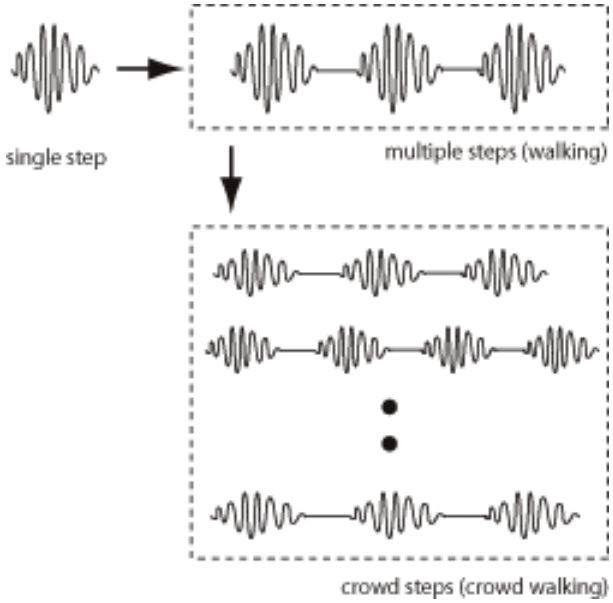

Figure 1. Illustrated concept of the synthesis of the footstep sounds of the crowd from single footstep sound

Traditional technique for authoring "Gaya" sound has another limitation that the authored sound is fixed by mixing down to sound tracks and which can not be applied to interactive situation.

In this study, footstep sounds of the crowd was focused and cognitive characteristics of footstep sound was described with psychophysical experiments, then those characteristics is applied to methodology for authoring effective crowd step sounds from single footstep sound (element sound), whose outline is described in Figure 1.

2 EXPERIMENT 1

In the experiment 1, the subjects evaluated a footstep sound whether they are heard as walking or running, varying tempo of recorded step sound. Thus, an appropriate tempo for walking (or running) footstep sound was obtained through the experiment.

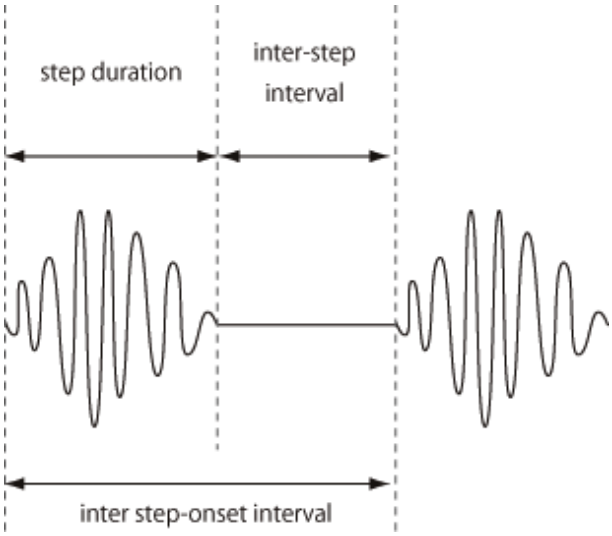

Figure 2. Time configuration of footstep-sound stimulus which is described by 3 parameters, step duration, inter-step interval and inter step onset interval.

2.1 Method

Stimuli: The stimuli for the experiment was digitally recorded footstep sound. The footsteps walking on the graveled path (crunch sound) was recorded by handheld digital recorder (TASCAM, DR-100) held by walker's hand, thus overall level of the footstep sound was almost constant through walking. The recording quality was 44100 Hz and 24 bit, not compressed.

The tempo of the recorded footstep sound was varied in 11 steps from 159 msec to 1750 msec in inter-footstep-onset interval (Figure 2.) with the manner of geometric progression (Table 1.).

Procedure: The 11 variations of the tempo of footstep sound were randomly presented to the subjects via headphone (Sennheiser, HDA200). In one trial, after 5-sec

Table 1. Duration in msec of inter step onset interval, step duration and inter step interval used in experiments.

