Exp Brain Res (2012) 221:403–412
DOI 10.1007/s00221-012-3182-9

RESEARCH ARTICLE

Effects of stimulus duration on audio-visual synchrony perception

I. A. Kuling · R. L. J. van Eijk · J. F. Juola ·
A. Kohlrausch

Received: 9 August 2011 / Accepted: 3 July 2012 / Published online: 22 July 2012

Abstract The integration of visual and auditory inputs in the human brain occurs only if the components are perceived in temporal proximity, that is, when the intermodal time difference falls within the so-called subjective synchrony range. We used the midpoint of this range to estimate the point of subjective simultaneity (PSS). We measured the PSS for audio-visual (AV) stimuli in a synchrony judgment task, in which subjects had to judge a given AV stimulus using three response categories (audio first, synchronous, video first). The relevant stimulus manipulation was the duration of the auditory and visual components. Results for unimodal auditory and visual stimuli have shown that the perceived onset shifts to relatively later positions with increasing stimulus duration. These unimodal shifts should be reflected in changing PSS values, when AV stimuli with different durations of the auditory and visual components are used. The results for 17 subjects showed indeed a significant shift of the PSS for different duration combinations of the stimulus components. Because the shifts were approximately equal for duration changes in either of the components, no net shift of the PSS was observed as long as the durations of the two components were equal. This result indicates the need to appropriately account for unimodal timing effects when quantifying intermodal synchrony perception.

I. A. Kuling · R. L. J. van Eijk · J. F. Juola · A. Kohlrausch (✉)
Eindhoven University of Technology, IPO 1.25, P.O. Box 513,
5600 MB Eindhoven, The Netherlands
e-mail: a.kohlrausch@tue.nl

A. Kohlrausch
Philips Research Europe, HTC 36, 5656 AE Eindhoven,
The Netherlands

Keywords Multisensory perception · Stimulus duration · Synchrony judgment

Introduction

The onset of a sound is important for its perceptual characteristics. For example, in speech, when the original onset consonant of a consonant-vowel syllable is removed, it is hard to identify the vowel correctly (e.g., Strange and Bohn 1998). When analyzing onsets of sounds, one needs to consider that the relation between physical and perceived onsets can vary considerably between stimuli; for example, speech stimuli that have a regular temporal distance of their *physical* onsets are not necessarily perceived as isorhythmic (e.g., Morton et al. 1976; Marcus 1981). In psychoacoustic research, it has been found that the perceived onset of a stimulus is increasingly delayed relative to the physical onset when stimulus duration increases (Schütte 1978; Schimmel and Kohlrausch 2008).

For short auditory stimuli, there is no shift in perceived onset, which means that the perceived onset occurs at the same relative temporal position as the physical onset. For longer stimuli, the perceived onset shifts by about 20 ms relative to the physical onset as stimulus duration increases from 10 to 400 ms [see Schimmel and Kohlrausch 2008, Fig. 1 (replotted from Schütte 1978)]. In their own data, Schimmel and Kohlrausch (2008) found a shift of the perceived onset by about 30 ms for stimulus duration variations from 5 to 350 ms (see their Table II). In modeling the relation between physical and perceived onset, the latter is regarded as a function of the duration and the temporal envelope of the sound. According to these concepts, the moment of onset perception reflects the buildup of the internal excitation after stimulus onset and is

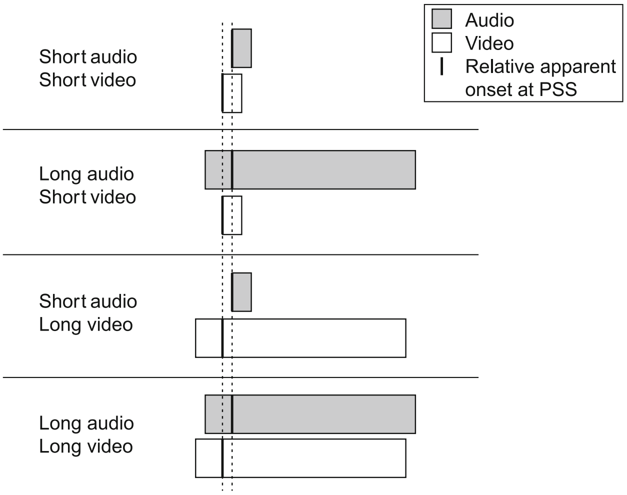


Fig. 1 Expected duration effects on PSS in an audio-visual synchrony judgment task. For two short stimuli of equal duration, a typical positive (video leading) PSS is expected, in which the midpoint of the subjective synchrony range corresponds to a later relative onset of the audio stimulus. When the duration of one of the components is increased, the relative apparent onset of this stimulus shifts to a later time; that is, the apparent onset occurs increasingly later than the physical onset. To keep the temporal distance between the perceived onsets constant, the PSS will shift toward more audio leading (more negative PSS values) for longer audio durations (2nd row), and in the opposite direction, more video leading (more positive PSS values) for longer video stimuli (3rd row). If the shifts are about equal for increasing durations in the two modalities, no change in the PSS should occur when both components have equal durations (compare *top* and *bottom rows*)

determined by the position after signal onset where the internal strength of the stimulus has reached a certain percentage of the maximum (e.g., Schütte 1978).

