Mobile Screen Size Limits Multimodal Synergy

Frans van der Sluis
Department of Information Studies
University of Copenhagen
Copenhagen, Denmark
f.vandersluis@acm.org

Egon L. van den Broek
Department of Information and Computing Sciences
Utrecht University
Utrecht, The Netherlands
vandenbroek@acm.org

Annemiek van Drunen
Amsterdam University of Applied Sciences
Amsterdam, The Netherlands
avandrunen@gmail.com

John G. Beerends
TNO
The Hague, The Netherlands
john.beerends@tno.nl

ABSTRACT

Available bandwidth is still a limiting factor for mobile communication applications. Multisensory communication has already been identified as an possibility to moderate this limitation. One of the strengths of mobile communication lies in its combination of visual and auditory modalities. However, one of the most salient features of mobile devices have are their small screen size. This paper explores how the potential for multimodal synergy relates to the small screen size. In an experiment with 54 participants, the intelligibility was tested using a standardized video-listening test. The videos had a signal-to-noise ratio of -9dB and were presented on three different screen sizes, whilst keeping the video and auditory signals equal. Intelligibility was found to be significantly higher when using a large screen in comparison to using either of both smaller screens. We conclude that multisensory synergy is key to mobile applications, yet that screen size is a substantial constraint to this synergy. We argue that knowledge about human sensory processing can alleviate this constraint and maximize the potential quality of service of mobile video technology.

CCS CONCEPTS

• **Human-centered computing** → **HCI theory, concepts and models**; **Empirical studies in HCI**; *Empirical studies in accessibility*;

KEYWORDS

multimodal, multisensory, mobile, screen size, intelligibility, Quality of Service (QoS), Field of View (FOV), Signal-to-Noise Ratio (SNR)

ECCE'18, September 5–7, 2018, Utrecht, Netherlands

ACM ISBN 978-1-4503-6449-2/18/09…$15.00
https://doi.org/10.1145/3232078.3232101

ACM Reference Format:
Frans van der Sluis, Egon L. van den Broek, Annemiek van Drunen, and John G. Beerends. 2018. Mobile Screen Size Limits Multimodal Synergy. In *ECCE'18: Proceedings of the 36th European Conference on Cognitive Ergonomics (ECCE2018), September 5–7, 2018, Utrecht, Netherlands.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3232078.3232101

1 INTRODUCTION

Mobile technology such as e-readers, laptops, smartphones, smartwatches and other wearables are rapidly gaining momentum [12, 13, 20, 23, 29]. Applications such as mobile video telephony, mobile television, mobile Internet, navigation and mobile games are growing in use [4, 11, 22]. However, mobile technology is similar to other technologies in that its industrial developments are dominated by both a technology push and commercial interests [cf. 19]. This approach runs the risk of not considering important knowledge about human cognition during the development. "Tools over solutions" [26] seems to be the dominant adage [16]. The work presented here argues for a different, human-centered, approach [27, 28]. With a foundation in cognitive ergonomics, this paper evaluates the potential multimodal synergy that can arise in mobile video technology and, correspondingly, be a significant factor to its success.

Multimodal synergy is the synergy that arises in perception when multiple modalities are joined. Multimodal synergy not only enhances the intelligibility of a message presented visually as well as auditory, but also enhances aspects such as memory and emotion [5, 12]. These effects are particularly beneficial when the presented is complex [30] or the user is under high cognitive load [2]. However, the bimodal advantages can benefit from or be restricted by several auditory, visual, and bimodal issues. Of the possible restrictions to bimodal synergy, the influence of screen size has received little attention within the context of mobile devices. Their small screen is, however, one of their most salient features and screen size is likely to influence the bimodal advantages and, thus, their user experience. Jung et al. [14] plead for more research on the influence of screen size, concluding that: "*a small screen (..) is considered the fatal disadvantage of mobile TV service*" (p. 129) [cf. 33].

Research on screen size and multimodal synergy indicates a potential to improve the Quality of Service (QoS) of mobile video technology [15, 16]. Findings on the influence of large screens support this plea, as larger screens have been found to influence variables such as arousal, sense of presence, attention and memory, and connectedness [9, 15]. For most of these variables, the effects can be summarized as intensifying the values. Hence, "the larger, the better" seems to hold. Within reasonable limits, spatial resolution or information throughput has been shown less important for bimodal synergy [8]. These findings on bimodal synergy suggest that optimal bandwidth utilization might not be the only critical factor to consider for the QoS of mobile video technology.

