Summary of P1 and P2

Summary of P1

Yea-Nay Personalized Radio offers satellite radio listeners a higher level of control over their music choices. The user population for this product was determined to be North-American, English speaking car owners or drivers who are willing and able to purchase a pre-installed satellite radio system. The tasks the user will want to perform on this system are: create a personalized listening profile, listen to music, adjust the listening environment, and vote on songs being played to enhance the user's profile.

The existing systems looked at were satellite radio, Music Match Music Discovery Engine, and iTunes. It was decided that the Yea-Nay radio system would be attached to an existing satellite radio system, so we could focus our design on the voting aspects, rather than worry about secondary music-listening tasks, such as adjusting volume. These secondary tasks are assumed to be built into the existing satellite radio.

Before coming up with design alternatives, we established important usability criteria to which the designs must adhere. The designs must be easily learnable so that the tasks involved in voting can become more automatic, so as not to distract the driver. They must also provide effective feedback, so the users know their vote was registered, and for the appropriate song. Finally, the designs must prevent errors and allow easy recovery from errors, so the users will feel comfortable casting votes, knowing that they can undo previous actions in multiple ways.

Summary of P2

Out of three informed brainstorming sessions, we came up with three strong alternative designs for our system. The iDrive controller with a rearview mirror HUD, Foot pedal controls with no display, and steering wheel control and display. We also considered using auditory feedback with any of the design choices. We also decided on the functional and non-functional requirements for our system. Some of the functional requirements are the system must allow voting, skip, multiple profiles, and preference matching. Some of the non-functional requirements are that thy system must be safe, easy to use, visible, and reliable.

At this time we also created a user scenario, which we later used to create our script for user testing. We mocked up our four best alternatives in Photoshop using real photos of the steering wheel of a car with our interface overlaid. We carefully considered the pros and cons of each of our alternatives. At this stage we also prepared a poster for our poster session. The feedback from our poster session helped us to decide to go with the steering display and control.

Evaluation Techniques/ Summary of P3

Users
Our participants were ten North American English speakers between the ages of 18 and 55. They all had at least a high school diploma. They were divided into two groups, five of which used the software prototype and five of which used the physical prototype.

The task
The tasks were to follow a script read aloud by one of the investigators and to respond using the prototype in the most appropriate way. The script told the participant how he or she wanted to vote for the song or when to change a profile. The tasks were a sampling of all of the functionality provided by the system. The actual script we used is attached to the end of this document, but here is an excerpt.

System begins playing “America Woman”

Investigator: “You like this song and would like to hear it more often. What would you do?”

User responds.

Investigator: “Normally, voting Yea on a song would result in the song playing all the way through, but for the sake of time, we’re going to move on to the next song.”

System begins playing “We Belong to the Night”

Investigator: “You do not like this song. You don’t want to hear this song now, and you want it to play less often in the future. What would you do?”

User responds.

Evaluation Plan

Our evaluation plan was made up of three parts: the software mockup, the physical mockup, and the attention task.

The Attention Task
Since we couldn’t actually have participants drive the car to test our physical mockup, we created an attention task that would require a similar amount of attention to complete while users were in a parked car working with our system. We created a PowerPoint slideshow that was a series of pictures of the Georgia Tech campus in which the target (a picture of Buzz) appears randomly. When the campus photo alone is presented, it remains on screen without a target present for 3-7 seconds. When a target is present, it remains on screen for 1-3 seconds. All durations were selected using a random number generator. There are a total of 120 slides, 60 of which presented a target. We showed the attention task on a lab and placed it on the hood of the car in front of the driver’s seat.

We chose to create the attention task using a laptop and PowerPoint as opposed to one of us doing something physically such as standing outside flashing numbered cards. We wanted it to be exactly the same for each user. We thought the potential for variability was too high if we were to engage one of the investigators in the physical task.

The Software Mockup
Group 1 was made up of five users who sat at a computer and used a software mock-up of the steering wheel. This group was able to devote their full attention to the Yea-Nay interface in order to accomplish the tasks read to them from the script by the investigator who was sitting nearby. The investigator also changed the music on another device according to the script.

Allowing the users to fully attend to the Yea-Nay interface enables us to more accurately determine whether users can understand the controls enough to operate them. We could also test the effect of the feedback provided by the display. The software mockup was created in Photoshop and HTML. It was a series of pictures of a steering wheel inside of a car with our display overlaid. This was the only testing we did where the users could see immediate feedback on their votes.

The participants who interacted with the software mockup also completed the attention task, but not at the same time as working with the prototype. We did this so that we could obtain a benchmark to measure against the group who simultaneously completed the attention task while interacting with the physical mockup.

The Physical Mockup

Group 2 was made up of five users who performed the attention task and the Yea-Nay radio task simultaneously. The users were told their *primary* task is to detect Buzz in the slideshow on the hood of the car. The users were told their secondary task was to follow the instructions given by the investigator and interact with the Yea-Nay steering wheel profile. The five users in this group were seated in the driver's seat in a parked car with the engine running. These participants performed the same tasks as those in Group 1. The script was read aloud to them by the investigator sitting in the passenger's seat. The investigator also changed the music according to the script while the other investigators either assisted or recorded data. At the end of the script, each participant answered the questions listed below.

The Questions

The post experiment interview questions did not change from P3, but here they are again as a reminder.

1. What did you think was the easiest thing to do on the interface?
2. What did you like best about the interface?
3. What did you think was the hardest thing to do on the interface? How would you make it easier?
4. What did you like least about the interface?
5. What do you think "happens" to the system when you vote "Yea" on a song?
6. How would you describe how the system works to someone who is thinking of purchasing it?