In the visual modality, the influence of duration differences on synchrony perception has been studied by Jaśkowski (1991). In both a temporal order judgment task (TOJ) and a synchrony judgment task (SJ), he found that there was a shift in perceived simultaneity of two visual flashes with different durations compared to equal-duration stimulus pairs. In order to be perceived in synchrony with longer stimuli, the onset of shorter stimuli had to be delayed by about 10–15 ms. Because the differences in overall duration were relatively small (for the majority of the data, stimulus durations were 110 and 150 ms), delaying the *onset* of the shorter stimulus also reduces the temporal distance between the two *offsets*. Jaśkowski concluded from the results that subjects judged two stimuli with different durations as synchronous when both onset and offset asynchronies were minimized, despite instructions to the subjects to focus only on the onsets. Brenner and Smeets (2010) performed a visual synchrony experiment, in which a 6-ms stimulus was compared with stimuli of up to 72-ms duration. The energy of each stimulus in a pair was kept constant; thus, the longer stimuli were presented with a lower luminance. Subjects were asked to adjust the temporal positions of the two stimuli such that they "appear to flash at the same time" (Brenner and Smeets 2010, pp. 1104). As the duration of the longer stimulus increased, the physical asynchrony of the adjusted temporal positions also increased. The shift was the same as observed by Jaśkowski: for perceptual synchrony, longer stimuli had to start relatively earlier than shorter ones. For the longest duration difference tested (6 vs. 72 ms), the measured onset asynchrony amounted to about 20 ms for their high-contrast stimuli. Thus, the data in the visual modality also indicate a systematic influence of stimulus duration on perceived synchrony, but this effect is interpreted in a different way than in the auditory modality, by referring to a combined effect of both onset and offset asynchronies.

In our study, we investigated the potential influence of stimulus duration on audio-visual synchrony judgments. In order to avoid difficulties in interpretation, we used stimuli of rather large duration differences in order to make it more likely that indeed the perceived onset of the individual components, and not also their offsets, is used in the subjects' judgments. Timing in multisensory events is a complex phenomenon based on multiple factors like temporal characteristics of the unimodal components, intersensory delays, and transmission differences in the different modalities (for a recent review see Vroomen and Keetels 2010). Therefore, it might not be straightforward to base predictions of stimulus duration effects in audio-visual synchrony perception on results from unimodal experiments.

As far as we know, Boenke et al. (2009) are the only researchers who have measured the effects of stimulus duration in an audio-visual synchrony perception task. In their experiment, participants made temporal order judgments for audio-visual stimuli with three different durations. The authors did not vary the duration of the auditory and visual components independently, but changed them together between 9 and 500 ms. The results indicated a clear shift in the point of subjective simultaneity (PSS) for most participants when the stimulus duration was increased, but the shift direction differed for different groups of subjects. The PSSs of the participants with the most positive (video has to be presented first) PSSs for the short stimuli moved to more negative values when the stimulus duration was increased, whereas the PSSs of the participants with the most negative (audio has to be presented first) PSSs became more positive when the stimulus duration increased. Across participants, this led to a decreased standard deviation (SD) of PSS estimates with increasing duration, but there was no effect of stimulus duration on the average PSS values.

The absence of an effect of stimulus duration on mean PSS estimates reported by Boenke et al. (2009) can be explained in two different ways. First, it is possible that unimodal variations in stimulus duration do not affect the percept of cross-modal synchrony in a systematic way. Alternatively, existing unimodal influences might be of equal magnitude and therefore lead to no net effect in a multimodal condition. The first explanation is unlikely for the reason that such a transfer of unimodal timing effects to multimodal timing perception has been shown for another stimulus parameter, the intensity. For example, Roufs (1963, 1974) compared different synchrony adjustment methods and found that in both the audio-visual 'eye and ear method' as in the purely visual 'double flash method', the perception lag decreased for increasing intensities of the visual stimuli. Also, Boenke et al. (2009) found a strong effect of intensity of the visual stimuli on audio-visual synchrony perception. From these results, the second explanation, in which the unimodal effects compensate for each other, seems to be more plausible.

To measure the effect of auditory and visual duration on perceived simultaneity, we designed an experiment in which both the auditory and the visual components were independently varied in duration between about 12 and 300 ms. A synchrony judgment task with three response categories (SJ3) was used, because this experimental task results in reduced variance of the PSS estimates between subjects when compared to those obtained in a TOJ task (van Eijk et al. 2008). Also, we have shown that the SJ task is less prone to individual differences and strategic processes that make it relatively more reliable and more suitable than the TOJ task for investigating new phenomena in bimodal perception (van Eijk et al. 2010).

Based on the results from the literature, we expected to find opposite effects of auditory and visual durations on the PSS estimates, and the relative strength of the duration-induced perceived onset shifts for auditory and visual stimuli should be reflected in duration-dependent PSS values for stimuli with equal-duration AV components (Fig. 1).

Methods

A simple bimodal stimulus was used, which consisted of a pair of gated auditory and visual stimuli. This stimulus type is commonly used to obtain perceptual temporal order and synchrony judgments as there are no context cues that could help participants predict or anticipate either stimulus (Sternberg and Knoll 1973; van Eijk et al. 2008).

Participants

Twenty-two participants took part (six females). Four of the participants were experienced in this research area (including the authors) and voluntarily joined the experiment. The other participants were naïve about the experiment and received a payment of 30 Euros. All participants reported (corrected-to-) normal vision and normal hearing. The participants varied in age from 19 to 68 years, with a mean of 33.5 years (SD = 16.5). The experiments conformed to the requirements of the World Medical Association as laid down in the Declaration from Helsinki 1964.

Stimuli

The visual part of the AV stimulus consisted of a white disk (97 cd/m^2 as measured using an LMT L1003 luminance meter) shown for one frame (12 ms), six frames (71 ms) or 25 frames (294 ms) at a central position on the screen. The disk had a diameter of 49 pixels and subtended an area of about 1.4° at an unconstrained viewing distance of about 60 cm. The presentation of the audio-visual stimuli happened within a 2-s period. This period was marked by the presence of four corners of a surrounding square, which also indicated the central location of the visual stimulus. The square was presented to give the participants spatial and temporal information about the upcoming flash and noise burst. The temporal onset of the flash was randomized, with the restriction that it occurred within the time window of 500–1,500 ms after the onset of the surrounding square. The acoustic part of the stimulus consisted of an 12, 71, or 294-ms white noise burst with a sound pressure level of 67 dB, which was presented identically to both ears (diotic presentation).

