The Potential of the Internet for Music Perception Research:
A Comment on Lab-Based Versus Web-Based Studies

HENKJAN HONING
Music Cognition Group, ILLC, Universiteit van Amsterdam

OLIVIA LADINIG
Music Cognition Group, ILLC, Universiteit van Amsterdam

ABSTRACT: While the discussion on the integrity of data obtained from Web-delivered experiments is mainly about issues of method and control (Mehler, 1999; McGraw et al., 2000; Auditory, 2007), this comment stresses the potential that Web-based experiments might have for studies in music perception. It is argued that, due to some important advances in technology, Web-based experiments have become a reliable source for empirical research. Next to becoming a serious alternative to a certain class of lab-based experiments, Web-based experiments can potentially reach a much larger, more varied and intrinsically motivated participant pool. Nevertheless, an important challenge to Web-based experiments is to control for attention and to make sure that participants act as instructed; Interestingly, this is not essentially different from experiments that are performed in the laboratory. Some practical solutions to this challenge are proposed.

WEB-BASED experiments are not novel. Since the availability of the Internet several initiatives have been developed using it as an alternative to lab-based experiments (Birnbaum, 2004; Johnson-Laird & Savary, 1999; Klauer et al., 2000; Munsch & Reips, 2000). However, in the domain of vision and audition the potential of Web-based experiments is still little used, if not simply avoided. There are at least two reasons for this. First, there is some doubt on how much control there is over the participant population and their sampling, as compared to orthodox experiments (cf. Johnson-Laird & Savary, 1999, p. 221). In orthodox experiments much biographical data is available about the participants, while in Web experiments, it is argued, there is no such control, potentially allowing participants to 'conspire' to generate the date one needs (Mehler, 1999). A second reason for some conservatism in doing Web experiments is the issue of replicability. Especially in the fields of experimental psychology and psychophysics there are serious concerns about the (apparent) lack of control one has in Web experiments as opposed to those performed in the laboratory. Where in the lab most relevant factors, including all technical issues, are under control of the experimenter (i.e. have a high internal validity; Campbell & Stanley, 1963) it is argued that Web experiments lack this important foundation of experimental psychology (Mehler, 1999). As a result of the first issue, it often proves to be problematic to convince University Review Panels to give permission when there is little insight in the environment in which participants tend to do these experiments (Auditory, 2007). As a result of the second issue, some high-impact journals made it a policy decision not to publish Web-based studies, as such discouraging Web experiments to be performed.

Skeptics of Web-based studies are mainly concerned with the question of how sure an experimenter can be that participants do not conspire towards a certain result, or in other ways try to deceive the purpose of the experiment (Mehler, 1999, p. 188). However, it is not clear what would motivate participants to deceive. In a laboratory experiment as well, a typical paid participant could well, for example, just press buttons and take little care in doing the instructed task. Actually, it can be an advantage that there is no experimenter present, because having participants completing tasks in spite of their anonymity, which would make it easy at any point to drop out, can be considered a valuable sign of

motivation. The absence of the experimenter also minimizes the performance according to social desirability, and eliminates possible experimenter biases or Pygmalion effects.

Nevertheless, in a Web experiment—just like in an orthodox experiment—one has to make sure the participants are doing what you asked them to do. One way of solving this issue is to ensure there is little reason for the participant to deceive.[1] Make the experiment challenging and fun to do, do not reward good answers (but simply participation), and make certain the participants feel involved. In music perception research this turns out to be relatively easy. Music lovers tend to like listening experiments and are usually very motivated, resulting in large numbers of responses (see Honing, 2006; 2007; Honing & Ladinig, 2006; under revision).[2] Furthermore, to make sure that a potential participant (who is typically sitting at home behind a computer screen) gets involved in the experiment, one can, for instance, use a *screen cast* [3]: a video showing what to expect and that presents the instructions in a compelling way. It generally makes the participants feel more involved and motivates them in really taking note of the instructions, more than when they are simply asked to read text from the screen. And finally, as an extra incentive, we often use a raffle of gift certificates among all respondents, independent of their responses.

However, in Web experiments there is always a participant group that is either just curious, didn't intend to do the full experiment, or that is simply not serious. Hence, one of the tasks of the Web experimenter is to distinguish between serious and unserious responses. Dropout of non-serious participants (typically around 30-40%; Reips, 2002) includes people that did not finish the experiment, did not fill-in all required information, did it too quickly (e.g., didn't listen through the full sound examples when instructed as such), or those that did the experiment too hastily. By, for instance, inserting foils (e.g., a question to test whether a participant is paying attention), consistency checks, and time logs in the experiment (e.g., using Javascript) one can easily check for this and filter out the serious from the non-serious responses.

To minimize the dropout of serious respondents, it is generally a good idea to (1) put a platform, browser and audio check first (to avoid frustration about possible technical problems later in the experiment, 2) put biographical questions at the beginning of the Web experiment (they tend to make the respondent feel more involved), followed by 3) engaging instructions (making sure the participants know what to expect), and 4) the actual experiment (that should not last longer than about fifteen minutes). In addition, it is important to give feedback about the results and the research context (what was the relevance of the experiment), along with the question of whether the respondent would like to participate in future experiments. The latter turns out to be a good index of 'seriousness', as well as an opportunity to build up a motivated participant pool that has the appropriate technical setup and a sincere interest in the topic area one is studying.