7. Do you have any suggestions for improvements?

We chose these questions because we wanted to find out a range of information from the user's preferences to the user's mental model of the system. From the first four questions, we hoped to learn about specific aspects of our design such as if the buttons were conveniently positioned, of a good size, and properly labeled. We also wanted to get a general idea of how users liked the design and were comfortable with its location. From the last three questions, we wanted to measure the user's ability to understand how the system works, or their mental model. This is a measure of the effectiveness of the system's display and the information it conveys.

Think Aloud
During both evaluations, we noted down the remarks the participants said aloud. During the evaluation of the software prototype, participants were asked to think aloud and those thoughts were recorded. During the evaluation of the physical prototype, we did not explicitly ask them to think aloud, but most seemed to anyway.

Design Rationale

Attention Task
The attention task on the laptop was originally supposed to be place on the dash inside the car. However, this location caused the screen to tilt down and reduce visibility. We were forced to place the laptop on the hood of the car, but this turned out to better serve our experiment. Drivers are used to focusing on activity outside of the car as opposed to inside the car, so placing the laptop on the hood of the car in front of the driver increased the focal point realism.

It turned out that this task accomplished what we wanted it to accomplish. It was of appropriate difficulty and required the user to keep a constant check outside the car, while achieving a level of monotony similar to typical driving. All of our in-car participants were primarily focused on not missing a Buzz. If a Buzz was missed, it was typically due to the participant watching the investigators change displays and not attributable to the difficulty of the secondary task of interacting with the Yea-Nay system.

Physical Prototype
We were forced to manually switch out the displays on the physical prototype because no one in our groups was aware or had access to any technology that was of the size we were looking for our display. Manually switching out displays is unavoidably cumbersome and pretty much

Change in arrow position was difficult for users to detect using the physical prototype.

ruled out any recognition by the participant of what the meter was displaying. The movements in the arrow were so fine that the participants could not detect the changes in their positions.

Script
We chose to provide users with a script because it was the surest way to achieve consistency and full usage of all of our system's functionality. The script told the user if he or she liked the

song and wanted to hear it now or not, but it did not tell the user which button to push. Our idea was to give the user the motivation we wanted him or her to have and then watch the actions knowing that motivation.

Dropped Timing Responses
We decided for a couple of reasons not to keep track of the time it takes a user to respond after we provide the instructions. First, we believed this data would be difficult to keep accurate. There would definitely have been some ambiguity around when to begin timing and when to stop timing. Also, we were not sure of the helpfulness of this data. We began to think it wasn't important to know how long it took someone to vote because it wasn't a true measure of uncertainty or confusion with the system.

Curtailed Song Play
We decided not to let the songs play out as normally would occur when using a real implementation of the system. This decision was due to time constraints and a desire not to bore our participants. We don't believe this had any impact on the results of our evaluation.

Scenes from the Evaluation of our Physical Prototype

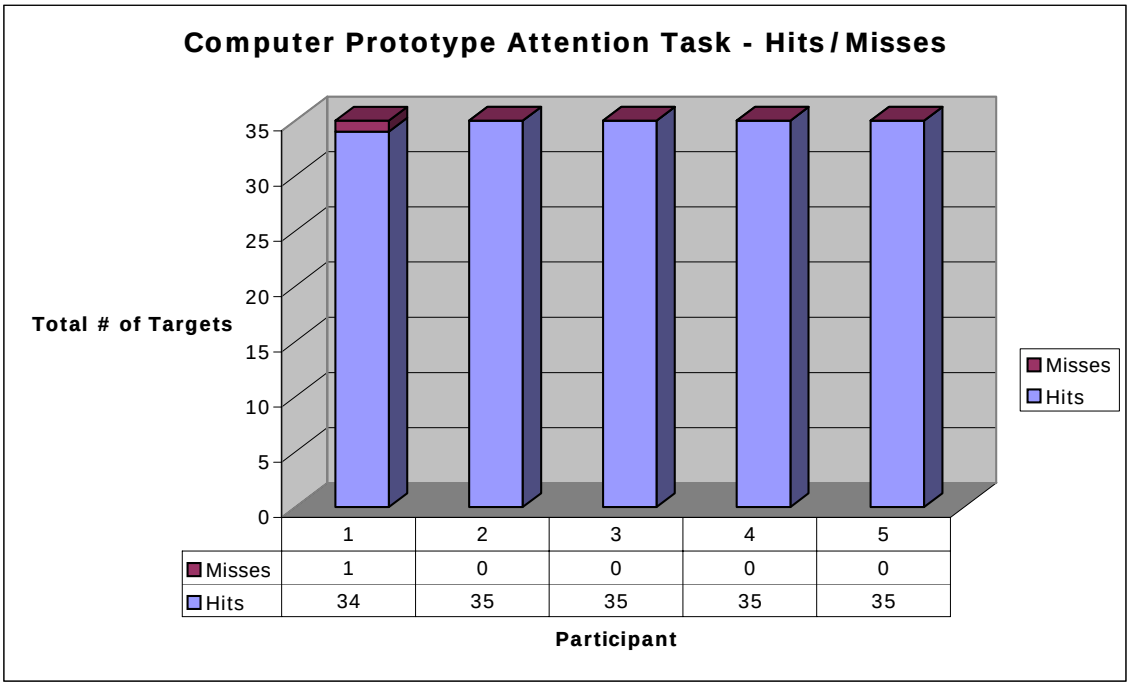
A participant in the car watching the attention task slideshow.

A participant in the car interacting with the physical prototype.