Apparatus

The visual stimulus was shown on a Dell D1025HE CRT monitor at a resolution of 1,024 × 768 pixels and at an 85-Hz refresh rate. The auditory stimulus was played through a Creative SB Live! sound card, a Fostex PH-50 headphone amplifier, and Sennheiser HD 265 linear headphones. Participants were seated in front of the monitor and responded using a keyboard. The setting was a dimly lit, sound-attenuated room (identical to van Eijk et al. 2008).

Design

We used a within-subjects design, in which all participants experienced all nine conditions produced by crossing all three duration values for both the flash and the noise burst. Each condition was divided into two runs with a short break in between (duration of the break was self-paced). Within each condition, the durations of the auditory and visual components of the stimuli were constant, but the relative stimulus onset asynchrony (SOA) between flash and noise burst was varied randomly. There were 15 SOA values, from −350 to 350 ms in steps of 50 ms, with negative values indicating audio-first and positive values indicating video-first presentations. An SOA of 0 ms is defined as physical synchrony (the Point of Objective Synchrony: POS).

Procedure

The participants received written instructions about their response options and the use of the keyboard. Participants had to judge the synchrony or asynchrony of the components of the audio-visual stimulus and responded by pressing a number on the numeric keypad on the keyboard. For all participants, the number "1" was used to indicate an audio-first decision, "2" was used for a synchronous responses, and "3" was used to indicate a visual-first judgment. The response was briefly shown on the screen, but without any feedback about the correctness of the response. First, the participants received a practice block, which consisted of 15 audio-visual stimuli with the same durations of flash and noise burst as in the upcoming trial run. In the practice block, each of the 15 audio-visual delays was presented once. After practice, the measurement trials were presented. Within each block of 225 trials (divided into two parts by a short break), all 15 delays were presented 15 times each, resulting in a total of 225 judgments per condition for each participant. Measuring one condition, a practice block + measurement block, lasted about 13–15 min, after which the participant could have a break, followed by a second condition, followed by another break, and then the third condition. The durations of the breaks were self-paced. Each session lasted about 45–55 min. The three sessions (each consisting of three conditions) were measured on different days within a 2-week period. The order in which the 9 conditions were presented was counterbalanced over all participants. After the order was determined, the first three conditions were presented in the first session, conditions 4–6 in the second session, and the last three conditions in the last session.

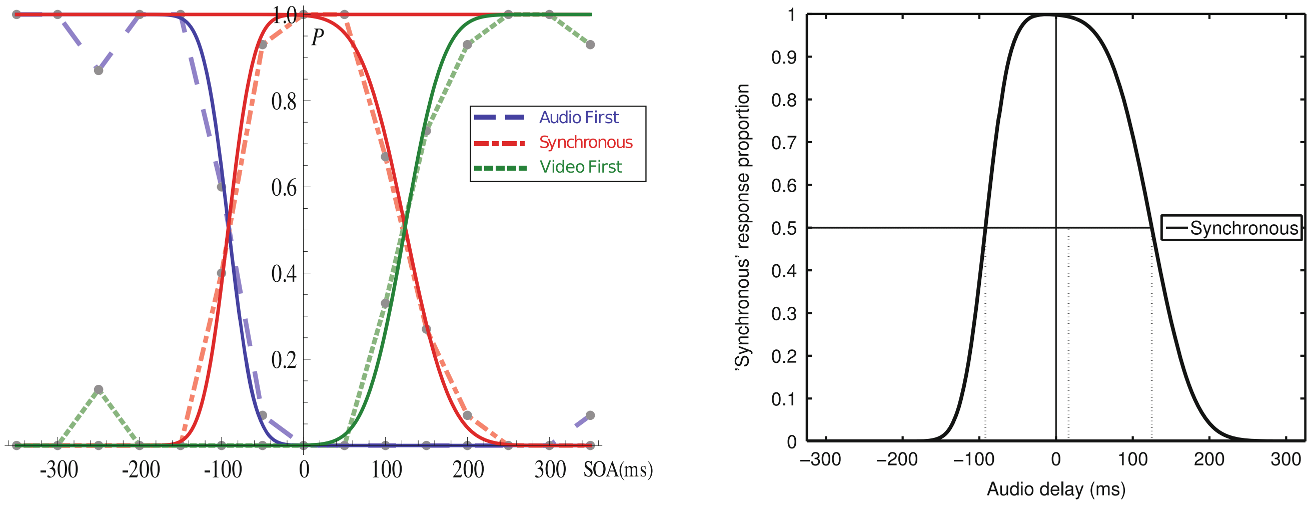
Results

To analyze the data, we used the same method as van Eijk et al. (2008). In this method, the response proportions for all three response categories were fitted with a psychometric function of the form $\gamma + (1 - \gamma - \lambda)F(\alpha, \beta)$, with F being the cumulative Gaussian distribution with mean α and standard deviation β. Fitting was done using the MATLAB psignifit toolbox, which implements the maximum-likelihood method described by Wichmann and Hill (2001). Stimulus-independent lapses (e.g., pressing the wrong key or blinking during stimulus presentation) were fitted by the γ (lower horizontal asymptote) and λ (higher horizontal asymptote) parameters. For the audio-first and video-first curves, the γ (gamma) and λ (lambda) parameters were later removed from the theoretical or underlying psychometric function, which is assumed to represent the actual perception of the participant (rather than the observed performance, van Eijk et al. 2008). The "synchronous" response category was fitted separately on each side to allow for possible asymmetry in the transition between "audio first" and "synchronous" and that between "synchronous" and "visual first." In these synchronous response fits, λ (lambda) was retained.