While for more socially oriented research the Web was used early on (Musch & Reips, 2000), for vision, audition and music perception studies the internet is only sparsely used. The main reason for this is, next to the issues discussed above, of a technical nature. One could, generally, not be sure of the video and sound quality at the user-end. However, several recent studies (e.g., Krantz, 2000; McGraw at al., 1999) showed that these technicalities are currently less of an issue. As an example, McGraw et al. (1999) found that, for some classical experiments in the visual domain, Web-based experiments are able to replicate the results obtained in the laboratory. They concluded that "existing technology is adequate to permit Web delivery of many cognitive and social psychological experiments [..] The added noise created by having participants in different settings using different computers is easily compensated for by the sample sizes achievable with Web delivery." (McGraw et al., 1999, p. 502). With comparable advances in audio presentation over the internet (using, for instance, file formats like MPEG4 that minimize loading time and guarantee optimal sound quality on different computer platforms at different transmission rates) there is little reason to think that Web experiments in music perception would be less reliable. Especially when the aim is to study music perception (as opposed to some psychophysical task that might have additional demands to the presentation of audio), the Web can even be preferred over lab-based experiments.

Web-based experiments have a much greater external validity as compared to lab-based experiments. While this sometimes results in losing some internal validity (cf. Auditory, 2007), in music perception studies this might actually be desirable: i.e., to have listeners respond in an environment in which they normally listen, including e.g. its noisiness and/or use of low-quality headphones or loudspeakers. For such a listener a lab situation—with high-quality audio, a soundproof booth, focused listening, and the pressure of having to perform—might be quite unnatural. In short, having listeners take part from their home in an arguably more natural environment and, as such, being less stressed, might actually positively influence the ecological validity of the results.

Another criticism of Web-based studies is that of the sample of participants. One can argue that participants with access to Internet belong to a special, technologically versatile, subgroup of the population, and that this could affect the representativity of a certain result. While this is changing rapidly (especially in the Western countries), it should be noted that such a (potential) restriction on the sample also applies to the typical psychology-student pool that biases most studies in psychology. As such this is not a specific drawback of Web-based studies.

In conclusion, while there are still plenty of challenges to online data collection, we believe that Web experiments generally do not generate more problems than an orthodox lab-based experiment. It might actually be a rich source for more ecologically valid and truly engaging studies in music perception. In the end it is up to the researcher to decide, and argue for each specific case, where to draw the line with regard to the trade-off between a higher *internal validity* of laboratory settings compared to a higher external validity of Web-based setups.

NOTES

[1] For most of the observations made in this comment, empirical support is (e.g., Reips, 2002) or can be made available. However, due to space restrictions we will not do this here.

[2] For the URLs to the online listening experiments on which these studies are based, see the reference section.

[3] For an example, see URL http://www.musiccognition.nl/e4/

REFERENCES

Auditory (2007). *Online listening tests and psychoacoustic experiments with large N.* Discussion on Auditory mailing list, see http://www.auditory.org/mhonarc/2007/threads.html#00531.

Birnbaum, M.H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, Vol. 55, pp. 803 -832.

Campbell, D.T., & Stanley, J.C. (1963). *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally.

Honing, H. (2006). Evidence for tempo-specific timing in music using a Web-based experimental setup. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 32, pp. 780-786. [See http://www.hum.uva.nl/mmm/exp2/]

Honing, H. (2007). Is expressive timing relational invariant under tempo transformation? *Psychology of Music*, Vol. 35, pp. 276-285. [See http://www.hum.uva.nl/mmm/exp1/]

Honing, H., & Ladinig, O. (2006). The effect of exposure and expertise on timing judgments: Preliminary results. *Proceedings of the International Conference on Music Perception and Cognition* (pp. 80-85). Bologna: Italy. [See http://www.hum.uva.nl/mmm/exp3/]

Honing, H., & Ladinig, O. (Under revision). Exposure influences expressive timing judgments in music. *Journal of Experimental Psychology: Human Perception and Performance.*
[See http://www.hum.uva.nl/mmm/exp3/]

Johnson-Laird, P.N., & F. Savary (1999). Illusory inferences: A novel class of erroneous deductions. *Cognition*, Vol. 71, pp. 191-229.

Klauer, K.C., Munsch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, 107, 852-884.

Krantz, J.H. (2000). Tell me, what did you see? The stimulus on computers. *Behavior Research Methods, Instruments, and Computers*, 32, 221-229.

McGraw, K.O., Tew, M.D., & Williams, J.E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, 11, 502-506.

Mehler, J. (1999). Editorial. *Cognition*, 71, 187-189.

Musch, J. & Reips, U. D. (2000). A brief history of Web experimenting. In M . Birnbaum (Ed.) *Psychological experiments on the Internet* (pp. 61-87). San Diego, CA: Academic Press.

Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-114). San Diego, CA: Academic Press.

Reips, U. D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49, 243-256.