Combining the findings on the effects of large screens as well as information throughput, this study examines whether multimodal synergy can be improved without the need for extra bandwidth capacity ([cf. 30]). This is examined in an experiment by increasing the Field of View (FOV) whilst keeping the information throughput constant. The experiment is performed above basic levels of visual acuity (i.e., one arc minute), to assure the participant does not acquire more information and any effects can be ascribed to enlarging multimodal synergy. Expected is that an increase in FOV enhances the bimodal advantages. To answer this hypothesis, the intelligibility of a message presented auditory as well as visually is measured. The relative importance of the visual compared to the auditory modality is increased by adding noise to the auditory channel. Consequently, changes in the visual channel are expected to have a greater effect on the intelligibility.

In the next Section, we present the research methods, including information on the participants, material, apparatus, and procedure as well as specifications on how the Field-Of-View (FOV) and Signal-to-Noise Ratio (SNR) are determined. Section 3 presents the results of the experiment. Subsequently, we close with Section 4, which presents a discussion of the results and its implications.

2 METHOD

To study the influence of FOV , a within-subjects design has been used evaluating the effects of screen size (i.e., small, medium, and large), video, and their sequence. The design served to counter-balance any order effects. This gave a total of 36 ($3! \times 3!$) different conditions, based on the possible number of combinations of three screen sizes and three videos, see also Table 1. The order of participation determined to which condition a subject was assigned. For example, the first subject was assigned to the first condition.

2.1 Participants

54 subjects (mean age: 20.3; range: 18-28) voluntarily participated in the research. 96.2% of the participants judged their level of English as either good or reasonable, which was verified by a standardized audio-only listening test of the Dutch Central Institute for Testing [3].

For all subjects, their English level was found to be sufficient (M = 8.89, SD = 1.98, the maximum score possible is 12). All participants had a (corrected to) normal vision and hearing.

2.2 Video and test material

A set of three videos were used from a standardized video listening test [3]. The videos were selected such that the face of the person talking was visible most of the time with the camera focused on the face; that is, talking-heads material, shown to be beneficial to bimodal presentation [21].

The questions from the standardized video listening test [3] were used to test for intelligibility. From the original set of questions, only those questions were selected that corresponded to a part of the video in which the speaker was visible. The resulting set contained two English three-choice questions per video. No restrictions were made on answering time. The test results were rated with the original CITO [3] scoring forms.

2.3 Apparatus

A computer was used with a 17" screen and a resolution of 1024×768. Three screen sizes were simulated by displaying the videos at 43.18 cm (1024×768 pixels), 17.00 cm (394×314 pixels), and 7.68 cm (178×142 pixels) diagonal.

A headset was used to present the videos with audio or audio alone. The subjects had to keep their heads between a square of ropes surrounding their head at forehead height, securing a fixed distance of 80 cm to the screen. In addition, the chair and keyboard were also placed at a fixed position.

All videos had a resolution of 178×142 pixels, the resolution of the smallest screen size, corresponding to an effective spatial resolution of 5.48 Pixels Per Centimetre (PPC) . This is considerably lower than the spatial resolution observable by the human eye [32].

To keep the amount of information constant, the smallest spatial resolution of 178×142 pixels was upscaled to the other two spatial resolutions. Hence, no extra information was given through the visual channel. The upscaling method used is the default algorithm of the Microsoft Windows video processing environment DirectShow [17], which is an enhanced bilinear method [24].

As an indicator of the FOV the Instantaneous Field of View (IFOV) was calculated for the azimuth (horizontal) direction. The IFOV combines the FOV of both eyes under a fixed head position [1].

Previous studies have shown that multimodal synergy increases when unimodal signal quality decreases [7]. For all videos, a Signal-to-Noise Ratio (SNR) of -9dB was used to enhance synergy to utilize this effect. This SNR was chosen in support of external validity. Speech elements below the noise level still contribute to the intelligibility [6] and compared to the SNR range of -6dB to -30dB used in other studies [7], the SNR has been kept relatively low. The expected increase in synergy was confirmed by a pilot study ($N = 6$).

Table 1: Accuracy scores with 5% confidence intervals per screen size, video, and sequence.