The raw data were plotted for each participant for each condition (an example plot for one subject is shown in Fig. 2), and then the parameters of the cumulative Gaussian functions of best fit were determined (see van Eijk et al. 2008 for more details of the procedure). Two participants did not report a "synchronous" response proportion above 50 % in all of the nine conditions, which is necessary to calculate the synchrony range and PSS. Two others did not report a transition between "synchronous" and "video-first" responses in all conditions. Therefore, the results of these participants were excluded from further analysis. The results of a fifth participant were excluded because of an experimenter error.

The data from the other 17 participants were analyzed separately for each participant and each condition. The PSS values are based on the synchrony range of the measurements. This range is determined by the synchrony boundaries, which are calculated by the intersections of the fitted asynchronous response curves (audio first and video first) with the fitted synchronous response curves. The PSSs are defined as the mean of the synchrony boundaries. This procedure was repeated for each condition for each participant separately. The PSS values for each condition of each participant can be found in Table 1, and the mean synchrony curves for each condition are shown in Fig. 3.

In the analysis, we first compared the results from the three conditions with equal stimulus durations, for which the mean PSS values were 16.4, 9.6, and 13.8 ms for stimulus durations of 12, 71, and 294 ms, respectively. Paired sample *t* tests showed no differences between PSS values of all three respective durations ($t(16) = 1.24$, $p = .23$, $t(16) = .95$, $p = .36$, $t(16) = .42$, $p = .68$). The constancy of mean PSS values with stimulus durations agrees with the results reported by Boenke et al. (2009). In the high intensity condition, they found PSS values of about 20 ms for all conditions, whereas our results for the equal-duration conditions are on average 7 ms smaller. Furthermore, the variation in the data they reported was larger than the present values (SEs of 10–15 ms with 22 participants in their data vs SEs of 5.0–6.6 ms with 17 participants in the present study, for equal-duration conditions). This



(a) Fitted + data curves

(b) Synchronous curve

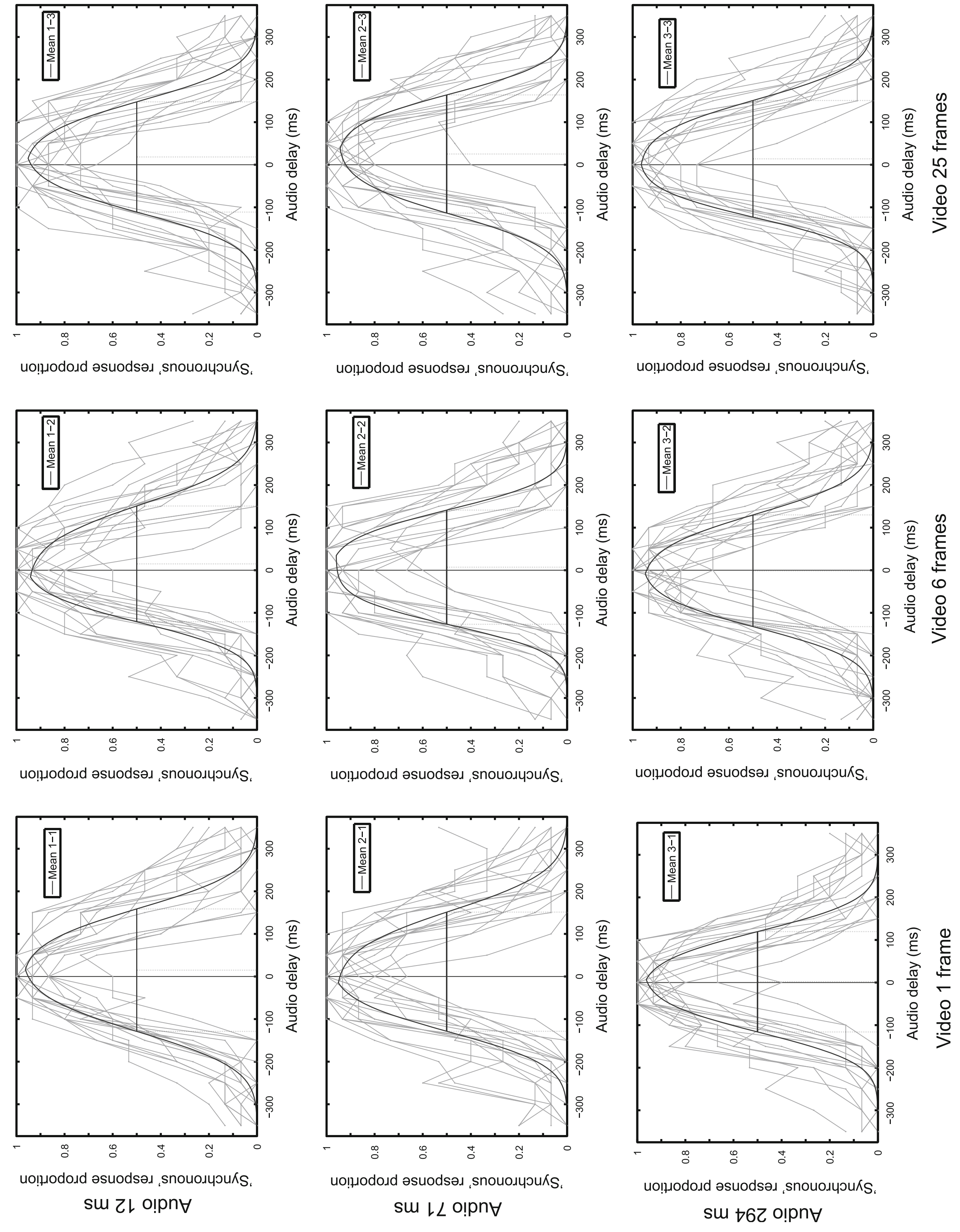
Fig. 2 Example of the fitted curves through the data points for one condition of one participant (audio 71 ms—video 294 ms) (*left*). The synchronous curve of the same condition of the same participant with indications of the synchrony boundaries and the PSS (*right*)