Inter step onset interval	step duration	inter step interval
159	70	89
175	77	98
194	86	108
219	97	122
250	110	140
292	129	163
350	155	195
438	193	245
583	258	325
875	387	488
1750	773	977

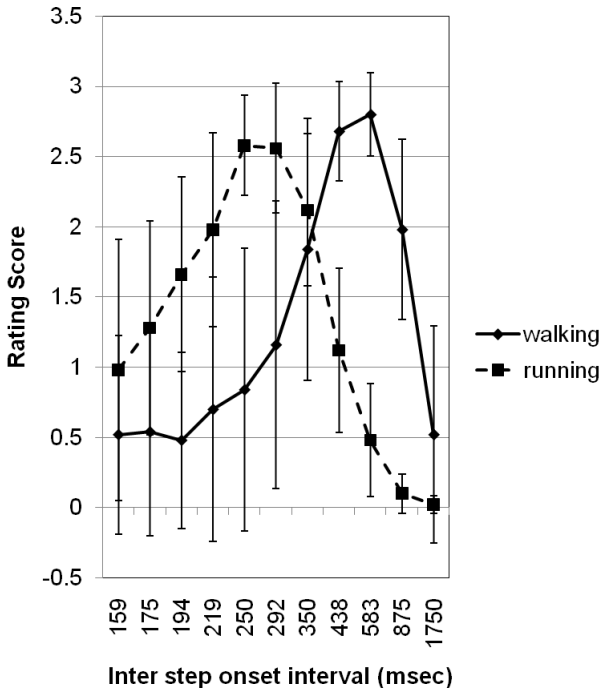

Figure 3. Results of experiment 1. The error bars indicate standard deviation.

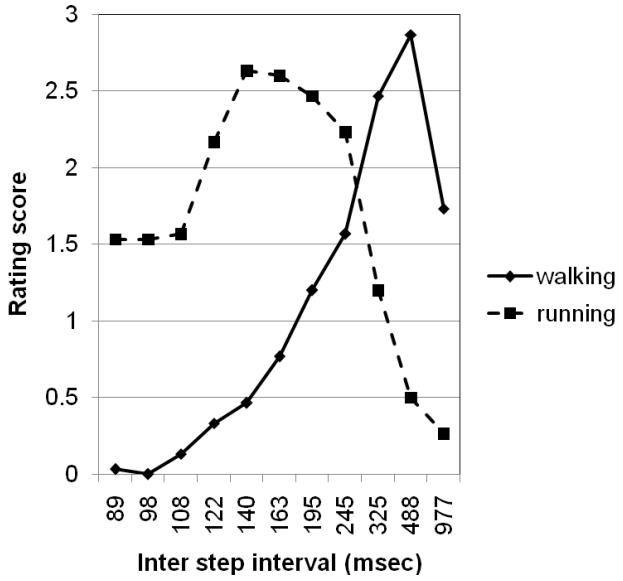

Figure 4. Results of experiment 2 for the step duration of 110 msec.

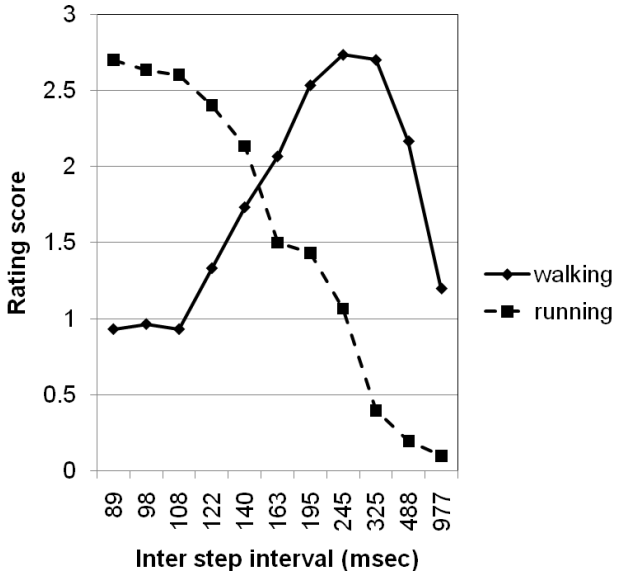

Figure 5. Results of experiment 2 for the step duration of 258 msec.

presentation of the sound of a condition, the subjects were asked to judge the reality of the sound with the rating

scores of 0 to 3 (4 steps), each number of which means 0 as "not walking (running)", 1 as "not so walking (running)", 2 as " probably walking (running)", 3 as "absolutely walking (running)". The judgment of the subjects was performed in terms of both the reality as "walking" sound and the reality as "running" sound for all 11 conditions. The judgment was iterated 5 times for each conditions and points of view of the judgment.

