Available online at www.sciencedirect.com
ScienceDirect
Neuroscience Letters 405 (2006) 132–136
Neuroscience Letters
www.elsevier.com/locate/neulet

Evaluating the influence of frame rate on the temporal aspects of audiovisual speech perception

Argiro Vatakis *, Charles Spence

Crossmodal Research Laboratory, Department of Experimental Psychology, University of Oxford, UK

Received 20 March 2006; received in revised form 13 June 2006; accepted 20 June 2006

Abstract

We investigated whether changing the frame rate at which speech video clips were presented (6–30 frames per second, fps) would affect audiovisual temporal perception. Participants made unspeeded temporal order judgments (TOJs) regarding which signal (auditory or visual) was presented first for video clips presented at a range of different stimulus onset asynchronies (SOAs) using the method of constant stimuli. Temporal discrimination accuracy was unaffected by changes in frame rate, while lower frame rate speech video clips required larger visual-speech leads for the point of subjective simultaneity (PSS) to be achieved than did higher frame rate video clips. The significant effect of frame rate on temporal perception demonstrated here has not been controlled for in previous studies of audiovisual synchrony perception using video stimuli and is potentially important given the rapid increase in the use of audiovisual videos in cognitive neuroscience research in recent years.

Keywords: Temporal perception; Frame rate; Video quality; Speech; Audition; Vision

Recent technological advances have rendered the transmission and observation of multimedia videos a routine part of daily communication for many people. Nowadays, the use of multimedia video clips to communicate and transmit information (e.g., for videoconferencing and entertainment) is no longer only experienced through the traditional medium of the television set but also through mobile devices (such as mobile phones and personal data assistants) as well as via computer systems. More importantly, multimedia video clips are increasingly being used to present more realistic and ecologically valid stimuli for behavioral and neuroimaging research (think, for example, of the use of audiovisual video clips for research on speech perception). In contrast to other technologies (such as text-based, image-only, or voice-only presentations), however, multimedia presentations reflect a more technologically demanding medium for communication (since they require more powerful computer and compression systems for transmission), one that is more sensitive to artifacts (i.e., the quality can be compromised by many factors, such as image size, compression, transfer speed, etc.). This higher demand (and consequent increased sensitivity to artifacts) is driven by the fact that multimedia presentations typically consist of two or more continuous information streams, normally one auditory and the other visual, which have to be delivered in such a way as to provide the viewer with a consistent, acceptable, and comprehensible message [5,13,19,24]. This means that any asynchrony between the content of the various information streams has to be kept within acceptable limits [28].

To date, the majority of research that has attempted to assess the effect of multimedia video quality on user perception has been conducted using high-quality systems (e.g., [19]), artificial participant groups (e.g., participants assigned randomly to tasks that were unrelated to their daily experiences; [28]), and/or by focusing on subjective measures of user satisfaction ([3,5,15,16,28]; though see [34,35], for physiological measurements). However, high-quality systems require powerful computer systems and relatively constant environments (e.g., stable lighting conditions, room acoustics, etc.) that are not available in many real-world situations [11]. In addition, user satisfaction may not provide an accurate measure of a user's ability to comprehend and utilize the informational content of the message being presented by the multimedia video in a meaningful manner (cf. [35]). What is more, it should also be noted that investigating the satisfaction of users who are unfamiliar with the experimental task may not necessarily provide an accurate

* Correspondence to: Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford OX1 3UD, UK. Tel.: 44 1865 271307; fax: +44 1865 310447.
E-mail address: argiro.vatakis@psy.ox.ac.uk (A. Vatakis).

0304-3940/$ – see front matter © 2006 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2006.06.041

measure of user satisfaction levels as compared to participants who are exposed daily to the task under investigation. Therefore, the study of how the manipulation of the parameters relevant to the presentation of multimedia videos (including the manipulation of frame rate) may affect human perception remains an important issue for applied research (see [7]).