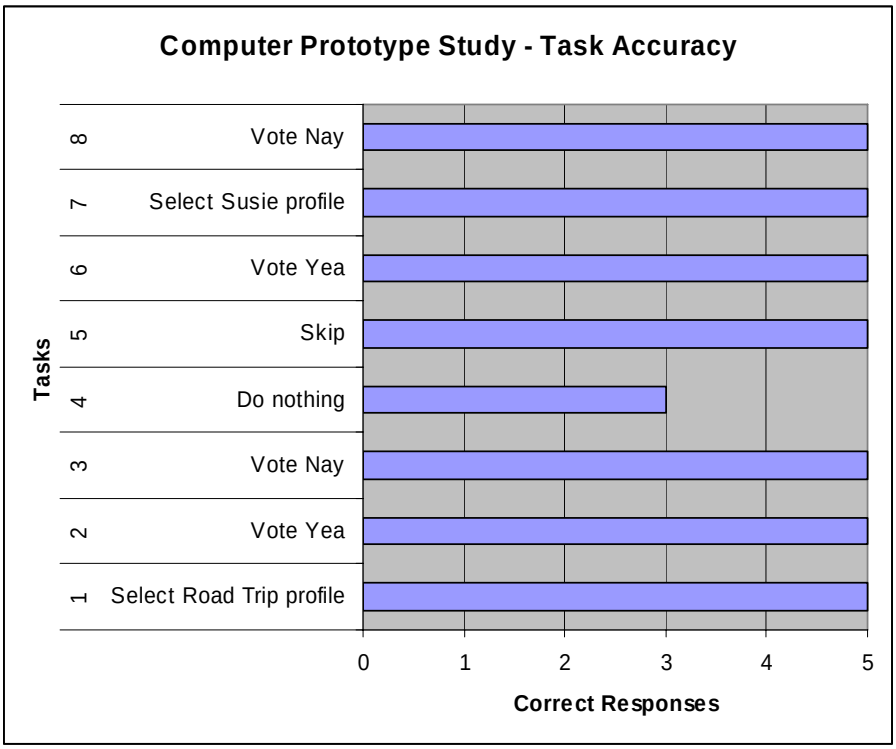
Courtney (Investigator) manually switching out the display.

Results of the Study

Computer Prototype Study Results

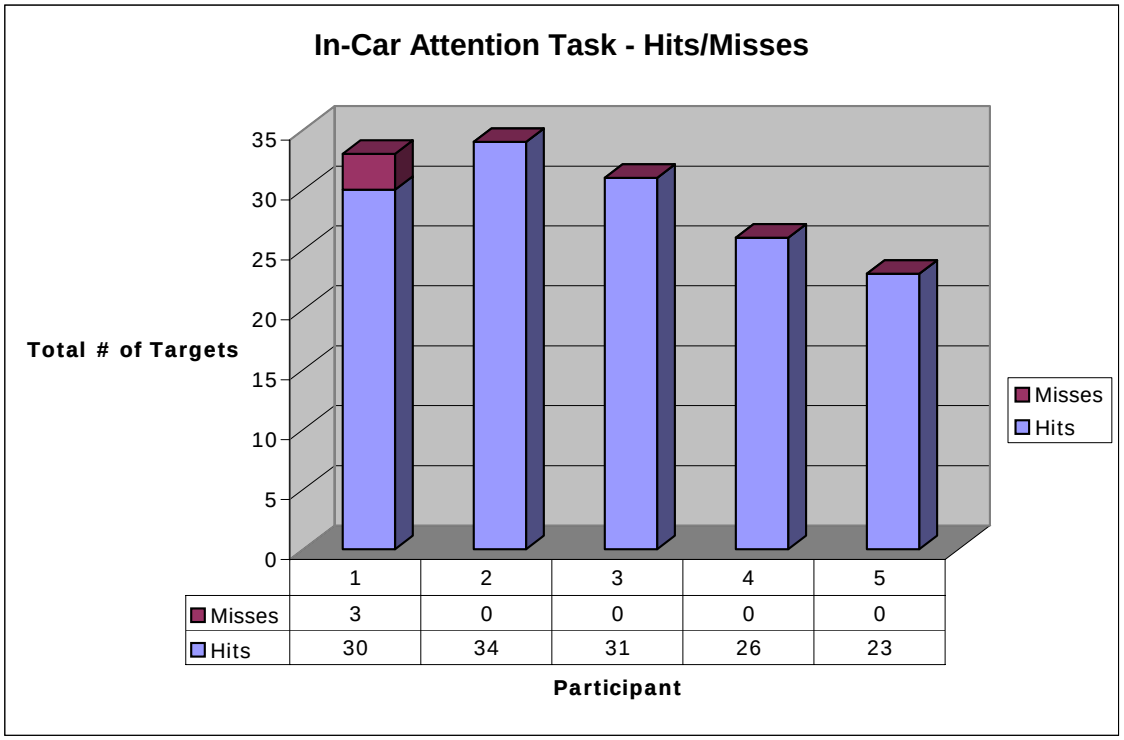


Participants were accurate in their detection of Buzz in the attention task.