Subjects: 10 graduate students of Miyagi university with ordinary hearing participated in experiment.

2.2 Results and Discussion

The averaged rating scores between all subjects and iterations was shown in Figure 3. Both judgment of walking reality and running reality as a function of the tempo of the footstep sound showed the bandpass-like property. The rating score of the judgment of walking reality was highest at the tempo of 583 msec in inter-footstep-onset interval. The rating score of the judgment of running reality was highest at the tempo of 250 msec in inter-footstep-onset interval. Those property indicated that there were typical tempo for both walking footsteps and running footsteps.

The standard deviation of the rating scores between subjects was bigger than those within subject, indicating judgment of walking and running reality of footstep sound was relatively stable within one subject.

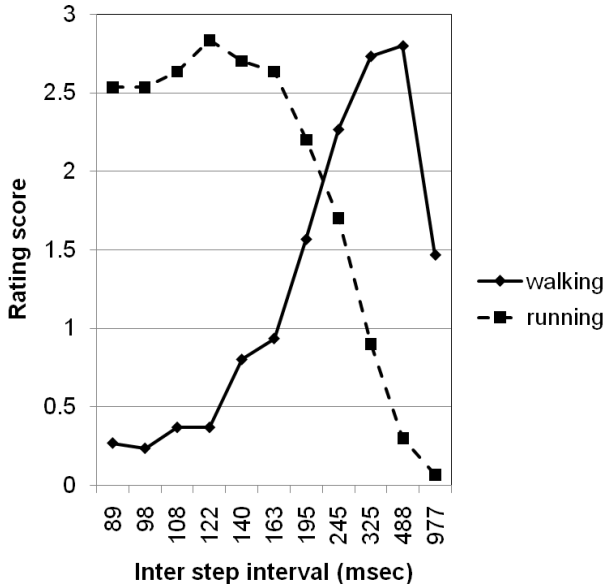

Figure 6. Results of experiment 2 for the step duration of 169 msec.

3 EXMERIMENT 2

In the experiment 2, the subjects evaluated a footstep sound whether they are heard as walking or running, varying inter-step interval with fixed step duration to examine the interaction effect between inter-step interval and step duration for the reality of walking or running sound.

3.1 Method

Stimuli: The footstep sounds for the experiment 2 was synthesized from single step sound (recorded for the experiment1) alternating with silence (inter-step interval). The duration of the single footstep sound was either 110 msec, 258 msec or 169 msec, which correspond to the condition derived highest rating score of walking or running cognition in the experiment 1 or geometric mean of them ("center" value between walking and running). The durations of the silence (inter-step interval) were the same as that of experiment 1 (Table 1.)s.

Procedure: The 11 variations of the inter-step interval (silence duration between steps) of footstep sounds for each 3 conditions of step duration were randomly presented to the subjects via headphone (Sennheiser, HDA200). In one trial, after 5-sec presentation of the sound of a condition, the subjects were asked to judge the reality of the sound with the rating scores of 0 to 3 (4 alternatives), which means 0 as "not walking (running)", 1 as "not so walking (running)", 2 as " probably walking (running)", 3 as "absolutely walking (running)". The judgment of the subjects was performed in terms of both the reality as "walking" sound and the reality as "running" sound for all 11 conditions. The judgment was iterated 5 times for each conditions and points of view of the judgment.

Subjects: 10 graduate students of Miyagi university with ordinary hearing participated in experiment.

3.2 Results and Discussion

The averaged rating scores between subjects was shown at Figure 4 (for step duration of 110 msec), Figure 5 (for step duration of 258 msec) and Figure 6 (for step duration of 169 msec). The profiles of rating score were different between figures, so an interaction between step duration and inter-step interval was obtained and difference of the scores between figures was also indicating the contribution of inter-step interval to the discrimination between "walking" perception and "running" cognition and fixed duration of footstep can be contributed to the reality for both "walking" and "running" cognition through wide range of inter-step interval.

4 EXPERIMENT 3

In the experiment 3, the subjects evaluated the perceived number of persons of crowd footstep sounds as a function of physical number of persons of footsteps to examine whether the perceived number of person increase monotonically as a function of physical number. If so, crowd footstep sounds can be synthesized by simply stacking single step sound. But if not, crowd footstep sounds can not be synthesized with "bottom up" manner, and there are some kinds of direct representation for crowd footstep sounds itself.

