A Survey on the Acceptance of Listening Context Logging for MIR Applications

Sebastian Stober, Matthias Steinbrecher & Andreas Nürnberger
Outline

- Survey Design & Context
- Survey Results
- Correlations?
- Conclusions
Motivation: What?

- conext: [Dey 2001]
  - any information that can be used to characterize the situation of a person, place or object of consideration

- four types of primary context [Dey & Abowd 2000]:
  - location
  - identity
  - time
  - activity
Motivation: Why?

- valuable information for building personalized and user-adaptive MIR applications
- e.g. learn personalized genres like “music I like when I do programming”
Motivation: How?

Possibilities for Automatic Context Logging:
- music metadata
- ambient noise
- GPS position
- keyboard & mouse events per minute
- currently running applications
- facial expression
- bio-information
- ambient light
- status
Question

How much context is too much?

or

Where do we encounter privacy issues?
Survey Design & Context

- 8 questions
- 4 groups:
  1. demographic information
     - gender, age & country of residence
  2. general relation to music
  3. use of (web-) applications that collect, access and expose to some extend private data of their users
  4. acceptance of logging information about the listening context

- paper questionnaire at CeBIT 2009
  - in German only, 156 participants
- online survey from March till June 2009
  - in English and German, 305 participants
### Demographic Information

#### Table 1. Countries with more than 5 participants in the survey

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of Participants</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>323</td>
<td>70.07%</td>
</tr>
<tr>
<td>USA</td>
<td>24</td>
<td>5.21%</td>
</tr>
<tr>
<td>Austria</td>
<td>14</td>
<td>3.04%</td>
</tr>
<tr>
<td>France</td>
<td>10</td>
<td>2.17%</td>
</tr>
<tr>
<td>Turkey</td>
<td>7</td>
<td>1.52%</td>
</tr>
<tr>
<td>Switzerland</td>
<td>6</td>
<td>1.30%</td>
</tr>
<tr>
<td>Spain</td>
<td>5</td>
<td>1.08%</td>
</tr>
</tbody>
</table>

(countries with at least 5 participants)

The last topic represents the main question of the survey whereas the others were added to be able to estimate a possible bias of the participants. Furthermore, correlations between the background of the participants especially in relation to music were expected that could be identified.

The survey was conducted in two parts: Between March 3rd and March 10th, a paper questionnaire in German was filled out by visitors of the German CeBIT. Based on this questionnaire, an online questionnaire was designed afterwards to extend the scope of the survey. It was open to the public from March 11th until June 12th. The questions of the online questionnaire were identical to those of the paper version. However, the questions were split across multiple screens and an English translation was added for international participants. 48 persons filled out the online questionnaire, resulting in 510 participants in total.

4.1 Demographic background of the participants

From the 510 participants of the survey, 267 were female and 243 were male. 2 persons did not answer this question. The average age was 29.39 with a standard deviation of 7.17. Table 1 shows the countries with more than 5 participants in the survey. In total, 384 persons from 10 countries participated, whereas 126 persons did not state their country of residence.

The majority of the participants was from Germany. This is primarily due to the fact that part of the survey was conducted amongst visitors of the German CeBIT. Furthermore, the survey was advertised at the Otto-von-Guericke-University. Thus, it can be assumed that there are many German students amongst the participants. Most of the international participants were probably informed about the survey by the announcements posted on popular MIR mailing lists such as the music-ir list maintained by IRCAM.
General Relation to Music

- **Please select all statements applicable for you!**

  - I use a mobile music player.
    - 66%
  - I am professionally involved with music.
    - 20%
  - I make music.
    - 40%
  - I have a large music collection.
    - 70%
  - I am very picky about the music I listen to.
    - 53%
  - I mostly listen to (internet-) radio stations.
    - 38%

- **How frequently do you listen to music?**

  - Frequently
  - Regularly
  - Occasionally
  - Rarely
  - Occasionally
  - Permanently
  - No answer
Use of (Web-) Applications…

… that collect, access and expose to some extend private data of their users

- Do you use the following (or comparable) applications?

- facebook / StudiVZ
- Gmail
- last.fm
- LinkedIn / XING
- flickr
- Google Docs
- wordpress / blogger
- delicious
- twitter
- wakoopa

![Bar chart showing usage frequencies for the listed applications.](chart.png)
Acceptance of Listening Context Logging

Would you allow your music player (as software or as a self-contained device) to log the following information in order to enable it to learn personalized genres for sorting your music collection?

- music metadata
- ambient light
- mouse & keyboard
- ambient noise
- twitter / IM status
- applications in use
- facial expression
- bio signals
- GPS position

Yes (unconditionally)  Yes, but only on my device  Yes, but only anonymized  Maybe  No
Looking for Correlations: Preprocessing

- conversion of answers for different logging options into numerical score:
  - unconditional acceptance = 2
  - conditional acceptance = 1
  - maybe = 0.5
  - no = 0

- sum of scores for different logging options:
  - values between 0 and 18

- binning into 6 equidistant groups as indicators for general acceptance
Looking for Correlations: Country

- USA: considerably fewer reservations against collecting context information
Looking for Correlations: Gender & Age

- majority: male persons between 20 and 30
- < 20: most drastic difference in distribution by gender
- < 20: surprisingly reluctant
Looking for Correlations: Relation to Music

(number of statements chosen)

- no significantly differing subgroups
  (neither for the aggregated affinity to music nor for the individual relations)
Looking for Correlations: Applications

- values aggregated (number of applications used):
  - low: \( \leq 2 \)
  - medium: \( 3..4 \)
  - high: \( 4..6 \)
  - very high: \( > 6 \)

- more applications used / involvement in online communities ➔ higher acceptance
Reasons

- concerns about privacy
- against any kind of data collection
- no logging unless for visible benefit
  - doubts about relevance of collected information
- fear of being patronized by player’s “intelligence”
- misuse of information for marketing purposes
- requirement of additional storage and processing power
  - increase of costs for hard- and software
- information leakage through hacking
Conclusions

- users must be:
  - fully informed about
    - when and
    - what information is logged,
    - where it is stored and
    - who has access to it
  - in full control of
    - the logging process
    - the adaptation of the MIR system
THANK YOU

Questions?