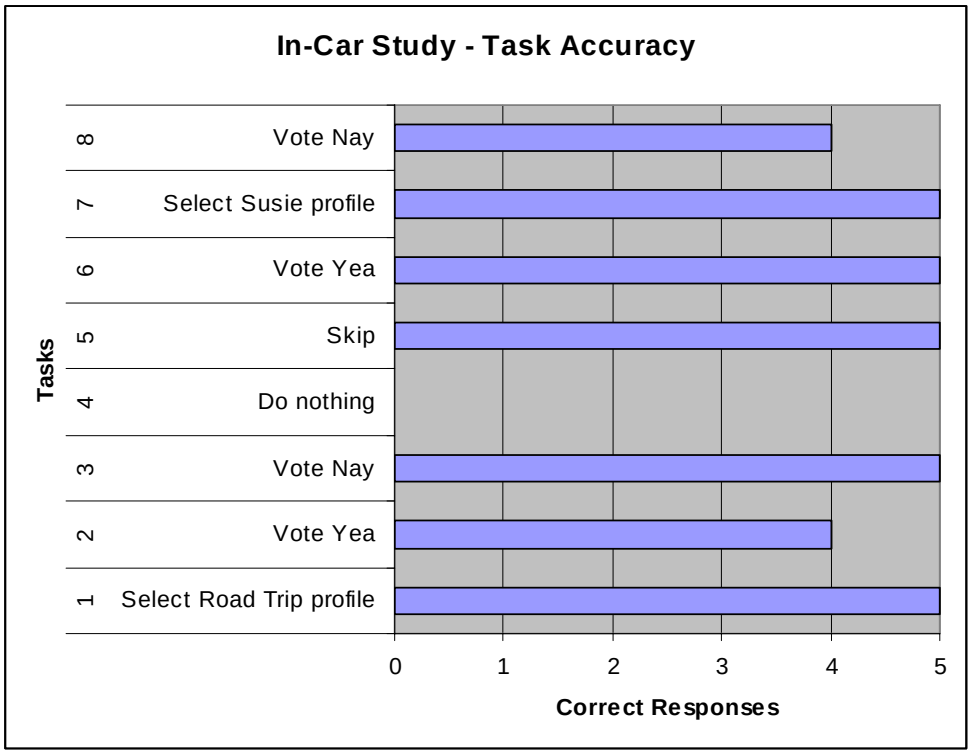


Task accuracy ranged from only three participants correctly responding (Task 4) to all participants responding correctly (all remaining tasks).

In-Car Study Results



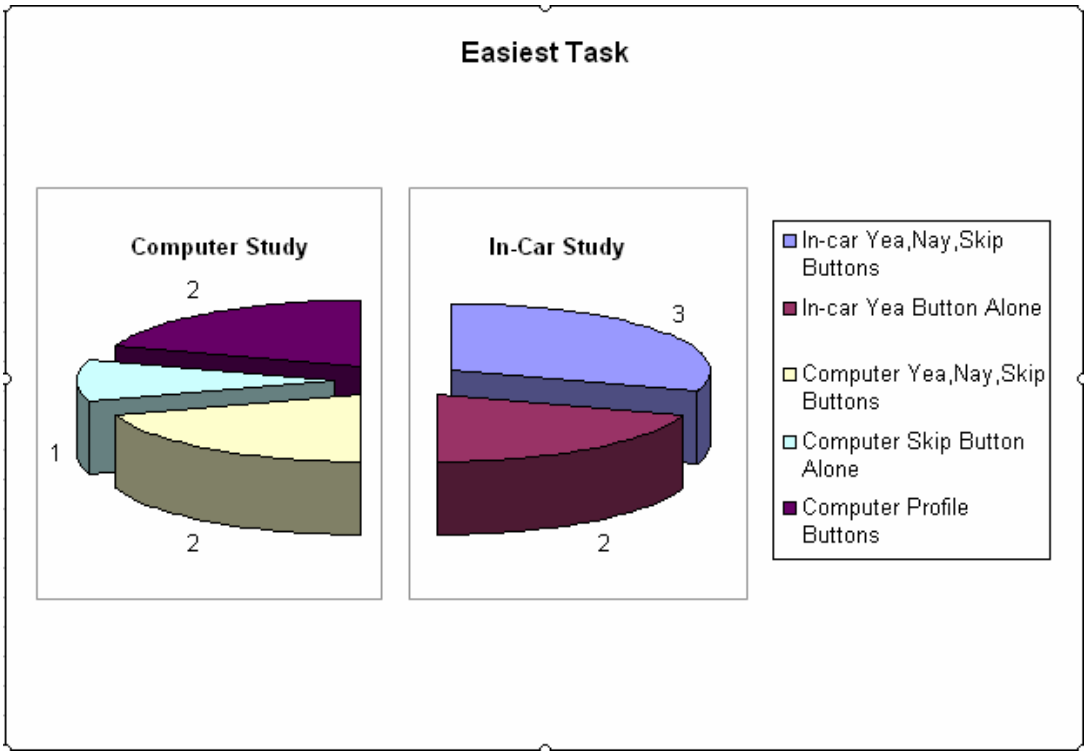
Participants were accurate in their detection of Buzz in the attention task.



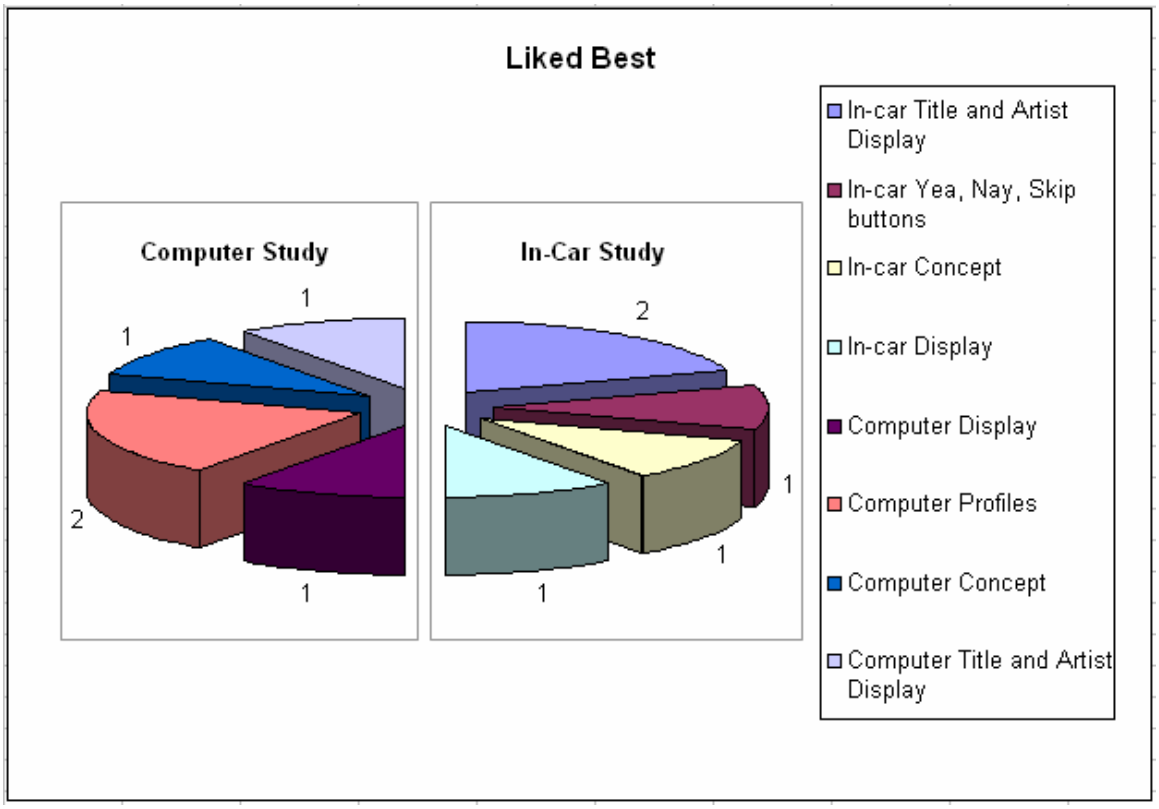
Task accuracy ranged from no participants correctly responding (Task 4) to all participants responding correctly (Tasks 1,3, 5-7).