Factor		Mean	Confidence Interval	
			LB	UB
Screen Size	IFOV			
Small	5.501°	0.57	.49	.66
Medium	12.13°	0.64	.55	.73
Large	30.21°	0.74	.66	.82
Video				
1		0.53	.43	.62
2		0.58	.51	.66
3		0.84	.78	.91
Sequence				
First		0.66	.56	.75
Second		0.64	.55	.73
Third		0.66	.57	.74
Mean (SD)		0.65 (0.32)		

Note. LB and UB denote respectively the Lower and Upper Bound.

2.4 Procedure

The subjects were told that they were conducting a listening and a video-listening test for which they should remember as much as possible from the video. Furthermore, they were told to sit still and keep their head stable. They were informed about a video camera that the experiment leader used to inspect the proper participation in the experiment. The total duration of the experiment was 30 minutes.

The experiment consisted of the following four phases: i) some questions concerning general demographic data were asked, namely: name, sex, age, occupation, and nationality; ii) the English listening test was assessed [3]; iii) three videos in three different screen sizes were shown in one of the 36 possible orders. Each video was followed by two multiple-choice questions to test for intelligibility; and iv) some questions were asked concerning the participant's experience with the experiment.

3 RESULTS

As expected, the SNR reduced the intelligibility of the standardized CITO video listening test. A one-tailed t-test showed a significant difference between the average norm results per question of the CITO [3] ($N = 359$; $M = 0.85$, $SD = 0.11$) and the current accuracy scores for the large screen size ($M = 0.74$, $SD = 0.16$); $t(53) = -2.80$, $p = .007$, $\eta_p^2 = 0.129$. With a reduction of 12.94%, this shows an overall modest influence of the added noise.

The descriptive statistics of accuracy scores per screen size, video, and sequence are shown in Table 1. The accuracy scores were normalized to a scale of 0 to 1, where 1 means all questions were answered correctly. The effect of screen size, video, and sequence on accuracy score were analyzed using a Multivariate Analysis of Variance (MANOVA) .

The MANOVA of accuracy score by screen size showed that screen size significantly influences intelligibility, $F(2, 135) = 4.60$, $p = .012$, $\eta_p^2 = .064$. Furthermore, the correlation between screen size and accuracy score was $r(160) = .213, p = .007$. Post-hoc Bonferonni comparisons for the effects of screen size on accuracy score revealed no significant results on the comparison between small (s) and medium (m) ($\Delta(s, m) = 0.07$, $SE = 0.06$, $p = .734$) and medium (m) and large (l) ($\Delta(m, l) = 0.10$, $SE = 0.06, p = .205$). The difference between and small (s) and large (l) was significant ($\Delta(s, l) = 0.17$, $SE = 0.06$, $p = .009$).

The three videos differed significantly in difficulty, as was also revealed by the MANOVA , $F(2, 135) = 18.36$, $p < .001$, $\eta_p^2 = .214$. The influence of sequence was non-significant, indicating that there was no learning effect within the different trials that each subject performed. Furthermore, the English level as tested with the standardized listening test did not correlate with accuracy scores on the video test. This indicates that differences in the level of English did not influence intelligibility.

4 DISCUSSION

In line with Sumby and Pollack [1954], the main hypothesis of this study stated that the intelligibility of a message presented visually as well as auditory reduces when the screen size is reduced. This was confirmed by a significant difference in accuracy scores on the standardized video-listening test for three different screen sizes, indicating that screen size is indeed an influential factor in intelligibility. Hence, through utilizing standardized intelligibility tests, this research specifies and quantifies a fundamental constraint that small screens place on the advantages of bimodal perception.

The effect of screen size appears to be robust, showing a quite consistent and gradual increase in synergy with an increase in screen size. Thus, even when it is possible to reduce the distance to the screen, the synergy is still likely to benefit from a larger screen size. This result does ask for further research, as to find the threshold above which an increase in FOV does not further increase bimodal synergy. In addition, this threshold might be different for variables such as emotional connectedness [12, 28], one of the key uses of mobile video telephony [18], and possibly of mobile television and mobile games.

The experiment conducted revealed an influential factor of mobile video technology: the limited synergy of audio and video with small screens. The presented constraint shows one of the possible reasons for the absence of a large scale success of mobile video telephony and supplied evidence for one of the possible threats to mobile television. But it also shows the potential of improving the QoS of mobile video technology, whilst using the same bandwidth. Multimodal synergy has the potential to alleviate auditory and visual issues that emerged in parallel with mobile technology [10, 29, 31]. And, when enhancing and highlighting auditory, visual, and bimodal features that benefit multimodal synergy, this opportunity comes free of charge.

