Light-synchronized tapping task as an objective method for estimating auditory detection threshold

Shigeto Furukawa1,*, Kazuki Onikura1,2,†, Shunsuke Kidani1,‡, Masaharu Kato1,§ and Norimichi Kitagawa1

1NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3–1 Morinosato-Wakamiya, Atsugi, 243–0198 Japan
2Graduate School of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, 819–0395 Japan

(Received 10 May 2017, Accepted for publication 22 August 2017)

Abstract: The accuracy of synchronous tapping with a sequence of fixed-interval light flashes (light-synchronized tapping, LST) is impaired by the presence of a sound sequence depending on its temporal relationship with the light flashes. The present study tested the possibility that the LST task can be used as an objective method of estimating auditory detection thresholds without requiring the listener to report directly his/her sensation as in standard audiometry. The experiment used tone bursts as distractor sequences and varied the frequency and level of the tones. The tone level had a statistically significant effect on the distraction level, but the effect of frequency was not significant. Significant distraction was observed for a tone level of, on average, as low as 15 dB above the detection threshold. In other words, once the lowest tone level of the distraction effect is identified, one can expect that the participant’s detection threshold would lie at around 15 dB below the level, regardless of the frequency. The results indicate that in principle, the LST task could be used to estimate auditory detection thresholds, although the reliability of the threshold estimation still has to be improved for it to be applicable to audiometry.

Keywords: Synchronized tapping, Audiometry, Detection threshold, Objective method

PACS number: 43.66.Cb, 43.66.Sr [doi:10.1250/ast.39.30]

1. INTRODUCTION

In standard audiometry, the audibility of a signal is assessed on the basis of the listener’s subjective report on whether or not they can hear the signal. Detection thresholds estimated in this way would be inaccurate when the listener intentionally (or not) makes reports that are inconsistent with his/her perception (e.g., malingering; functional hearing loss), which is a welfare policy issue in Japan. Physiological measures, such as oto-acoustic emissions and auditory brainstem responses can indicate (mal-)functioning of particular physiological mechanisms, but they do not directly represent subjective audibility. The results of behavioral methods based on the subject’s spontaneous utterances, such as the Lombard test and delayed auditory feedback test, are difficult to use to quantify the audibility or make comparisons with a standard audiogram [1]. Pure-tone delayed auditory feedback (PDAF) is considered to be a reliable quantitative, behavior-based method: the subject is requested to finger-tap repeatedly with a particular temporal pattern. The tapping behavior is disturbed by the introduction of tones with the same pattern but with a relative delay to the tapping. The disturbances are known to occur even for barely audible tones (5 dB above the threshold) and thus could be a good indicator of the detection threshold [2]. This principle not only enables objective estimation of detection threshold, but has a critical advantage over the standard audiometric method, particularly in situations in which malingering is suspected. It is difficult to perform the task correctly when the sound is audible. Pretending that the sound is inaudible would be difficult because the subject would have to perform the task better to demonstrate that it indeed is.

Here, we propose a new tapping-based behavioral method for objectively estimating the detection threshold. The method is based on a light-synchronized tapping (LST)
task, in which a subject is instructed to finger-tap synchronously with a regular-interval sequence of flashing light. As in the PDAF task, the presentation of asynchronous audible sounds distracts the subject (see, e.g., [3]). The subject’s performance can be evaluated as the amount of deviation in reference to the light timing. This enables a quantitative evaluation of the subject’s behavior that is easier than in PDAF, which is based on the subject tapping at his or her own pace.

The specific aim of this study was to determine the lowest sound level at which the sound induces the distraction effect in the LST task. The results were evaluated mainly in terms of the applicability of the LST task to audiometry. Repp and Penel [3] observed an effect even for “soft” sounds (i.e., low in sound pressure level), but did not specify the levels of those sounds. We used the task paradigm of Repp and Penel [3], in which the tap timing relative to the flash timing is measured while the sound timing relative to the flash timing is systematically varied. Tone bursts were used as the stimuli, and the frequency of the tone was 250, 1,000, or 4,000 Hz. These tone stimuli are most frequently tested in audiometry.

2. MATERIALS AND METHODS

2.1. Participants

Six, right-handed adults (four females and two males; 28–42 years old) participated in the experiment. None has a history of diagnosed hearing deficit. All participants were informed of the aims and risks of the experiment and gave written informed consent. The experiments were designed and performed with the approval of the local ethics committee of NTT Communication Science Laboratories.