The subjective quality of multimedia video clips has been shown to depend on both the dynamic nature of the motion picture (referred to as its 'temporality') and on their 'watchability' (e.g., [3]). For example, footage of a football match is considered as being high in 'temporality' due to the rapid change of scene sequences, while footage of a news report is considered as low in 'temporality' due to the relatively static nature of the speakers. The 'watchability' of a video clip refers to the clarity and acceptability of the auditory signal, the continuous sequence of the visual images (i.e., frames), and the extent to which the visual signal is synchronized with the auditory signal [3,27–29]. Frame rate (how quickly consecutive frames are presented) is one of the most important parameters that affects both the 'temporality' and 'watchability, of video clips [3,22].

Previous research on the evaluation of multimedia video quality in terms of variations in frame rate is, however, surprisingly scarce. The few studies that have been conducted to date have primarily focused on the effect of varying the frame rate of the video stream on communication using *subjective* report measures (e.g., [3,15,16,19]; though see [34,35]) or by asking the participants to report the word or consonant uttered in a speech token or in McGurk-type (i.e., inconsistent auditory and visual) presentations (e.g., [7,20,23,33]). Therefore, as yet, no *objective* perceptual measure of the effects of variations in the frame rate of a video stream on a user's perception of audiovisual synchrony has been reported. This might be due to the widespread assumption that people's performance will be unaffected by variations in frame rate if they are not consciously aware of such differences (i.e., according to subjective report data, once the frame rate of a video presentation exceeds 15 frames per second, fps; the synchronization of audiovisual speech is perceived, while full motion is perceived at 25 fps; e.g., [2,17,22,35]).

In the present study, therefore, we investigated whether or not variations in the frame rate of video clips would affect audiovisual perception of temporal order. Examining the consequences of manipulating video frame rate on temporal perception is of critical importance from an applied perspective, given that demonstrating such effects would imply that the acceptable temporal window for synchrony perception for audiovisual stimuli is dependent on the frame rate at which the stimuli are presented [20,23]. The possible influence of frame rate on temporal perception is also important from a theoretical perspective given that the use of more ecologically valid stimuli in the form of video clips is becoming increasingly popular in laboratory-based cognitive neuroscience research on multisensory perception. As yet, however, little consideration has been given to the consequences of changes of frame rate on human perception. This issue deserves consideration given the recent increase of neuroscience interest in the perception of synchrony, and the fact that many imaging studies have now started to utilize audiovisual video presentations with varying degrees of asynchrony (e.g., [6,21]).

Given the importance of this research question, we therefore decided to conduct an experiment using video clips displayed at different frame rates while keeping all other parameters of the clips constant. We used brief audiovisual speech video clips, that were chosen due to their low dynamic motion content (i.e., there were few event changes from one frame to the next, thus making the stimulus parameters easier to control; e.g., in terms of resolution, auditory stream composition, etc.; [13]). In addition, the very low informational content of the video clips ensured that their content did not draw participants' attention away from the physical attributes of the video clips themselves [4]. The perception of the speech stimuli was assessed using a temporal order judgment (TOJ) task with a range of stimulus onset asynchronies (SOAs) presented using the method of constant stimuli (see [26]).

Our use of a TOJ task to assess the effects of video frame rate variation on audiovisual temporal perception enabled us to test two predictions regarding participants' temporal discrimination performance, as measured by the just noticeable difference (JND) and the point of subjective simultaneity (PSS). The JND provides a standardized measure of the accuracy with which the participants were able to correctly discriminate the temporal order of the auditory and visual speech signals. The PSS indicates the amount of time by which one sensory modality had to lead the other in order for synchrony to be perceived (i.e., for participants to make the 'speech-sound first' and 'visual-speech first' responses equally often). In terms of the JND, one might predict that any reduction in frame rate would, if anything, result in an increase in the JND (i.e., worse performance), driven by the lowering of the 'quality' of fine-grained spatiotemporal information that would be available visually (e.g., [12,20,23,33]). However, the low motion stimuli utilized in this study and the resilience of speech intelligibility in noisy environments might also lead to the prediction of a null effect on the JND.

