

The influence of psychophysical procedure and stimulus type on estimates of human performance in detecting audio-visual asynchrony

Rob L.J. van Eijk¹
R.L.J.v.Eijk@tm.tue.nl

Armin Kohlrausch^{1,2}
Armin.Kohlrausch@philips.com

Steven van de Par²
steven.van.de.par@philips.com

James F. Juola^{1,3}
juola@ku.edu

¹Eindhoven University of Technology, ²Philips Research Laboratories, Eindhoven, and ³University of Kansas

Abstract

Human perception of audio-visual synchrony is typically characterized by two quantities: the point of subjective equality (PSE) and human sensitivity to asynchrony. The PSE can be derived from the 50% point in a temporal order judgment (TOJ) procedure or from the average of both intersections of the synchronous response curve with the non-synchronous response curve(s) in a synchronous-successive response paradigm. Sensitivity is derived from the steepness of the response curve in the TOJ paradigm. In this contribution we present data that show that PSE estimates derived from TOJ measurements are much more variable across observers than those based on synchronous-successive data. Two synchronous-successive methods were used with different response rules: (1) 2 alternatives: asynchronous, synchronous, and (2) 3 alternatives: audio first, synchronous, video first. Both synchronous-successive methods yielded similar results. Two stimuli were used: (1) a flash-click stimulus, and (2) a simple animation of a bouncing ball. PSE estimates derived from synchronous-successive data were larger for the bouncing ball (impact) stimulus. Furthermore, we analysed discriminability values obtained with a 2-alternative forced-choice procedure. Discriminability is better near the edge than in the middle of the synchronous response category. This suggests that categorical perception might play a role in audio-visual synchrony perception.

Results: synchrony judgment

Using psignifit [2] a cumulative normal distribution was used to fit individual TOJ data (video first response proportions) and the left and right halves of the individual synchronous response curves obtained from SJ procedures (see Table 1). All analyses are based on fitted parameters.

The PSE is defined as the middle of the synchrony range (i.e. the audio delay range between the two intersection points of the synchrony curve with its neighbouring curves).

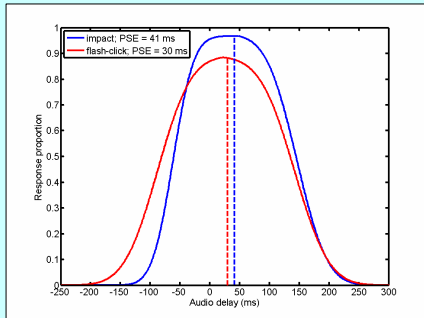


Figure 2: Average synchrony curves for impact and flash-click stimuli. The curves show the average probability of a 'synchronous' response as a function of the audio delay (negative audio delays indicate audio preceding video).

Table 2: Mean PSEs for impact and flash-click stimuli for different response rules. Standard errors are between brackets.

	SJ2	SJ3	TOJ
Impact	45 (7)	37 (6)	8 (13)
Flash-click	31 (8)	29 (8)	33 (11)

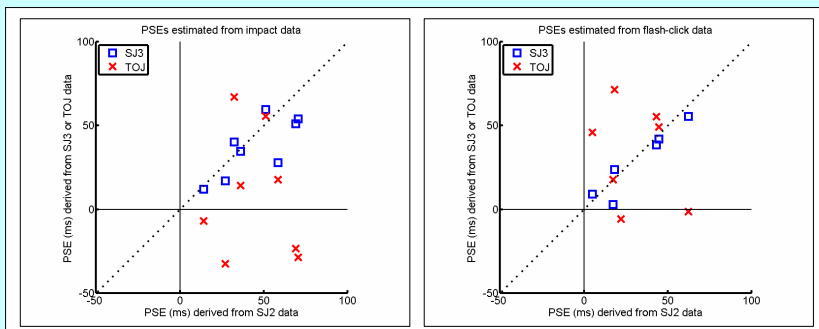


Figure 3: Scatter plot of PSEs for impact stimuli (left) and flash-click stimuli (right) estimated using different response rules. The dotted line indicates identical performance across response rules. In line with earlier results [1], we also sometimes observe negative TOJ PSEs, whereas SJ PSEs are always positive and less variable. Pearson's correlations for impact stimuli are 0.76 between SJ2 and SJ3, -0.16 (n.s) between SJ2 and TOJ, and 0.28 (n.s) between SJ3 and TOJ PSE pairs. Pearson's correlations for flash-click stimuli are 0.94 between SJ2 and SJ3, -0.24 (n.s) between SJ2 and TOJ, and -0.20 (n.s) between SJ3 and TOJ PSE pairs.

Experimental methods

Two stimulus types were used (impact & flash-click), which were both used in two different types of experiment (synchrony judgment & duration discrimination thresholds for audio-visual delay) resulting in a total of four experiments.