4.1 Method

Stimuli: The crowd footstep sounds were synthesized from recorded single footstep sound as material. First, footstep sound for one person was synthesized with most typical combination of step duration and inter step interval (silence duration) which can be strongly perceived as "walking" in experiment 2, 169 msec and 488 msec. Then, the footstep sound for one person was copied to 20 tracks (persons) of multi-track audio editor (Audacity) and those footstep sounds in each tracks was varied in tempo and pitch each other.

Procedure: The crowd footstep sounds which consist of from 1 to 10 persons of footsteps were randomly presented to the subjects via headphone (Sennheiser, HDA200). In one trial, after 5-sec presentation of the sound of a condition, the subjects were asked to respond perceived number of persons of footstep sounds directly. The judgment was iterated 5 times for each conditions.

Subjects: 10 graduate students of Miyagi university with ordinary hearing participated in experiment.

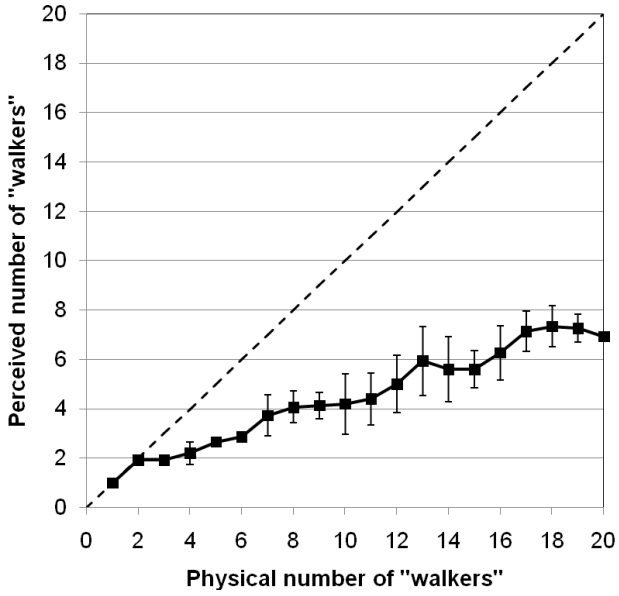

Figure 7. Results of experiment 3. The error bars indicate standard deviations.

4.2 Results and Discussion

The averaged perceived number of persons of footstep sounds between subjects was shown at Figure 7, indicating the perceived number of persons of footstep sounds does not increase monotonically as a function of physical number of that and underestimated. The same kind of underestimation of the perceived number of talkers from simultaneous utterances of speech sounds was reported in previous study [1]. The results suggest that there are some kind of a direct (or a static) representation for crowd footstep sounds itself, such as the skewness of the luminance histogram as the direct representation of the glossiness of an object surface in vision [2].

5 CONCLUSION

From the cognitive characteristics point of view, it can be concluded that footstep sound can be synthesized by merely lining recorded single step sound (element sound) alternating with appropriate duration of silence (inter-step interval). Without any other audio effect (e.g. randomizing amplitude or duration, etc.), simply lined single step can be heard as "walking" or "running".

The discrimination between "walking" perception and "running" cognition was mainly depend on inter-step interval (silence duration). There was interaction between step duration and inter-step interval, and relatively long step duration can be applied to wider range of inter-step interval.

The crowd footstep sounds could not be synthesized merely by stacking footstep sound for one person. The stack of footstep sound with varying tempo and pitch between sound (between audio track in multi-track sound editor) up to for 10 person could not monotonically increase perceived number of person of crowd, and saturated at 3 or 4 persons. The "underestimation" of stacking footstep sounds suggests that here are some kind of direct representation for crowd footstep sounds itself.

In fact, sound effect engineer's empirical technique to synthesize "Gaya" (crowd sound effect) is known that effective crowd footstep sounds consist of direct crowd sound (like pink noise) as background and a few footstep sound as foreground. Thus, this study could explain those empirical techniques.

REFERENCES

[1] M. Kashino and T. Hirahara. Judging the number of concurrent talkers for sentence stimuli. Proceedings of Spring Meeting of Acoustical Society of Japan 1997, 369-370, March 1997 [in Japanese]

[2] I. Motoyoshi, et al. Image statistics and the perception of surface qualities. Nature, 447:206-209, May 2007