Post Interview Analysis – Computer and In-Car Study Combined

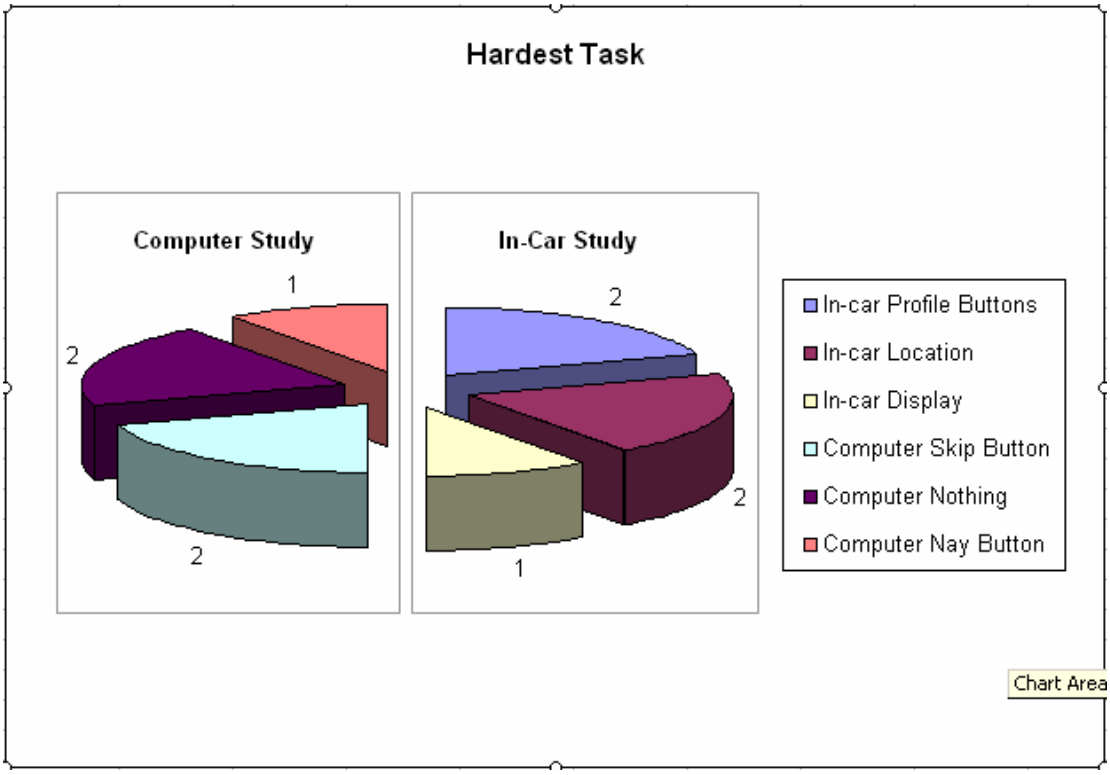
Question 1. What did you think was the easiest thing to do on the interface?