Impact stimulus

This stimulus consisted of a white ball falling down toward a white bar and rebounding from it (see Figure 1) accompanied by an impact sound. The stimulus lasted for 2 seconds including two randomly determined periods of 0-500 ms at the beginning and end of the stimulus during which the ball was not moving (i.e. "floating").

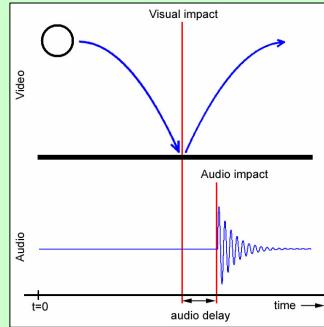


Figure 1: Impact stimulus. Please note that the horizontal axis is a time axis; the animation thus consisted of a vertically moving ball with a constant acceleration. This example has a positive audio delay (video leads audio).

Flash-click stimulus

This stimulus consisted of white circle (identical to the ball of the impact stimulus) shown during 1 frame (12 ms) accompanied by a 'click' (noise burst of 12 ms).

Results: discrimination thresholds

For each audio-visual reference delay two thresholds were measured: one in which audio is further delayed relative to the reference delay (called positive threshold), and one in which audio is further advanced relative to the reference delay (negative threshold). These positive (green bars) and negative (red bars) thresholds are indicated in Fig. 4.

The three reference delays are chosen to be in the left half, middle, and right half of the synchrony curve (see Fig. 2).

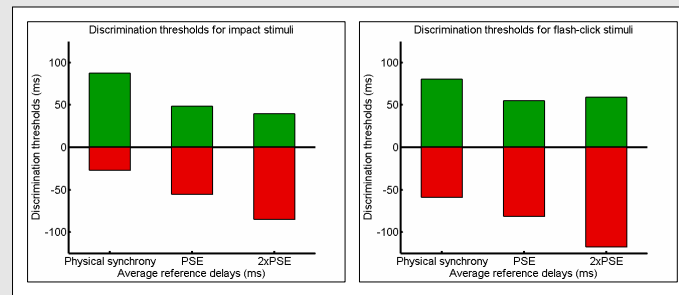


Figure 4: Discrimination thresholds for three reference delays for impact (left) and flash-click (right) stimuli. The bars indicate the threshold delay relative to the reference delay. Green bars indicate positive thresholds (audio further delayed) and red bars indicate negative thresholds (audio further advanced).

Synchrony judgment

Subjects (n = 8 for impact; n = 7 for flash-click, all of whom were also participants in the impact study) judged the (a-)synchrony of the audio-visual stimulus using 3 different response rules (in different sessions) as shown in Table 1. Stimuli were presented using the method of constant stimuli and contained audio delays varying from -350 ms (audio leading) to +350 ms with steps of 50 ms. Each subject judged each audio delay 60 times.

Table 1: Response rules with the corresponding response alternatives used in the synchrony judgment experiment.

Response rule	Response alternatives
Temporal Order Judgment (TOJ)	Audio first, video first
Synchrony Judgment with 2 alternatives (SJ2)	Asynchronous, synchronous
Synchrony Judgment with 3 alternatives (SJ3)	Audio first, synchronous, video first

Discrimination thresholds

Subjects (n = 7; all participants were also in the synchrony judgment experiments) were presented with groups of three animations (3 intervals). The first interval (the reference) had an audio-visual delay that was constant during a measurement, while either the second or third animation had a different audio-visual delay. Subjects had to indicate the deviant animation (the target). The audio-visual target delay was adjusted using a 1-up, 2-down adaptive procedure, resulting in a threshold indicating the 71%-correct point. Thresholds were determined by taking the median of the last eight reversals.

Reference delays were individually determined from synchrony judgment data and equalled physical synchrony (0 ms), PSE, and 2xPSE.

Thresholds are (approximately) symmetric at the PSE reference delay, but strongly asymmetric at physical synchrony, and 2xPSE. Thus, thresholds are lower near the boundaries of the synchronous range, and larger within it.

This procedure sheds some doubt on the widely accepted procedure of using the steepness of the TOJ response curve (just-noticeable difference; JND) as an optimal measure of sensitivity to changes in audio-visual delay.

Conclusion

PSEs are much more influenced by response rule (i.e. SJ2 or SJ3 vs. TOJ) than by stimulus type.

Thresholds for discriminating between audio-visual delays are smaller near synchrony boundaries.

Acknowledgements

Thanks to Martin Boschman, Constant Hak, and Christophe Stoelinga for assistance.

References

- [1] van de Par, S., Kohlrausch, A., and Juola, J. F. (2002). Some methodological aspects for measuring asynchrony detection in audio-visual stimuli. In Calvo-Manzano, A., Perez-Lopez, A., and Santiago, J. S., editors, *Proceedings of Forum Acusticum 2002 (CD)*, Sevilla.
- [2] Wichmann, F. A. & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception and Psychophysics*, 63(8), 1293-1313.