In terms of the modality leads/lags required for the PSS to be attained, one might predict that the amount of time by which the visual stream has to lead the auditory stream (for the PSS to be achieved) should decrease as the video frame rate increases [20,23]. Such a modulation of the PSS would be expected given the small delays (in terms of the physical onset of the visual-speech with respect to the speech-sound) that necessarily result from any reduction in frame rate: Over the course of the video, the visual stream would, on average, be physically delayed by approximately half of the duration of each frame, given that the visual image and the auditory signal would be more or less synchronous at the onset of a new visual frame, but the visual stream would gradually fall behind the on-going auditory stream until the onset of the next new visual frame, at which point audiovisual synchrony would once again be restored.

Eleven participants (6 male and 5 female) aged between 18 and 33 years (mean age of 25 years) took part in the experiment. All of the participants were naïve as to the purpose of the study and all reported having normal or corrected-to-normal hearing and visual acuity. The experiment was performed in accordance

with the ethical standards laid down in the 1964 Declaration of Helsinki, as well as the ethical guidelines laid down by the University of Oxford. The experiment took 40 min to complete.

The experiment was conducted in a completely dark sound-attenuated booth. During the experiment, the participants were seated comfortably facing straight-ahead. The visual stimuli were presented on a 17-inch (43.18 cm) CRT monitor (75-Hz maximum refresh rate), placed at eye level, approximately 98 cm in front of the participants. The auditory stimuli were presented by means of two loudspeaker cones (11.1 cm in diameter), one placed 27.6 cm to either side of the center of the monitor. The audiovisual stimuli consisted of black-and-white video clips presented on a black background, using Presentation (Version 9.90; Neurobehavioral Systems, Inc., CA). The video clips (380 × 400-pixel, Cinepak Codec video compression, 16-bit audio sample size, 24-bit video sample size) were processed using the Adobe Premiere 6.0. The video clips consisted of a female British-English speaker (frontal view, including the head and shoulders), looking directly at the camera, and uttering the Vowel-Consonant-Vowel (VCV) syllables: /aba/ and /aga/ (the open vowel /a/ was used in order to provide high levels of visible contrast relative to the closed mouth in the rest position at the start and end of each video clip, thus ensuring that even at low frame rates the closed lip event was captured; cf. [23]) at one of four different frame rates (6, 12, 24, and 30 fps).

The 'Exporting' function in Adobe Premier was utilized to construct each video clip, in order to manipulate the number of frames displayed per second for each clip without changing the speed of action or any other feature (i.e., duration) of the original clip. The participants responded using a standard two-button computer mouse, which they held with both hands, using their right thumb for 'visual-speech first' responses, and their left thumb for 'speech-sound first' responses (or vice versa, the response buttons were counterbalanced).

Nine SOAs between the auditory and visual signals were used: ±300, ±200, ±133, ±66, and 0 ms. Negative SOAs indicate that the speech-sound stream was presented first. The participants completed one block of eight practice-trials before the main experimental session in order to familiarize themselves with the task and the video clips. The practice-trials were followed by 10 blocks of 72 experimental-trials, consisting of a single presentation of each of the eight video clips at each of the nine SOAs (presented in a random order) in each block of trials.

The participants were informed that they would have to decide on each trial whether the speech-sound or visual-speech stream appeared to have been presented first. The participants were also informed that the task was self-paced, and that they should respond only when confident of their response. The participants were instructed prior to the experiment not to move their heads and to maintain fixation on the center of the monitor throughout each block of trials. Upon completion of the experiment and before debriefing took place the participants were asked if they observed any differences between the clips in terms of their quality (e.g., resolution, jitter, speed, etc.). None of the participants had reported noticing any differences in the quality of the video clips presented except in relation to the auditory and visual stream leads and lags (i.e., none of the participants noticed the changes taking place in the frame rate; cf. [35]).[1]

The proportions of 'visual-speech first' responses (see Fig. 1A) were converted to their equivalent z-scores under the assumption of a cumulative normal distribution [14]. The data were used to calculate best-fitting straight lines for each participant for each condition, which, in turn, were used to derive values for the slope and intercept. These two values were used to calculate the JND (JND = 0.675/slope; since ±0.675 represents the 75 and 25% point on the cumulative normal distribution) and the PSS (PSS = −intercept/slope) values (see [10], for further details). For all of the analyses reported here, the data derived from the /aba/ and /aga/ video clips were collapsed. Note that no differences were obtained for the JND/PSS values of these two speech tokens (although see [9,30]) and Bonferroni-corrected t-tests (where $p < .05$ prior to correction) were used for all post hoc comparisons.