This study adds to a body of research that predominantly shows the effects of big screens on a range of psychological variables. It specifies and quantifies the constraint that small screens place by showing the effects of screen size on multimodal synergy. As such, this study places a fundamental but commonplace constraint on basic human multimodal perception in the context of the field of mobile video technology. This constraint shows both the vulnerability and strength of mobile video technology: When the constraints are met, the mobile user experience can fully benefit from the potential bimodal advantages.

ACKNOWLEDGMENTS

This work has been done within the scope of the NWO IPPSI-KIEM project "Adaptive Text Mining: Information in the Eye of the Beholder" (643.000.002). The CITO [3], in particular Jan van Thiel, is gratefully acknowledged for their generous cooperation in selecting and, subsequently, preparing suitable video-listening tests. In addition, we thank Ronald van Eijk, Johan de Heer, and Sorin Iacob for their contributions to this study. Last, we thank all subjects for their voluntary participation in this study.

REFERENCES

[1] J. R. Banbury. 1983. Wide field of view head-up displays. *Displays* 4, 2 (1983), 89–96.
[2] Y. Cao, F. van der Sluis, M. Theune, R. op den Akker, and A. Nijholt. 2010. Evaluating informative auditory and tactile cues for in-vehicle information systems. In *Proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI'10)*, A. K. Dey, A. Schmidt, S. Boll, and A. L. Kun (Eds.). New York, NY, USA: ACM, Pittsburgh, PA, USA, 102–109.
[3] CITO. 2018. http://www.cito.com/. [Last accessed on June 15, 2018]. (2018).
[4] C. Cobârzan, K. Schoeffmann, W. Bailer, W. Hürst, A. Blažek, J. Lokoč, S. Vrochidis, K. U. Barthel, and L. Rossetto. 2017. Interactive Video Search Tools: A detailed analysis of the video browser showdown 2015. *Multimedia Tools Applications* 76, 4 (2017), 5539–5571.
[5] B. de Gelder and J. Vroomen. 2000. The perception of emotions by ear and by eye. *Cognition & Emotion* 14, 3 (2000), 289–311.
[6] R. Drullman. 1995. Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level. *The Journal of the Acoustical Society of America* 98, 3 (1995), 1796–1798.
[7] N.P. Erber. 1975. Auditory-visual perception of speech. *Journal of Speech and Hearing Disorders* 4, 40 (1975), 481–492.
[8] H. Frowein, G. Smoorenburg, L. Pyters, and D. Schinkel. 1991. Improved speech recognition through videotelephony: Experiments with the hard of hearing. *IEEE Journal on Selected Areas in Communications* 9, 4 (1991), 611–616.
[9] M.E. Grabe, M. Lombard, R.D. Reich, C.C. Bracken, and T.B. Ditton. 1999. The role of screen size in viewer experiences of media content. *Visual Communication Quarterly* 6, 2 (1999), 4–9.
[10] S. Harper, Y. Yesilada, and T. Chen. 2011. Mobile device impairment ... similar problems, similar solutions? *Behaviour & Information Technology* 30, 5 (2011), 673–690.
[11] J. Hess, H. Knoche, and V. Wulf. 2014. Thinking beyond the box: Designing interactive TV across different devices. *Behaviour & Information Technology* 33, 8 (2014), 781–783.
[12] J. H. Janssen, P. Tacken, J. J. G. de Vries, E. L. van den Broek, J. H. D. M. Westerink, P. Haselager, and W. A. IJstelstein. 2013. Machine beats human emotion recognition through audio, visual, and physiological modalities. *Human Computer Interaction* 28, 6 (2013), 479–517.
[13] C. Johnson and P. Grainge. 2015. *Promotional screen industries*. Oxon, OX, UK: Routledge / Taylor & Francis Group.
[14] Y. Jung, B. Perez-Mira, and S. Wiley-Patton. 2009. Consumer adoption of mobile TV: Examining psychological flow and media content. *Computers in Human Behavior* 25, 1 (2009), 123–129.
[15] K. J. Kim. 2017. Shape and size matter for smartwatches: Effects of screen shape, screen size, and presentation mode in wearable communication. *Journal of Computer-Mediated Communication* 22, 3 (2017), 124–140.