2.2. Stimuli

A device manufactured by Tucker-Davis Technologies (RM1 module) was used to generate visual and auditory stimuli and to record the timing of the stimuli and tapping. The visual stimuli were flashes, presented with a red light-emitting diode (LED). The flashes had a 100-ms duration and were presented in 655-ms intervals. The auditory stimuli were tone bursts, which were digitally synthesized with a sampling rate of 24 kHz and a resolution of 24 bits, digital-to-analogue converted, and presented to the right ear through a calibrated earphone. Similarly to the visual flashes, the tone bursts each had a 100-ms duration (including 10-ms raised-cosine ramps) and were presented in 655-ms intervals. Note that the flash and tone-burst intervals were intended to be 640 ms when the experiment was designed, but post-hoc examinations of the stimulus records revealed that they were actually 655 ms due to the system’s additional processing time. The tone bursts were varied in frequency (250, 1,000, or 4,000 Hz), sensation level (sound pressure level ranging from −15 to 45 dB in 15 dB steps relative to the detection threshold; see the next section), and onset time (sound delay relative to the flash onset; 0, 80, 160, 240, 320, 400, 480, and 560 ms). The conditions with a sensation level of −15 dB were regarded as control conditions in which the sound was considered to be inaudible.

2.3. Procedures

The experiments were conducted in a sound-proof room. Prior to the formal experiments concerned with tapping, the detection threshold for each participant and tone frequency was measured using standard psychophysical methods, i.e., a two-alternative forced-choice method and the transformed up-down method (three-down one-up procedure; [4]). One run of the adaptive tracking procedure in the transformed up-down method was terminated after 12 reversals. The step size was 5 dB for the first four reversals and 2 dB after that. The detection threshold was defined as the average sound pressure level of the last eight reversals. The obtained thresholds are summarized in Table 1. In the rest of the paper, sound levels are represented as sensation levels, i.e., relative level to those thresholds.

A sequence of stimuli (flashes and tone-bursts) and obtained responses is illustrated in Fig. 1. One trial of the experiment consisted of a sequence of 32 presentations of flashes and tone-bursts. The parameters of the auditory stimuli (i.e., frequency, level, and sound delay) were fixed within a trial. Three trials were run per subject per parameter set. The participants were instructed to tap with their right index finger synchronously with the light flashes, while ignoring the auditory stimuli. The tapping responses were obtained with a touch-sensitive capacitance switch located close to the LED, and their timings were recorded.

2.4. Data Analyses

We defined the tap delay and sound delay as the delays in the tap and sound-onset timings, relative to the onset of the nearest flash in time. The tap timing data for the first four taps within each trial were excluded from the analyses. Missing data (due to the participant’s miss-tapping or the failure of the touch-sensitive switch to indicate a response)
were ignored. Due to the cyclic nature of the stimuli and responses, sound and tap delays were expressed as phases relative to the flash period. A tap delay of $-0.25$, for example, indicates that the tap was made at $655 \times 0.25$ ms before (or equivalently $655 \times 0.75$ ms after) the flash.

For the same reason, circular statistics [5] were used to represent the central tendency and dispersion of the tap timings, namely the circular mean and the circular standard deviation, respectively. Individual tap delays were expressed as a unit vector with the angle corresponding to its phase. The phase corresponding to the angle of the mean vector is defined as a circular mean. The circular standard deviation is computed as $\sqrt{-2 \log R}$, where $R$ is the resultant length of the vector averaging. Hereafter, the circular mean and standard deviation of the tap delays within a trial are referred to as the “mean asynchrony” and “s.d. of asynchronies,” respectively.