Repeated measures analysis of variance (ANOVA) on the JND data with the factor of Frame Rate (6, 12, 24, or 30 fps) revealed no main effect [$F(3, 63) < 1$, n.s.]. That is, the accuracy of participants' judgments of the temporal order of the auditory and visual speech signals was unaffected by the frame rate at which the video clips were presented (see Fig. 1B). The JND values for the speech video clips were all in the range of 66–76 ms, which is well within the range observed (i.e., 59–95 ms) in previous studies on audiovisual temporal synchrony perception using discrete speech stimuli ([31,32]).

Analysis of the PSS data, however, revealed a significant main effect of frame rate [$F(3, 63) = 6.13$, $p < .01$], with a larger visual lead being required for video clips presented at 6 fps (mean PSS = 42 ms) than for the video clips presented at 24 (mean PSS = 17 ms) or 30 fps (mean PSS = 1 ms) [$p < .05$, for both comparisons] (see Fig. 1C). No differences were obtained between the 12 fps (mean PSS = 30 ms) clip and any of the other video clips. Paired samples t-tests revealed that the PSS values for the 6 and 12 fps video clips were significantly different from 0 ms, while those of the 24 and 30 fps clips failed to reach statistical significance [$t(21) = 1.33$, $p = .20$, $t(21) < 1$, n.s., respectively].

These results provide empirical evidence concerning how the quality of a multimedia video (in terms of variations in frame rate) affects audiovisual temporal perception (though see [20,23,33–35], for studies investigating other perceptual consequences of changes in frame rate). Modulating the quality of the video clips by using high (i.e., 24 or 30 fps) versus low (i.e., 6 or 12 fps) frame rates did not affect the participants' ability to

[1] We also conducted a follow-up study in order to examine whether people were able to correctly identify the frame rate of the audiovisual speech video clips used in the present study at above chance levels. 10 new participants were presented with the video clips used in the main experiment and had to try and identify the frame rate of each clip by pressing one of four buttons on a standard computer keyboard (corresponding to 6, 12, 24, and 30 fps, respectively; each condition was presented five times in a randomized order). The participants were, however, unable to correctly identify the frame rate at which the video clips were presented. The mean accuracy of their discrimination responses (26% correct) did not differ significantly from chance (i.e., from 25% correct) [$t(9) = 1.80$, $p = .86$]. The participants were poor at discriminating all four frame rates (i.e., 31, 18, 25, and 28% correct for 6, 12, 24, and 30 fps, respectively).