Table 1 PSS for each condition of each participant. Conditions are indicated by the duration of the auditory (A) and visual (V) components of the stimuli; $A_1 = V_1 = 12$ ms, $A_2 = V_2 = 71$ ms, $A_3 = V_3 = 294$ ms

Audio component	A_1	A_1	A_1	A_2	A_2	A_2	A_3	A_3	A_3
Video component	V_1	V_2	V_3	V_1	V_2	V_3	V_1	V_2	V_3
Participant									
P1	30.9	40.8	20.1	47.2	42.3	30.6	23.3	20.4	13.4
P2	3.4	1.3	−8.7	3.0	−58.1	−53.0	−20.3	−27.9	4.2
P3	−11.4	4.4	1.6	1.1	−10.8	16.3	−31.1	−15.3	11.9
P4	−44.9	−79.4	−11.5	−51.9	−37.4	−3.4	−25.1	−26.1	−23.2
P5	8.2	19.7	4.1	−27.6	4.8	22.4	−4.7	−31.5	−9.5
P6	15.8	41.0	54.1	43.6	28.4	61.1	22.9	2.9	40.4
P7	52.6	12.0	16.6	59.5	1.1	46.5	28.6	29.7	12.1
P8	11.7	16.3	13.6	7.5	4.6	20.4	−7.7	3.3	−2.7
P9	32.1	42.8	31.5	14.8	15.2	58.8	−5.0	−13.4	11.1
P10	41.6	49.4	61.5	58.4	44.2	65.8	28.4	34.5	52.3
P11	−14.8	−7.1	25.7	−2.7	−7.5	30.1	4.0	14.7	20.5
P12	76.2	9.8	4.1	7.0	4.4	29.0	27.8	1.0	23.1
P13	18.7	−3.4	11.0	0.8	32.7	28.0	−6.8	−5.9	12.4
P14	3.6	3.0	3.1	8.0	−4.2	−6.2	−5.0	−17.5	9.8
P15	15.9	53.0	22.1	6.9	18.2	14.0	1.4	15.2	29.9
P16	25.5	12.0	37.7	22.4	22.1	17.4	−5.1	−8.4	34.7
P17	13.6	11.6	48.8	2.5	1.8	32.9	0.4	7.1	26.8
Mean	16.4	13.4	20.3	11.1	9.6	23.9	−0.4	−0.6	13.8
SE	6.6	7.4	5.0	7.0	5.0	7.1	5.5	4.7	5.1

difference might be caused by differences in experimental procedures, that is, SJ (here) versus TOJ in Boenke et al., and the way specific conditions were blocked (here) or not (Boenke et al.).

Analyses of the width of the synchrony windows and the steepnesses of the synchronous curves did not show any significant differences across the conditions. As proposed by Boenke et al. (2009), the individual data of the 17 participants (Fig. 4) were analyzed for the relationship between the individual PSS values for the stimuli with intermediate duration and the amount of PSS shift with stimulus duration (difference in PSS for the longest and the


Audio 12 ms
Audio 71 ms
Audio 294 ms
Video 1 frame
Video 6 frames
Video 25 frames
Mean 1–1
Mean 1–2
Mean 1–3
Mean 2–1
Mean 2–2
Mean 2–3
Mean 3–1
Mean 3–2
Mean 3–3
'Synchronous' response proportion
Audio delay (ms)
0
0.2
0.4
0.6
0.8
1
−300
−200
−100
0
100
200
300

◀ **Fig. 3** The *panels* show synchrony curves for all nine combinations of audio and video durations. Durations are indicated along the axes. The parameters of the *solid curves* in each panel are the means of the fit parameters derived separately for the individual subjects. The *light grey lines* indicate the raw data of all participants for the corresponding condition. The *horizontal line* indicates the width of the synchrony window. The *dashed lines* indicate the synchrony boundaries and the PSS

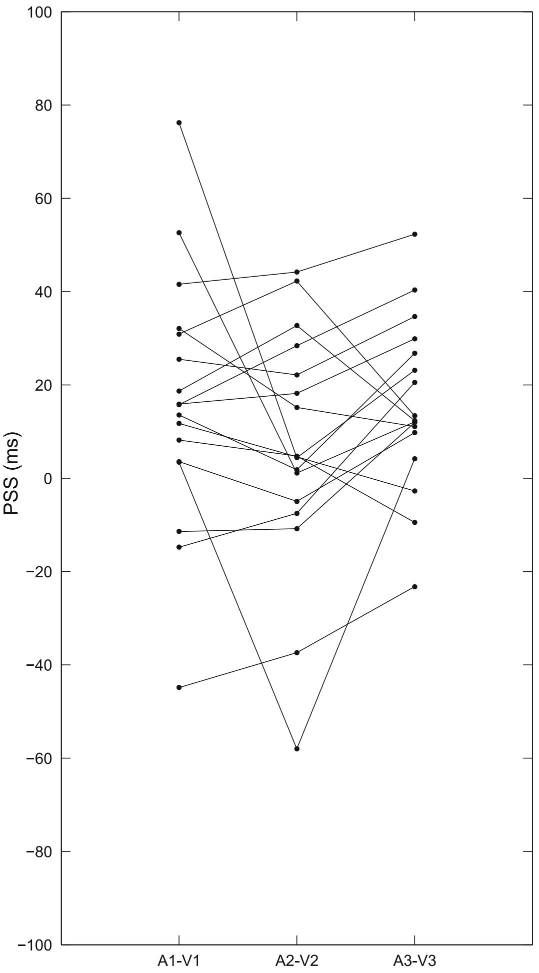


Fig. 4 The PSS values for all participants for the three equal-duration conditions. Duration values are the following: $A_1 - V_1 = 12$ ms, $A_2 - V_2 = 71$ ms, $A_3 - V_3 = 294$ ms

shortest stimuli). We did not observe a significant correlation in our results ($r(16) = .14$, $p = .60$), and this result remained true if individual data sets with a nonmonotonic relation between PSS and duration were removed (for example, by excluding subject P2 and/or P12).