3. RESULTS

3.1. Mean Asynchrony

Figure 2 plots mean asynchrony as a function of sound delay (referred to as the sound-delay function) for the 1,000-Hz tone. Each panel represents the results of one participant and one tone level. Each circle within a panel represents one trial. It should be noted that the mean asynchronies (which were circular) and the ordinate were adjusted so that the tap delays for the 0 sound-delay appear around the middle of the ordinate range. The thin solid lines are fifth order polynomials fitted to the data points after this adjustment. The fitting was made to indicate the general trend [3]. Although there were appreciable inter-individual variabilities in reproducibility and in the patterns of sound-delay effect, we can see a general trend in which SEI sine increased with sound level. The functions for the three tone frequencies generally overlapped with a few exceptions. A within-subject analysis of variance with factors stimulus frequency and sensation level indicated the main effect of sensation level ($p < 0.0001$; $F(4, 20) = 19.26$) but no effect of frequency ($p = 0.103$; $F(2, 10) = 2.884$) or interaction of sensation level and frequency ($p = 0.068$; $F(8, 40) = 2.025$). A post-hoc comparison of mean SEI sine for each sensation level (except the $-15$ dB) with that for the $-15$-dB sensation level revealed a significant difference for the sensation levels at or above 15 dB ($p < 0.01$; Tukey’s honestly significant difference test).

It can be argued that a single common function cannot fully capture the effects of sound delays on individual participants. As observed in an earlier study [3], detailed patterns of the sound delay functions varied appreciably among participants, and there were cases in which the fitted sine function deviated from the polynomial fits. Therefore, we examined an alternative index to quantify the effect of sound delay on mean asynchrony. Namely, we computed
First, for each tone frequency, subject, and sound level, the mean asynchronies for each sound delay were circular-averaged across trials. Then, we calculated the circular standard deviation of the averaged mean asynchronies across sound delays. The circular standard deviation obtained in this way is referred to as SEI\textsubscript{cstd}. A larger value of SEI\textsubscript{cstd} indicates a greater dependence of the tap timing on the sound delay. It should be noted that SEI\textsubscript{cstd} is hypothesis-free, i.e., not sensitive to whether the sound-delay dependent variation is systematic (e.g., sinusoidal) or not.

The middle row of the panels in Fig. 4 represent SEI\textsubscript{cstd} as a function of sensation level. Each panel represents one subject, and the results for the three tone-frequencies are represented by different symbols within each panel. The results for SEI\textsubscript{cstd} were generally consistent with those for SEI\textsubscript{sine}: again, we found a general trend in which SEI\textsubscript{cstd} increased with sound level. The SEI\textsubscript{cstd} functions for the three tone frequencies substantially overlapped. This indicates that the sensitivity of the SEI\textsubscript{cstd} to the sound level is invariant across frequencies, given the level was adjusted in reference to the detection threshold or audibility. A within-subjects analysis of variance with factors stimulus frequency and sensation level indicated the main effect of sensation level ($p < 0.0001$; $F(4, 20) = 13.57$) but no effect of frequency ($p = 0.666$; $F(2, 10) = 0.423$) or interaction of sensation level and frequency ($p = 0.440$; $F(8, 40) = 1.02$). A post-hoc comparison of mean SEI\textsubscript{cstd} for each sensation level (except the $-15\text{dB}$) with that for the $-15\text{dB}$ sensation level revealed a significant difference for the sensation levels at or above $15\text{dB}$ ($p < 0.01$; Tukey’s honestly significant difference test).

### 3.2. S.D. of Asynchrony

Figure 3 represents the s.d.’s of asynchronies (the circular standard deviation of tap delay within each trial; see Materials and Methods). Smaller values indicate more stable entrainment to flash or tone-burst intervals. Although the range and pattern of the s.d.’s of asynchronies varied markedly among subjects, there was a general tendency that the s.d. of asynchronies at around zero sound delay was smaller for higher sensation levels than for lower ones. This tendency can be most clearly seen in subject #6. We linear-averaged across trials the s.d. of asynchronies at around zero sound delay, and denoted it as SEI\textsubscript{entrain}. The s.d. of asynchronies for a large absolute sound delay exhibited a sensation-level dependence that was inconsistent among the subjects.
The bottom panels of Fig. 4 plot SEI\textsubscript{entrain} as a function of sensation level. Each point represents the average across multiple trials for the same condition. The general trend mentioned above appears: SEI\textsubscript{entrain} tended to decrease with increasing sensation level. The functions for the three different frequencies substantially overlapped.

**Fig. 3** The s.d. of asynchronies as a function of sound delay. The results for the 1,000-Hz tones are shown. The meanings of the other symbols and lines are the same as in Fig. 2.

**Fig. 4** Indices of the sound delay effect as a function of sensation level. Symbols represent tone frequencies as indicated in the key. Filled symbols indicate the values at the sensation level were statistically significantly different from those at the $-15$-dB sensation level.