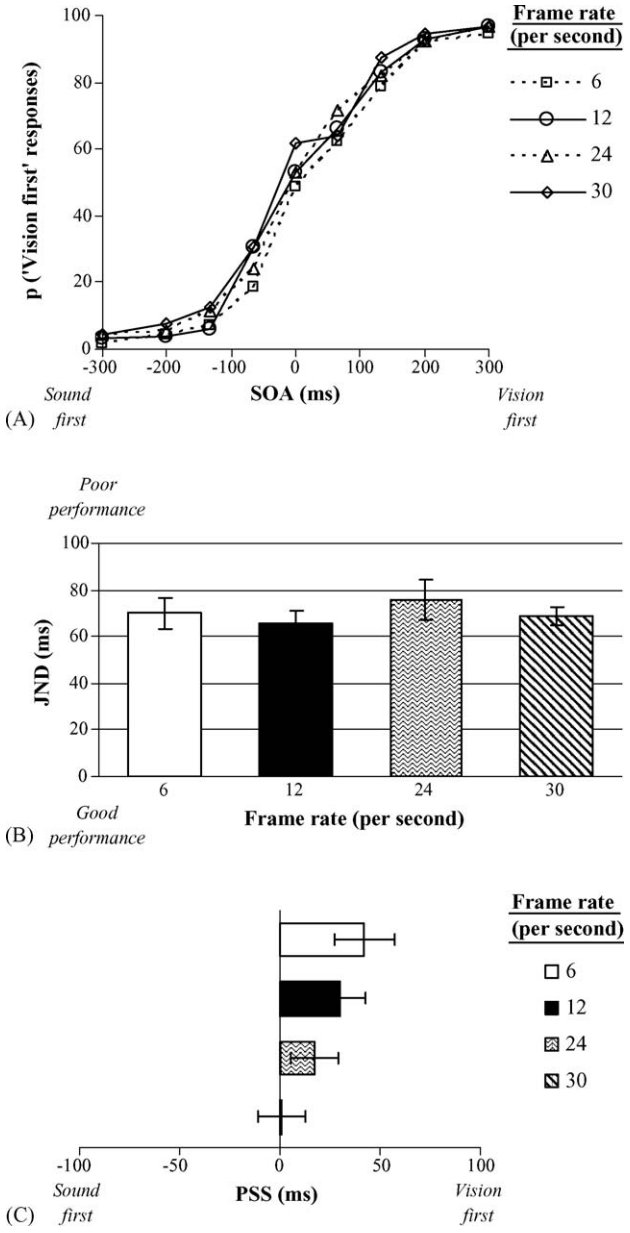

Fig. 1. (A) Mean percentage of 'vision-first' responses plotted as a function of stimulus onset asynchrony (SOA) for each of the four frame rates used in the experiment. (B) Mean JND and (C) mean PSS values for the audiovisual speech video clips as a function of frame rate. The error bars represent the standard errors of the means.

accurately judge the temporal order of the audiovisual speech stimuli (an outcome that is consistent with the majority of previous research on this topic; e.g., [20,23]). However, variations in frame rate did influence the modality leads/lags required for the PSS to be achieved. Specifically, low frame rate video clips (i.e., 6 fps) required larger visual leads for the PSS to be reached as compared to the higher frame rate video clips (i.e., 24 and 30 fps; [20,23]).

The ability of participants to accurately discriminate the temporal order of video clips of varying frame rate (as measured by the JND) and their failure to notice any differences between the clips presented (both subjectively, as measured by post-experimental questioning, and objectively; see Footnote 1) might be taken to imply that the use of high-quality multimedia technologies is unnecessary or that our stimuli (which were relatively 'low in event changes') did not require fine-grained processing of the visual stream.[2] However, our findings indicate that there are actually perceptual differences between high and low frame rate multimedia videos in terms of the modality leads/lags required for the PSS to be achieved (see also [20,23,35]). Specifically, at lower frame rates, video clips require larger visual leads for the PSS to be reached, while at higher frames much smaller visual leads are required. The PSS values obtained for the VCV syllables used in the present study involved visual leads ranging from 1 to 42 ms, well within the range obtained in previous audiovisual studies using syllables as stimuli (visual lead range of 19–27 ms; [31,32]).

The broadcasting industry, well aware of the fact that people are sensitive to asynchrony in audiovisual stimuli, has established a maximum acceptable audiovisual asynchrony for broadcasts, stating that the auditory signal should not lead by more than 45 ms or else lag by more than 125 ms [18]. Research suggests that within this temporal window only a minimal deterioration in program intelligibility will be observed (cf. [25]). However, the present results suggest that this estimate of the temporal window might not always be appropriate, since any change in the frame rate at which a broadcast is presented may modulate the mid-point of this temporal window (i.e., by changing the PSS). On the basis of our results (and assuming a 50% threshold for acceptability), the acceptable temporal window of asynchrony at 6 fps would appear to be in the range of auditory leads of up to 28 ms or lags of up to 112 ms, whereas at 30 fps the window ranges from auditory leads of up to 68 ms or lags of up to 70 ms. Such variations in the acceptable window of asynchrony clearly need to be borne in mind given the variety of different frame rates used to transmit audiovisual videos in currently-popular electronic devices, such as video-enabled mobile phones where frame rate varies from 1.6 to 100 fps (see [1]).