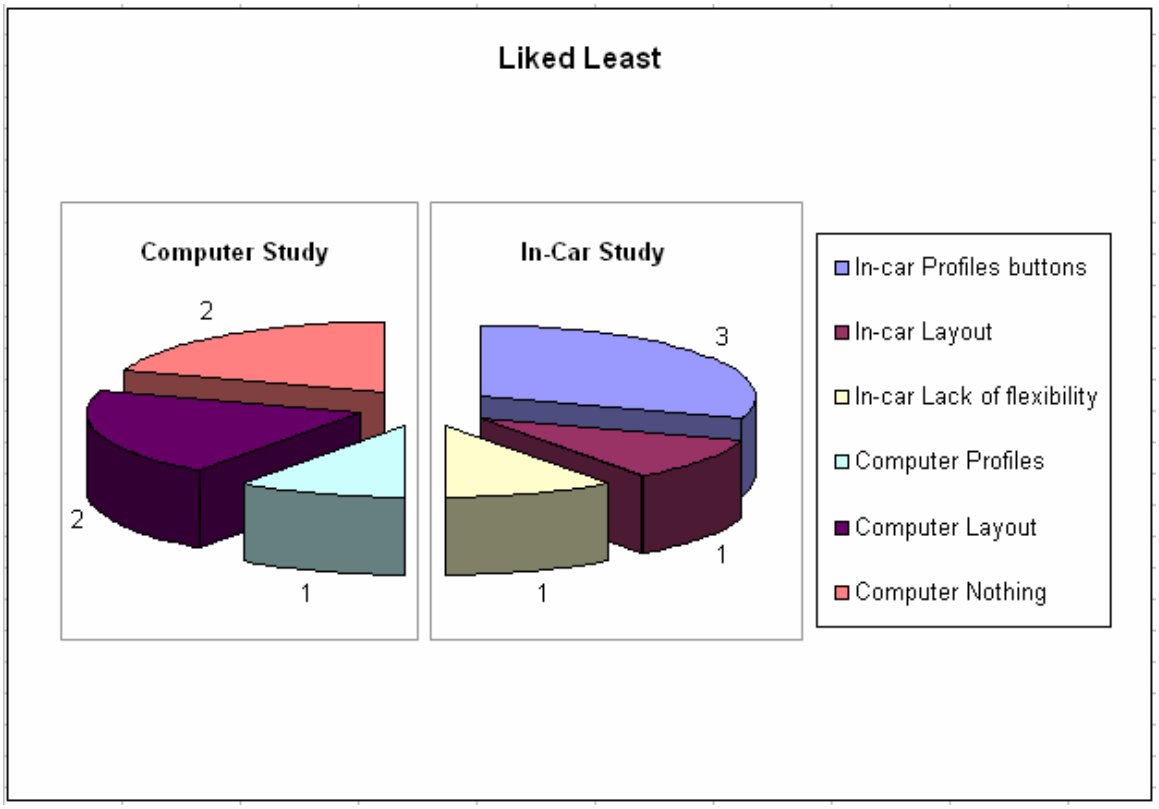
Question 2. What did you like best about the interface?



Question 3. What did you think was the hardest thing to do on the interface?



Question 4. What did you like least about the interface?



Remaining Post Interview Questions and Responses – Computer Study.

These results are so varied that they did not lend themselves to presentation in graphical format.

Question 5. What do you think "happens" to the system when you vote "Yea" on a song?

a. Records song and keeps it in a separate file
b. Logs some kind of entry about the song or genre. Could also log year info, uses some kind of program or algorithm to figure out profile
c. Moves it up on a priority list, randomizes the list, and then plays songs
d. Registers another positive response
e. Logs a vote and assigns a numerical value to that song, giving it a higher ranking so it plays more often

Question 6. How would you describe how the system works to someone who is thinking of purchasing it?

f. Neat thing on your steering wheel that keeps track of the music you prefer and you can easily play what you want to hear
g. It would be ok, but dislikes machines detecting preferences and locking into system
h. Like the customizable internet radio, in your car
i. Pretty easy steering wheel panel that allows user to choose music and frequency it will play
j. Delivers personalized targeted music. Plays songs you know and like and based upon those, sends you songs you may like which you can vote upon.

Question 7. Do you have any suggestions for improvements?

k. Make profile names bigger
l. Offer screen savers
m. Use a toggle switch for Yea/Nay/Skip. Allow for multiple votes at once.
n.

Remaining Post Interview Questions and Responses – In-Car Study.

These results are so varied that they did not lend themselves to presentation in graphical format.

Question 5. What do you think "happens" to the system when you vote "Yea" on a song?

a. Hear that song more often
b. Alters type of music you like, maybe adds a little more of that kind of music or artist to profile
c. Saves song like regular computer, puts it on random and plays songs
d. Keeps in favorites, like web browser
e. Moves song and genre up in profile

Question 6. How would you describe how the system works to someone who is thinking of purchasing it?

f. Vote on songs as they play so you wouldn't have to go through stations to find liked songs.
g. You can pick and choose the songs you like, and the ones you don't like you can get rid of permanently
h. Like downloading your own music. MP3s competition wouldn't have to buy MP3's off Internet, nor cds, placing them on specific profiles.
i. Gives random songs to begin with, and it plays them for you, you vote your preferences, and it categorizes the song and updates itself to play more of that kind every once in a while playing new songs just to see how you like them and keep them up to date.

Question 7. Do you have any suggestions for improvements?

j. Weird in steering wheel. Better where normal radio is.
k. Larger profile buttons, wider, not longer. Screen may interfere with airbags or horn. Maybe at top and airbag at bottom. Display and buttons on airbag are less distracting, and should remain.
l. No need for Yea-Nay scale
m. Everything should be in one small area, so it only takes 1 hand to control it. Would have more control with a wheel. Simpler. Move display to dash.
n. Back button. Make Yea-Nay scale clearer, because people don't like to read directions. Maybe vote history. Something clearly descriptive.

Discussion of the Results

Attention Task Results

The attention task being used to simulate a driver's concentration upon the road was easily followed by most of the participants in the study. Only one participant from each study group missed any of the targets at all. In the Computer Prototype Study, one participant missed 1 out of 35 targets. In the In-Car Study, a single participant missed 3 out of 33 targets. The participant who missed some of the targets in the in-Car Study missed most of these due to a button falling off of the prototype. All the other participants got 100% of the targets that appeared upon the screen during the time it took them to complete the scripted task. This helped in representing that the participants were following the task properly, and primarily watching the attention task, as directed.

Few issues were noticed in relation to the attention task. The first issue was the button falling of the prototype, causing reason for the user to pause during the testing. The second major issue

was from the participants attempting to keep identifying the target as it appeared, while also attempting to speak aloud their choices of buttons to push.

Accuracy of Choices
Each participant appeared to easily identify the usage of the Yea, Nay, and Profile buttons. Other then slight pauses as the participants familiarized themselves with the systems Yea, Nay, and Profile Buttons, few paused in identifying the usage of these buttons. Both the Computer Prototype and the In-Car Study helped to show easy understanding of the usage of these buttons.