We next compared the PSS values obtained for the nine different duration conditions (Fig. 5). The PSS values were analyzed with a 3 × 3 ANOVA. The analysis revealed a significant main effect for both audio (F (2, 15) = 8.2, $p < .01$) and video ($F(2, 15) = 4.9$, $p < .05$) duration. The interaction effect was not significant ($F(4,13) < 1$). Post hoc comparisons showed no significant differences in PSS values nor any interactions when analyzing the short (12 ms) and medium (71 ms) duration conditions. The data were therefore combined into a single “short duration” condition to be compared with the data from the 294-ms duration condition.

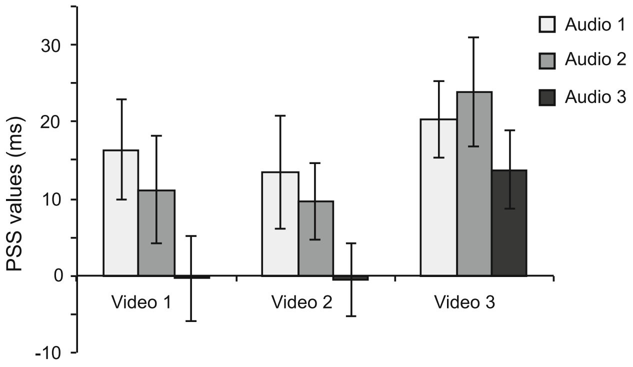


Fig. 5 Effects of video duration on PSS for short (*light grey*), middle (*grey*), and long (*black*) audio durations. Means and SE based on data of 17 participants

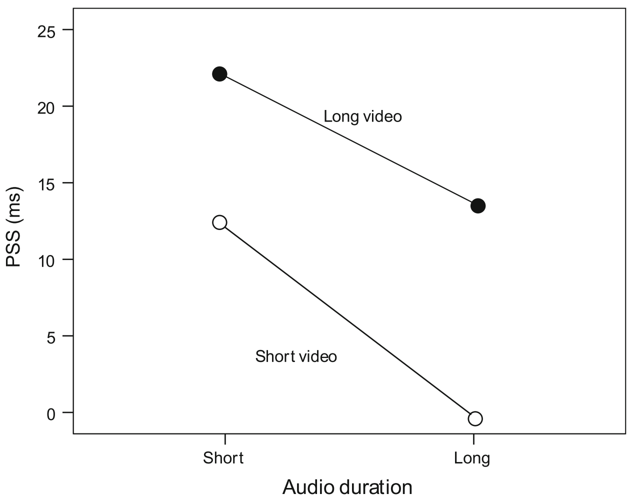


Fig. 6 Effects of audio duration on PSS for short (*open circles*) and long (*filled circles*) video durations

These PSS values were analyzed with a 2 × 2 ANOVA. The analysis showed a significant main effect for both audio ($F(1, 16) = 18.8$, $p < .01$) and video ($F(1, 16) = 13.6$, $p < .01$) duration. The interaction effect was not significant ($F(1,16) < 1$). It can be seen in Fig. 6 that for the short video condition, the PSS shifts from about 13 ms to 0 ms when the duration of the auditory stimulus is increased. In the long video condition, the result is a bit smaller: the PSS shifts from about 22 to 14 ms with increasing audio duration. The results show that as audio duration increases, the PSS shifts toward less positive audio delays. This implies that longer auditory stimuli need to be presented physically earlier, compared to short stimuli, in order to compensate for the delayed perceived moment of onset for longer stimuli. As a result, synchrony is perceived at a smaller visual leading SOA. As video duration increases, the PSS shifts toward more positive audio delays, and the opposite effect can be seen; synchrony is now perceived at a larger SOA.

Discussion

In the present study, the influence of stimulus duration on audio-visual synchrony perception was investigated. The main goal was to establish whether the unimodal timing effects of stimulus duration demonstrated for both auditory and visual stimuli also occur in multimodal, audio-visual conditions. We observed that indeed changes in the duration of the visual component and the audio component both affect the PSS in a systematic way. In addition, we observed no net effect on the averaged PSS value as long as the durations of the audio and video components were equal, in perfect agreement with the results of Boenke et al. (2009). One obvious difference in the finding between the two studies is the variability in PSS estimates. Boenke et al. found that with increasing stimulus duration, variability in PSS estimates decreased significantly from an SD of 69 ms (9-ms stimuli) to an SD of 46 ms (for 500-ms stimuli). In contrast, the variability in our study with a somewhat smaller sample number($N = 17$, vs. $N = 22$ in Boenke et al.) remained constant and the SD had values between 26 (12-ms stimuli) and 20 ms (294-ms stimuli). It appears that, in particular, the variability for the shortest stimulus duration in the data by Boenke et al. is high, also in comparison with many other studies which had estimated PSS values with short stimuli (10–20 ms) using a TOJ procedure. Values found in the literature are the following: SD = 33 ms ($N = 10$, stimulus duration 12 ms, van Eijk et al. 2008); SD = 18 ms ($N = 8$, 8 ms, Spence et al. 2003); SD = 24 ms ($N = 10$, 20 ms, Vroomen et al. 2004); SD = 17 ms ($N = 9$, 9 ms, Zampini et al. 2003, experiment 1). The paper by Zampini et al. contains an interesting additional observation. The low SD value of 17 ms was found when the task of the subjects was to answer, which *modality* came first. In a second experiment, subjects were asked to respond *which side came first*, the same experimental procedure used by Boenke et al. For the two conditions where light and sound stimuli were presented from opposite sides, the mean SD was 45 ms, that is, about a factor of 3 higher. Thus, there are two possible procedural aspects that might have contributed to the relatively high SD in the short-duration data of Boenke et al. One aspect is that in their study, conditions with different stimulus durations and with two different light intensities were presented interspersed in a pseudo-randomized way, while in all other studies, as well as in the present study, these stimulus parameters were kept constant within a measurement block. The second is the response paradigm to judge across-modal synchrony, where the response "which modality came first" seems to lead to less variability between subjects, then the paradigm "which side came first". Thus, dividing attention between different spatial locations might lead to additional noise in the obtained estimates of timing parameters (see also below).