[16] C. Kühnel. 2012. *Quantifying Quality Aspects of Multimodal Interactive Systems*. Berlin/Heidelberg, Germany: Springer-Verlag.
[17] Microsoft. 2018. About DirectShow. URL: [Last accessed on March 16]. (2018).
[18] K. O'Hara, A. Black, and M. Lipson. 2006. Everyday practices with mobile video telephony. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, R. Grinter, T. Rodden, P. Aoki, E. Cutrell, R. Jeffries, and G. Olson (Eds.). ACM Press, Montreal, Quebec, Canada, 871–880.
[19] S. Oviatt, R. Coulston, S. Tomko, B. Xiao, R. Lunsford, M. Wesson, and L. Carmichael. 2003. Toward a theory of organized multimodal integration patterns during human-computer interaction. In *Proceedings of the 5th International Conference on Multimodal interfaces (ICMI-PUI'03)*, S. Oviatt, T. Darrell, M. Maybury, and W. Wahlster (Eds.). New York, NY, USA: ACM, Vancouver, BC, Canada, 44–51.
[20] H. Powell. 2017. *Always On: Mobile culture and its temporal consequences*. London, UK: World Scientific Publishing Europe Ltd, Chapter 4, 99–117.
[21] A.N. Rimell, N.J. Mansfield, and D. Hands. 2007. The influence of content, task and sensory interaction on multimedia quality perception. *Ergonomics* 51, 2 (2007), 85–97.
[22] S. Shaheen, A. Cohen, and E. Martin. 2017. *Smartphone app evolution and early understanding from a multimodal app user survey*. Cham, Switzerland: Springer International Publishing AG, Chapter 10, 149–164.
[23] L. Shoukry and S. Gobel. [in press]. Reasons and Responses: A Multimodal Serious Games Evaluation Framework. *IEEE Transactions on Emerging Topics in Computing* ([in press]).
[24] S. Srinivasan, P. J. Hsu, T. Holcomb, K. Mukerjee, S. L. Regunathan, B. Lin, J. Liang, M.-C. Lee, and J. Ribas-Corbera. 2004. Windows media video: Overview and applications. *Signal Processing: Image Communication* 19, 9 (2004), 851–875.
[25] W.H. Sumby and I. Pollack. 1954. Visual Contribution to Speech Intelligibility in Noise. *The Journal Of The Acoustical Society Of America* 26, 2 (1954), 212–215.
[26] N. Tractinsky. 2004. Tools over solutions? Comments on Interacting with Computers special issue on affective computing. *Interacting with Computers* 16, 4 (2004), 751–757.
[27] E. L. van den Broek. 2005. *Human-Centered Content-Based Image Retrieval*. Ph.D. Dissertation. Nijmegen Institute for Cognition and Information, Radboud University Nijmegen.
[28] E. L. van den Broek. 2011. *Affective Signal Processing (ASP): Unraveling the mystery of emotions*. Ph.D. Dissertation. Human Media Interaction (HMI), Faculty of Electrical Engineering, Mathematics, and Computer Science, University of Twente, Enschede, The Netherlands.
[29] E. L. van den Broek. 2017. ICT: Health's best friend and worst enemy?. In *BioSTEC 2017: 10th International Joint Conference on Biomedical Engineering Systems and Technologies, Proceedings Volume 5: HealthInf*, E. L. van den Broek, A. Fred, H. Gamboa, and M. Vaz (Eds.). Porto, Portugal: SciTePress – Science and Technology Publications, Lda., Porto, Portugal, 611–616.
[30] F. van der Sluis, E. L. van den Broek, R. J. Glassey, E. M. A. G. van Dijk, and F. M. G. de Jong. 2014. When Complexity becomes Interesting. *Journal of the American Society for Information Science and Technology* 65, 7 (2014), 1478–1500.
[31] K. Škařupová, K. Ólafsson, and L. Blinka. 2016. The effect of smartphone use on trends in European adolescents' excessive Internet use. *Behaviour & Information Technology* 35, 1 (2016), 68–74.
[32] G. Westheimer. 1979. The spatial sense of the eye. Proctor lecture. *Investigative Ophthalmology & Visual Science* 18, 9 (1979), 893–912.
[33] P. C. Yuen, Y. Y. Tang, and P. S. P. Wang. 2002. *Multimodal interface for human-machine communication*. Series in Machine Perception and Artificial Intelligence, Vol. 48. River Edge, NJ, USA: World Scientific Publishing Co. Pte. Ltd.