The bottom panels of Fig. 4 plot SEI\textsubscript{entrain} as a function of sensation level. Each point represents the average across multiple trials for the same condition.
A within-subject analysis of variance with factors stimulus frequency and sensation level indicated the main effect of sensation level ($p < 0.006$; $F(4, 20) = 4.98$) but no effect of frequency ($p = 0.440$; $F(2, 10) = 0.892$) or interaction of sensation level and frequency ($p = 0.991$; $F(8, 40) = 1.89$). A post-hoc comparison of mean SEI_{entrain} for each sensation level (except the $-15$ dB) with that for the $-15$ dB sensation level, however, did not exhibit a significant difference at any sensation level ($p > 0.05$; Tukey’s honestly significant difference test).

3.3. LST-based “Thresholds” of Individual Participants

The above analyses concerned general trends among participants. When audiology is the intended application, it is important to evaluate the extent to which the detection thresholds or audibilities of individual participants can be estimated on the basis of the LST. Thus, we sought the minimum sensation level required to detect the sound effect on the tapping behaviors of individual subjects. A sensation level found in this way can be regarded as an estimate of the “audibility threshold” based on the LST task.

As described earlier, the goodness of fit of a sine function is one index of the sound-delay effect. The filled symbols in the top panels of Fig. 4 (SEI_{sine} as function of sensation level) indicate that the regressions were statistically significant ($p < 0.01$; uncorrected). A bootstrap method was used to make a statistical analysis of SEI_{cstd}: the mean asynchrony at each sound delay was calculated on the basis of resamples of trials with replacement (the number of resamples was the same as the original samples), and SEI_{cstd} was derived from the resampled data. This operation was repeated 10,000 times to generate bootstrap samples of the SEI_{cstd}. The filled symbols in the top panels of Fig. 4 indicate that the SEI_{cstd} at that sensation level was statistically significant (one tailed; $p < 0.05$, Bonferroni-corrected) greater than that at the sensation level of $-15$ dB. A similar bootstrap analysis was applied to SEI_{entrain}. Sensation levels at which the values were significantly smaller than that at the $-15$ dB sensation level (one tailed; $p < 0.05$, Bonferroni-corrected) were found and are marked as filled symbols in the lower panels of Fig. 4. Table 2 summarizes the lowest sensation levels with significant SEI_{sine}, SEI_{cstd} or SEI_{entrain} (for SEI_{sine}, the results for the $15$ dB sensation level are excluded). The symbol “-” indicates that there was no sensation level with a significant difference under that condition.

### 4. DISCUSSION

The present results replicated the earlier findings that the performance of the LST task is influenced by the timing of simultaneously-presented (audible) auditory stimuli. Although there were marked inter-individual variabilities in reproducibility, the patterns of sound-delay effect were generally consistent with those reported in the earlier study [3]: Around a sound delay of zero, advancing (delaying) the sound timing relative to the flash resulted in advanced (delayed) tap timing. The presence of sound sequences synchronized to flashes tended to improve flash rhythm entrainments (appears as a decrease in s.d. of asynchronies). The present and earlier studies are also consistent in that the tap timings were often in advance of the flashes (and the sounds when the sounds were synchronized to the flashes), as indicated by negative values of mean asynchrony, which is known as anticipation tendency [6].

The present study has two major new findings. First, the sound-delay effects are observable for sound levels, on average, at or above $15$ dB relative to the detection threshold. Most of the “audibility thresholds” in SEI_{sine} and SEI_{cstd} of the individual participants (Table 2), when successfully obtained, were $15$ dB in sensation level. Second, the pattern of the level dependence was generally invariant among the tone frequencies when the level is expressed as the sensation level.