The results concerning the mean PSS values clearly indicate that the perceived onsets of both auditory and visual stimuli were influenced in quantitatively similar ways by changes in duration. This resulted in equivalent, although opposite effects on the PSS in a synchrony judgment task. This observation also suggests that unimodal onset extraction precedes cross-modal timing comparisons by some significant amount of processing time. It is well known that auditory signals excite midbrain cells about 50 ms faster than visual signals (e.g., King and Palmer 1985), and multimodal cells in the superior colliculus show a range of SOAs around this value over which they respond in a more or less additive way to auditory and visual inputs (e.g., Sanford et al. 2005; Stein and Meredith 1993). However, without cortical involvement, multisensory enhancement does not occur (Sanford et al. 2005), and cortical cross-modal interactions in humans do not occur until about 165 ms after stimulus onset in the occipito-temporal ventral stream, and not until about 220 ms in the peri-sylvian cortex (Teder-Sälejärvi et al. 2002). More recently, Naue et al. (2011) found evidence of auditory effects in frontocentral and occipital cortices about 50–200 ms after the auditory component of audio-visual pairs was presented. These delays in processing multimodal components of simple stimuli support the idea that stimulus onsets are processed before multimodal integration, and unimodal timing phenomena should be rendered intact in multimodal synchrony judgments.

In previous unimodal research, the effects of duration were somewhat larger than the effects we found in the present multimodal experiment. In auditory experiments, the perceived stimulus onset was shifted by up to 30 ms as stimulus duration increased from 5 to 350 ms (Schimmel and Kohlrausch 2008). We found an effect of about 10 ms for a stimulus duration increase from 12 to 294 ms. Also, in visual research, somewhat larger shifts of the perceived stimulus onset have been found, about 20 ms for a stimulus duration increase from 6 to 72 ms (Brenner and Smeets 2010), while we found an effect of 12 ms over a larger range of durations.

The differences in the amount of perceived onset shift between the unimodal studies from the literature and our cross-modal study could be due to several factors. First, in our experiment, the stimuli of different durations were presented with the same intensity, thus the total energy in the stimuli increased in proportion to their duration. In contrast, the results in the literature were obtained for stimuli for which the intensity was increased for shorter stimulus durations with the intention of keeping the apparent brightness, or loudness, respectively, constant. In the visual study, this was achieved by adapting the stimuli according to Bloch's law (Brenner and Smeets 2010). Bloch's law states that at least for relatively dim visual stimuli shorter than 100 ms, the perceived brightness is equal to the product of intensity and stimulus duration (Bloch 1885). This implies that also (nearly) all stimuli in the study by Jaśkowski (1991) had the same brightness because, with one exception, all stimuli used by him had durations greater than 100 ms. In the auditory study (Schimmel and Kohlrausch 2008), stimulus levels were adjusted such that the overall loudness was the same for all durations (for the relation between stimulus duration and perceived loudness, see, for example, Florentine et al. 1996). Thus, in contrast to the conditions from the two cited studies, the brightness and loudness of our shortest stimulus was certainly lower than the one of intermediate duration. This is an important experimental difference, because it has been shown with various paradigms that the perceived onset of auditory and visual stimuli occurs relatively later for stimuli with lower intensity (e.g., Roufs 1963, 1974 and Boenke et al. 2009, for visual stimuli; Schimmel and Kohlrausch 2008 for auditory stimuli—but compare Roufs 1963, who reported only a very weak dependence of the perceptual lag for varying the intensity of an auditory stimulus over a range of 20 dB). This experimental choice might explain why we did not observe a systematic difference in the PSS values between the two shortest stimuli. According to our initial hypothesis based on stimulus duration, the perceived onset for the 71-ms stimuli should be delayed relative to the perceived onset for the 12-ms stimuli. On the other hand, because the 71-ms stimuli have a higher overall loudness/brightness than the 12-ms stimuli, they should be processed faster in the corresponding perceptual systems, which should to some extent compensate the expected duration effect. In order to have quantitative support for this way of interpreting our data, more independent results on the effects of brightness and loudness on the perceptual lag would be needed.

Another explanation for our smaller duration effects could be based on the increased effort involved in trying to focus on two modalities rather than on one. For example, in the Brenner and Smeets paper, participants had to compare two visual stimuli next to each other, both near the center of the visual field. A comparison of two stimuli in the same modality near the center of attention is much more common and therefore should be easier to focus on for participants than the stimuli used in our experiment, in which input from two sensory modalities had to be compared. Differences in results between unimodal and multimodal experiments could be due to differences in attentional allocation, as full attention is presumably directed to the single modality in the unimodal case, but is inevitably shared in some proportional way in the multimodal case. In studies of prior entry, in which relative attention between two components of a stimulus pair is manipulated, differences between unimodal and multimodal stimuli have been found in TOJ and SJ tasks (see Spence et al. 2001, for haptic-visual comparisons and Zampini et al. 2005, for audio-visual comparisons). Zampini et al. found, in an audio-visual SJ task, that the results for unimodal pairs were more accurate (SD = 51 ms) than for bimodal stimulus pairs (SD = 97 ms). Divided attention between modalities thus seems likely to increase noise and variability in synchrony judgments, but it is not obvious how this effect alone could contribute to reduced duration effects. To our knowledge, direct comparisons of duration effects have not been made between unimodal and multimodal studies, and it is obvious that such comparisons are needed to assess whether duration effects are reduced in multimodal studies due to attentional limitations or to early sensory interactions that could mitigate duration differences for multimodal stimuli.