The Computer Prototype Test provided information that revealed people could understand the difference between the Skip Button and the option to let a song play without a vote. During the in-car study, all the users stumbled over these portions of the script, confusing situations where they should press the Skip button with situations where they should allow a song to play. In one case, the participant pressed the Nay button instead. It is the opinion of the investigators that these incorrect responses may have occurred due to the actual wording of the instructions in the script rather than to the labels or placement of Skip Button or the concept of letting a song play through.

Follow-up Interview Answers

Display
The survey questions provided some good feedback on the system. The questions made apparent the display itself was easily understood. The main point of confusion with the display arose from the participant's understanding of the Yea-Nay Meter, which shows vote history upon a song. One complaint on the display dealt with the names of the profiles being too small. Many of the In-Car participants did not seem to fully grasp the usage of the meter. In both studies, it seemed participants questioned the need for the meter. It is believed that part of the reason that the meter was not fully understood during the in-car testing was due to the nature of the mock-up displays. With the displays being pictures that were changed out, it seemed that the users were unable to easily picture the meter moving as it changed to different songs. The listing of the artist and song title received very strong feedback, and was well liked by most of the participants.

Buttons
The Yea and Nay buttons had positive feedback from many of the participants. Most of the participants enjoyed the interaction with the song choices. One of the complaints about the Yea/Nay buttons was that they could be wider, to make them easier to press. One participant expressed a dislike for the idea that the system would record personal preferences. Several participants expressed that it was easier to use the Yea button, relating it to the fact that they were right handed. One suggestion concerning these buttons was to place them all in one spot, making it easier to access all at once.

The Profile buttons received some negative feedback. One complaint about the profile buttons related to their size. It was also expressed that the participants had to look away from the road to much to find the proper profile buttons. Most of those who expressed any dislike in the profile buttons also said that they assumed they would be able to memorize the profile positions during normal usage of the system.

System Comprehension
The feedback also showed that most of the participants had some comprehension of the system usage. Many of the users did not have a perfect idea of how the system worked, but the mental model they had formed was generally close to the way the system actually performed. The mental models created by the In-Car participants seemed to be further off than those created by the software participants. This could again be related to the environment and the fact that the nature of the physical prototype was less interactive than the software prototype.

Suggestions
The personal suggestions of the participants provided a variety of answers to the studies. There was some discomfort with the placement of the controls and the display. One participant expressed that it would get in the way of airbags. It was also expressed that the display should be moved up to the dashboard. One participant expressed the wish for the controls to be brought together in one location, and a couple of participants similarly expressed interest in a single toggle switch or a wheel for all the controls. It was even expressed by one participant that they would prefer the controls being in the position of the regular radio.

Some additional suggestions provided options to be considered that could be added to the system. One such option was to add a back button, so that a song could be repeated if the user liked it. It was also suggested to add the option to vote multiple times upon one song.

The suggestions also helped back up that the Yea-Nay meter needed clarification. It appeared that many of the participants didn't see a reason for it, or felt that their needed to be more of an understanding of what the meter's purpose was.

How Our Evaluation/Prototype Could Be Improved

Although we were able to garner much valuable feedback from our user testing and our prototypes, there are several things we would like to have done differently.

More pathways
In both the software and physical testing, the only paths we built through the system was the correct one. Therefore, the user was restricted to the correct path. For example, in the software testing, the only hot spot on each page was the button that the user should press. It's difficult to know if a user has picked up on that and let it guide their interaction. We also didn't ask that question explicitly in our post interview. In the car, this mattered a little less because we could tell which button the user was attempting to press. However, if the user pressed the wrong button and we moved on to the next step, the feedback on the display we manually replaced confirmed an action different to what the user did. This may have let the user know that action was not what we expected and may have made him or her feel bad or have changed his or her future interaction.

As an improvement, we would like to have been able to build more pathways through the system such that we could more easily know the user's first impression. It would have been helpful, especially in some of the less obvious interaction situations such as "Skip" and "Change Profiles", if the user would have pressed the buttons we expected or not.

Buttons
We received some interesting feedback on the profile buttons on each side. In the car, some users felt the profile buttons were too small. We would like to make them larger. They could also be reshaped to be wider or taller. We also could have put them all along the bottom of the screen similar to the buttons on a traditional radio.

There was one suggestion from a user that all of the buttons be on one side instead of requiring her to have both hands on the wheel. That was not something we had considered before, but may have been a good solution. Many people drive with just one hand and our system really assumes a two-handed interaction.

We would also rethink our placement of the Yea and Nay buttons. We placed the Nay button on the left and the Yea button on the right. Our rationale was that we felt like most people use their right hand to drive, which would make the Yea vote (theoretically the most common vote) more convenient. However, this may not be intuitive to the user. Since we read left to right in this

country, and one would expect the Yea vote to come first due to the name of the system and convention, it may be a better design choice to put Yea on the left and Nay on the right.

Another suggestion we would like to have incorporated is changing the look of the Skip button in some way. Since the functionality of this button is different from the Yea and Nay buttons in the sense that it is not logging a vote, it makes sense that the button should look different.

Other suggestions were a back button that would allow a recovery from Skip or a way for the user to hear a song again if it is one they really like. This suggestion was a good one because it gives users a way to rapidly change the status of a song in their playlist if desired. After hearing a new song the user likes, he could hit the back button to hear it again and log another vote. This could also be used to send a new song the user hates into oblivion very quickly.

Display (not including meter)
Most users found the display to be very effective. They especially liked being able to see the title and artist. We did receive criticism on the size of the profile names during the in-car evaluation. We would like to have made those bigger. We aren't sure if the users could distinguish between the active profile and the other profiles because we failed to ask that question. We didn't receive any criticism on our choice of colors.