These variations in the audiovisual perception of synchrony highlighted here may also have important implications for psychophysical and neuroimaging studies where audiovisual videos are increasingly being used as stimuli (e.g., [8]) and where the audiovisual perception of synchrony is becoming an increasingly popular topic for investigation (e.g., [6,21]). Additionally, recent attempts by researchers to acquire dynamic magnetic resonance imaging image sequences (i.e., MRI movies) in order to investigate the movements of the speech organs during speech

[2] We conducted an additional follow-up study in order to examine whether frame changes for events of high temporality (e.g., for videos containing a large number of event changes from one frame to the next) might be detected more easily by participants [3,22]. Twelve participants were presented with a video clip of a male and a female dancing and had to try and identify the frame rate of each clip (just as in Footnote 1). The participants were able to correctly identify the frame rate of the dancing video clips on 48% of the trials. This level of discrimination performance was significantly above chance (i.e., 25% correct) [$t(11) = 2.82$, $p \leq .05$]. The participants discriminated the four frame rate conditions presented at 32, 38, 55, and 68% correct for 6, 12, 24, and 30 fps, respectively.

are highly dependent on video clip frame rate and acquisition techniques.

Acknowledgements

We would like to thank R. Campbell and G.A. Calvert for providing the speech stimuli used in this study. A.V. was supported by a Newton Abraham Studentship from the Medical Sciences Division, University of Oxford.