At present, we are not able to conclude with certainty which explanation is most appropriate for the differences in the amount of perceived onset shift between the unimodal findings from literature and our multimodal results for short durations up to 70 ms. Nevertheless, we consider our data to reveal a relevant additional factor, which needs to be considered in multimodal synchrony studies with complex audio-visual stimuli. For complex stimuli like speech or music, the perceived onsets can have quite different relative shifts compared to the physical onsets of the stimuli. Given the increasing use of such stimuli in intermodal timing experiments, one needs to be aware that there might exist systematic unimodal differences between physical and perceived onsets for the experimental stimuli. Such relative within-modality shifts could quite well lead to quantitative shifts in intermodal timing parameters, like the PSS. Only when those unimodal effects are quantitatively accounted for can one correctly interpret the (remaining) temporal effects as reflecting true intermodal properties.

Conclusion

We studied the role of stimulus duration in an audio-visual synchrony judgment task. We found both audio and video duration effects on the PSS estimated from synchrony judgments. The effects were as predicted from the results of unimodal experiments. When the components have unequal durations, a shift in perceived synchrony was found. This shift in perceived synchrony was observed in the expected negative direction for longer audio durations and in a positive direction for longer video durations. For equal durations of the components, the absolute durations of the stimuli have no net influence on their perceived synchrony. These results thus demonstrate that unimodal changes in perceived onset timing due to changes in stimulus duration are also reflected in estimates of the PSS in multimodal synchrony experiments. Given that such changes in perceived onsets are known from auditory research with speech and music stimuli, these unimodal effects might be of relevance for the interpretation of cross-modal synchrony experiments using stimuli with a complex acoustic structure like speech or music.

Acknowledgments We would like to thank the three reviewers for their extensive and valuable remarks on earlier versions of this manuscript.

References

Bloch AM (1885) Expériences sur la vision. C R Seances Soc Biol (Paris) 37:493–495

Boenke LT, Deliano M, Ohl FW (2009) Stimulus duration influences perceived simultaneity in audiovisual temporal-order judgment. Exp Brain Res 198:233–244

Brenner E, Smeets JBJ (2010) How well can people judge when something happened?. Vis Res 50:1101–1108

Florentine M, Buus S, Poulsen T (1996) Temporal integration of loudness as a function of level. J Acoust Soc Am 99:1633–1644

Jaśkowski P (1991) Perceived simultaneity of stimuli with unequal durations. Perception 20:715–726

King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. Exp Brain Res 60:492–500

Marcus SM (1981) Acoustical determinants of perceptual center (P-center) location. Percept Psychophys 30:247–256

Morton J, Marcus S, Frankish C (1976) Perceptual centers (P-centers). Psychol Rev 83:405–408

Naue N, Rach S, Strüber D, Huster RJ, Zaehle T, Körner U, Herrman CS (2011) Auditory event-related response in visual cortex modulates subsequent visual responses in humans. J Neurosci 31:7729–7736

Roufs JAJ (1963) Perception lag as a function of stimulus luminance. Vis Res 3:81–91

Roufs JAJ (1974) Dynamic properties of vision—V. Perception lag and reaction time in relation to flicker and flash tresholds. Vis Res 14:853–869

Sanford TR, Quessy S, Stein BE (2005) Evaluating the operations of underlying multisensory integration in the cat superior colliculus. J Neurosci 25:6499–6508

Schimmel O, Kohlrausch A (2008) On the influence of interaural differences on temporal perception of noise bursts of different durations. J Acoust Soc Am 123:986–997

Schütte H (1978) Ein Funktionsschema für die Wahrnehmung eines gleichmäßigen Rhythmus in Schallimpulsfolgen. Biol Cybern 29:49–55

Spence C, Shore DI, Klein RM (2001) Multisensory prior entry. J Exp Psychol: Gen 130:799–832

Spence C, Baddeley R, Zampini M, James R, Shore DI (2003) Multisensory temporal order judgments: when two locations are better than one. Percept Psychophys 65:318–328

Stein BE, Meredith MA (1993) The merging of the senses. MIT Press, Cambridge

Sternberg S, Knoll RL (1973) The perception of temporal order: fundamental issues and a general model. In: Kornblum S (eds) Attention and performance IV. Academic Press, New York, pp 629–685

Strange W, Bohn OS (1998) Dynamic specification of coarticulated German vowels: perceptual and acoustic studies. J Acoust Soc Am 104:488–504

Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An anlysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Cogn Brain Res 14:106–114

van Eijk RLJ, Kohlrausch A, Juola JF, van de Par S (2008) Audiovisual synchrony and temporal order judgments: effects of experimental method and stimulus type. Percept Psychophys 70:955–968

van Eijk RLJ, Kohlrausch A, Juola JF, van de Par S (2010) Temporal order judgment criteria are affected by synchrony judgment sensitivity. Atten Percept Psychophys 72:2227–2235

Vroomen J, Keetels M, de Gelder B, Bertelson P (2004) Recalibration of temporal order perception by exposure to audio-visual asynchrony. Cogn Brain Res 22:32–35

Vroomen J, Keetels M (2010) Perception of intersensory synchrony: a tutorial review. Atten Percept Psychophys 72:871–884

Wichmann FA, Hill NJ (2001) The psychometric function: I. Fitting, sampling, and goodness of fit. Percept Psychophys 63:1293–1313

Zampini M, Shore DI, Spence C (2003) Audiovisual temporal order judgments. Exp Brain Res 152:198–210

Zampini M, Shore DI, Spence C (2005) Audiovisual prior entry. Neurosci Lett 381:217–222