These are preferable features for audiology: a sensation level of $15$ dB is relatively close to the detection threshold. When the sound-delay effect is observed for a given sound, one can expect that the participant’s “true” detection threshold should be at or below $-15$ dB relative to that sound level. This feature is particularly useful for detecting malingering or functional hearing loss. No significant interaction of tone frequency and sensation

---

### Table 2

| Freq. (Hz) | #1  | #2      | #3  | #4  | #5  | #6      |
|-----------|-----|---------|-----|-----|-----|---------|
|           |     |         |     |     |     |         |
| 250       | 15/-/-/-  | 15/15/30 | 15/15/-/-  | 15/15/30 | 15/-/-/-  | 30/-/-/-/45 |
| 1,000     | -/-/-/-/15 | 15/15/-/-  | 15/15/-/-  | -/-/-/-/-  | 30/30/-/-  | 45/-/-/-/15  |
| 4,000     | 15/-/-/-  | 0/15/-/-  | 0/0/15 | 15/15/0 | 15/15/45 | 15/-/-/-/15  |
level implies that the effect of tone sensation level is comparable between different frequencies. Thus, one could evaluate the plot of the LST-based “audibility threshold” as a function of frequency, in the same manner as a standard audiogram, which also plots the detection threshold as a function of frequency.

In order for the LST task to be applied to audiometry, it is necessary to ensure the reliability of the threshold estimation for a variety of listeners. In any of the three indices (SEI<sub>sine</sub>, SEI<sub>cstd</sub>, and SEI<sub>entrain</sub>), there were cases in which we failed to identify an “audible threshold.” SEI<sub>sine</sub> appears to be the best of the three indices in terms of the number of successfully identified cases (although caution is needed here because the method for evaluating statistical significance varies among the indices) but was susceptible to deviation of the pattern of the sound-delay dependence from a sine function (e.g., participants #1 and #4). This susceptibility is also a cause of occasional misses of a significant sound-delay effect at relatively high sound levels (e.g., the function for 250-Hz of participant #4, and those for 4,000 Hz of participants #5 and #6; data not shown). SEI<sub>cstd</sub> and SEI<sub>entrain</sub> are hypothesis-free measures that reflect dependence on sensation level with good consistency across tone frequencies. However, they appeared to have weaker sensitivities (fewer cases of significant sound-delay effects even at relatively high levels).

Increasing the number of trials is a possible solution to the problem. In the current experimental design, the net time required to test the sound-delay effect (i.e., to plot a single point in Fig. 4) was about 8 min. (640-ms flash intervals × 32 flashes × 3 trials × 8 sound delays). Spending any longer for measurements would be impractical in clinical settings. Nevertheless, we hope that the measurement time can be reduced to a reasonable duration while increasing the number of trials, by optimizing experimental parameters (e.g., sound-delays to examine) or by choosing a more efficient task [7].

Combining more than one index of sound-delay effect may also help to improve reliability. This possibility is anticipated by cases in which a failure to obtain a threshold in one index was sometimes complemented by success in another index.

ACKNOWLEDGMENTS

This work was conducted as part of a research project commissioned by the Ministry of Internal Affairs and Communications, Japan (MIC) under the Strategic Information and Communications R&D Promotion Programme (SCOPE; 121803022), and was funded in part by internal research funds of NTT Corporation. Portions of this study were presented at the 38th Mid-Winter Meeting of the Association for Research in Otolaryngology, Baltimore, Maryland, February, 2015, and at the 2015 Spring Meeting of the Acoustical Society of Japan, Tokyo, Japan, March 16, 2015. There are no conflicts of interest.

REFERENCES

[1] F. N. Martin and J. G. Clark, “Nonorganic hearing loss,” in Handbook of Clinical Audiology, J. Kats, M. Chasin, K. English, L. J. Hood and K. L. Tillery, Eds. (Wolters Kluwer Health, Philadelphia, 2014), pp. 617–629.
[2] H. B. Ruhm and W. A. J. Cooper, “Delayed feedback audiometry,” J. Speech Hear. Disord., 29, 448–455 (1964).
[3] B. H. Repp and A. Penel, “Rhythmic movement is attracted more strongly to auditory than to visual rhythms,” Psychol. Res., 68, 252–270 (2004).
[4] H. Levitt, “Transformed up-down methods in psychoacoustics,” J. Acoust. Soc. Am., 49, 467–477 (1970).
[5] K. V. Mardia and P. E. Jupp, Directional Statistics (John Wiley & Sons, Chichester, 2000).
[6] G. Aschersleben, “Temporal control of movements in sensorimotor synchronization,” Brain Cogn., 48, 66–79 (2002).
[7] B. H. Repp, “Sensorimotor synchronization: A review of the tapping literature,” Psychon. Bull. Rev., 12, 969–992 (2005).