References

[1] Abstract Worlds Ltd. (2006). The StrangeMaze 3D v1.1 benchmark. Retrieved on January 25th, 2006, from http://www.abstractworlds.com/Strangemaze/index. php?id=bench.
[2] A. Anderson, L. Smallwood, R. MacDonald, J. Mullin, A. Fleming, O. O'Malley, Video data and video links in mediated communication: what do users value? Int. J. Hum. Comput. Stud. 52 (2000) 165–187.
[3] R.T. Apteker, J.A. Fisher, V.S. Kisimov, H. Neishlos, Video acceptability and frame rate, IEEE Multimedia 2 (1995) 32–40.
[4] M.D. Basil, Multiple resource theory. I. Application to television viewing, Commun. Res. 21 (1994) 177–207.
[5] J.G. Beerends, F.E. de Caluwe, The influence of video quality on perceived audio quality and vice versa, J. Audio Eng. Soc. 47 (1999) 355–362.
[6] D. Bergmann, C. Spence, H.J. Heinze, T. Noesselt, Neural correlates of synchrony perception using audiovisual speech stimuli, 7th Annual International Multisensory Research Forum. Dublin, Ireland 17–21 June, 2006.
[7] A. Blokland, A.H. Anderson, Effect of low frame rate on intelligibility of speech, Speech Commun. 26 (1998) 97–103.
[8] G.A. Calvert, E.T. Bullmore, M.J. Brammer, R. Campbell, S.C. Williams, P.K. McGuire, P.W. Woodruff, S.D. Iversen, A.S. David, Activation of auditory cortex during silent lipreading, Science 276 (1997) 593–596.
[9] B.L. Conrey, D.B. Pisoni, Lipreading and auditory-visual asynchrony detection, Abstracts Psychon. Soc. 9 (2004) 89.
[10] S. Coren, L.M. Ward, J.T. Enns, Sensation & Perception, sixth ed., Harcourt Brace, Fort Worth, 2004.
[11] O. Daly-Jones, A. Monk, L. Watts, Some advantages of video conferencing over high-quality audio conferencing: fluency and awareness of attentional focus, Int. J. Hum. Comput. Stud. 49 (1998) 21–58.
[12] H. de Paula, H.C. Yehia, D. Shiller, G. Jozan, K.G. Munhall, E. Vatikiotis-Bateson, Linking production and percpetion through spatial and temporal filtering of visible speech information, in: Proceedings of the 6th International Seminar on Speech Production, Sydney, Australia, 7–10 December 2003, 2003, pp. 37–42.
[13] R. Finger, A.W. Davis (2001), Measuring video quality in videoconferencing systems. Report SN187-D.
[14] D.J. Finney, Probit Analysis: Statistical Treatment of the Sigmoid Response Curve, Cambridge University Press, London, 1964.
[15] G. Ghinea, J.P. Thomas, QoS Impact on user perception and understanding of multimedia video clips, in: Proceedings of ACM Multimedia 1998, Bristol, UK, 13–16 September, 1998, pp. 49–54.
[16] G. Ghinea, J.P. Thomas, Quality of perception: user quality of service in multimedia presentations, IEEE Trans. Multimedia 7 (2005) 786–789.
[17] Y. Inazumi, T. Yoshida, Y. Sakai, Y. Horita, Estimation of the optimal frame rate for video communications under bit-rate constraints, Electron. Commun. Jpn. 86 (2003) 54–67.
[18] ITU-R BT.1359-1 (1998), Relative timing of sound and vision for broadcasting (Question ITU-R 35/11).
[19] M. Jackson, A.H. Anderson, R. McEwan, J. Mullin, Impact of video frame rate on communicative behaviour in two and four party groups, in: Proceedings of CSCW 2000, Philadelphia, PA, 2–6 December, 2000, pp. 11–20.
[20] H. Knoche, H. de Meer, D. Kirsh, Compensating for low frame rates, in: Proceedings of the Conference on Human Factors in Computing Systems (CHI 2005), Portland, Oregon, USA, 2–7 April, 2005, pp. 1553–1556.
[21] E. Macaluso, N. George, R. Dolan, C. Spence, J. Driver, Spatial and temporal factors during processing of audiovisual speech perception: a PET study, Neuroimage 21 (2004) 725–732.
[22] G. Mastoropoulou, A. Chalmers, The effect of music on the perception of display rate and duration of animated sequences: an experimental study, in: Proceedings of Theory and Practice of Computer Graphics (TPCG 2004), 2004, pp. 128–134.
[23] K. Nakazono, Frame rate as a QoS parameter and its influence on speech perception, Multimedia Syst. 6 (1998) 359–366.
[24] T.N. Pappas, R.O. Hinds, On video and audio data integration for conferencing, in: Proceedings SPIE: Human Vision, Visual Processing, and Digital Display VI, Vol. 2411, San Jose, CA, 6–8 February, 1995, pp. 120–127.
[25] S. Rihs, The influence of audio on perceived picture quality and subjective audio-visual delay tolerance, in: R. Hamberg, H. de Ridder (Eds.), Proceedings of the MOSAIC Workshop: Advanced Methods for the Evaluation of Television Picture Quality, Eindhoven, 18–19 September, 1995, pp. 133–137.
[26] C. Spence, D.I. Shore, R.M. Klein, Multisensory prior entry, Journal of Exp. Psychol.: Gen. 130 (2001) 799–832.
[27] R. Steinmetz, Human perception of jitter and media synchronization, IEEE J. Selected Areas Commun. 14 (1996) 61–72.
[28] J.C. Tang, E. Isaacs, Why do users like video? Studies of multi-media supported collaboration, Comput. Supported Cooperative Work 1 (1992) 19–34.
[29] S. Van de Par, A. Kohlrausch, Sensitivity to auditory-visual asynchrony and to jitter in auditory-visual timing, in: Proceedings of SPIE conference: Human Vision and Electronic Imaging V, SPIE conference volume 3959, 2000, pp. 234–242.
[30] V. Van Wassenhove, K.W. Grant, D. Poeppel, Visual speech speeds up the neural processing of auditory speech, Proc. Nat. Acad. Sci. 102 (2005) 1181–1186.
[31] A. Vatakis, C. Spence, Audiovisual synchrony perception for speech and music using a temporal order judgment task, Neurosci. Lett. 393 (2006) 40–44.
[32] A. Vatakis, C. Spence, Crossmodal binding: evaluating the 'unity assumption' using audiovisual speech stimuli, submitted for publication.
[33] M. Vitkovitch, P. Barber, Effect of video frame rate on subject's ability to shadow one of two competing verbal passages, J. Speech Hear. Res. 37 (1994) 1204–1210.
[34] G. Wilson, M.A. Sasse, Investigating the impact of audio degradations on users: subjective vs. objective assessment methods, in: C. Paris, N. Ozkan, S. Howard, S. Lu (Eds.), Proceedings of OZCHI 2000: Interfacing Reality in the New Millenium, 2000, pp. 135–142.
[35] G. Wilson, M.A. Sasse, Do users always know what's good for them? Utilizing physiological responses to assess media quality, in: S. McDonald, Y. Waern, G. Cockton (Eds.), Proceedings of HCI 2000: People and Computer XIV—Usability or else!, Sunderland, UK, 5–8 September, 2000, pp. 327–339.