At first, we hesitated to make our display larger, but it seems like now that we should have. It may have been better to go with a shorter and longer display rather than the shape we did. We could have used more space and therefore made our smallest fonts larger. One problem with our evaluation was that it occurred in the evening (thanks to an IRB mishap). Of course, our printouts for the display weren't illuminated so we had to rely on the interior light of the car. None of our in-car participants complained about the lighting itself, but it could be the reason why the profile labels were difficult to read. It's important to note that none of the computer users had trouble changing profiles or reading the labels.

Of course, we would like to have been able to automate the changing of the in-car display rather than having to do it manually. This was sometimes distracting to the user and caused him or her to miss the appearance of Buzz in the attention task slide show.

Meter
The Yea-Nay meter was not as obvious to the users as we though it might be. We discussed several different ways to show users how they had voted in the past on a certain song and to give them an idea of the comparative frequency of play of each particular song. We thought of displaying a hard number, percentage, a gauge, or some other kind of graphic. We knew it had to be an image that could rapidly convey to the user where that song ranked in their playlist.

In our in-car testing, we found that the meter did not get across to the users what we hoped it might. We believed the main reason was the lack of immediate adjustment in the meter after the user voted. This was due to the unavoidable lag that was caused by the investigators manually changing out the display. As expected, users of the software mockup had less trouble with the meter because it changed in a way the user could see it move up or down depending on the vote.

However, to facilitate a more noticeable change, we think we could have added tick marks or maybe a flashing arrow or some other attention grabber to indicate the arrow has moved. We also think a label may help users understand more of the purpose of this meter.

Also, we may want the vote on a song for the very first time to carry more weight than subsequent votes. Otherwise, it may be very difficult for new songs to get a prominent position on the playlist.

Evaluation Techniques/Questions
Although we gained a lot of valuable information from our user evaluation, there are things we would have done differently now that we reflect on it. For example, we realized in the course of our first evaluation that one of our scenario commands in our script was improperly worded. The command "You like this song and want to listen to it right now, but you don't want it to play more or less often in the future," was followed by the question "What would you do?" The correct response is to do nothing and let the song play out. However, we believed the wording was awkward and lead the user to want to press something. We would have dropped the first part so the command read "You want to hear this song right now, but you do not want to hear it more or less often in the future.

Looking back, there is additional information we would like to have learned from the user. We could have accomplished this by verbally asking more questions after our testing or asking the user to fill out a questionnaire before the testing. We also could have run one or two pilot subjects in order to work out kinks in our testing plan and generate more ideas for questions.

Some questions we would like to have asked:

- Can you tell from the screen which profile you are currently in?
- Do you think six profiles would appropriately accommodate your moods and lifestyles?
- Do you have a car radio that displays the title and artist?
- In the car, do you mostly listen to cds/tapes/mp3 or do you listen to the radio?
- If you listen to the radio regularly in the car do you find yourself switching stations every 5, 10, 15, 20, or do you mostly listen to one station?
- Would coding the Yea button green and the Nay button red have reduced the possibility of accidentally pushing the wrong button?
- Did the positioning of Yea on the right and Nay on the left make sense to you?

Also, we could have asked some additional questions in the software mockup. For example, since none of the computer users negatively commented on the meter, then we can only assume it is because they could see the changes in it. Maybe we should have asked this question explicitly. Also, with the in-car participants, we could have asked if they felt like there would be too much involved with the system to user while driving.

Random thoughts and suggestions from users
It is interesting to note that several of our users came up with ideas that we had never considered.

- The fact that the displays and controls were in the place where the horn and airbags are typically found was pointed out by one of our users.
- It was suggested that the display should be on the dashboard so that the driver doesn't have to look down at the steering wheel frequently.
- Should be able to assign multiple votes at once so that a new song could be moved more prominently to the Yea or Nay side.

Yea-Nay Radio Usability Testing Script

Description of the system to the users:

Recent years have shown a change in the interests of people and the way they wish to listen to music. The growing popularity of satellite radio and MP3 players shows a new trend in the way users wish to enjoy their musical listening experience. Users are attempting to take a more active role in the choice of music that they listen to.

The purpose of the Yea-Nay Personalized is to offer the users a higher level of control over their music choices. Through a process of voting upon songs, users create a database of their musical preferences. The system will then play songs for the user from both their personal selections as well as new selections that users with similar profiles preferred. Therefore, users can hear the songs they like, while still being introduced to new songs they may also enjoy. This grants the user a more personalized music experience.

A user begins to interact with our system by tuning the car radio to the Yea-Nay channel. At this time, the steering wheel display lights up with the words “Welcome to Yea-Nay Radio. Please select a profile.” The profile names are listed alongside the corresponding buttons that will activate each one. The user selects the profile he wants at which time a play-list generated by the Yea-Nay engine for that profile begins to play. The user votes on the songs he likes or dislikes by pressing the Yea or Nay buttons respectively. The more Yea votes that a user gives a song, the more often he/she will hear it. Likewise, the more Nay votes given to a song, the less often the listener will hear it. If the user does not want to log a vote for a song, he can listen to the song play out by not pressing any buttons. Or, the user can skip the song by pressing skip to move to the next song. Pressing skip will also not log a vote on the song. The user can log as many or as few votes as he chooses. He may change to any of the other five profiles at any time by pressing the desired profile button.

GROUP 1

Today we are going to ask you to test out our Yea-Nay radio system by listening to a number of songs and voting on them according to the preferences we give you. As a song plays, we will tell you whether or not you like the song and whether or not you want to hear it more or less often. After we tell you your “preferences,” we ask that you vote or abstain accordingly. We also ask that you “think aloud” as you interact with the system. After you are done testing the Yea-Nay interface, we are going to have you perform an attention task.

GROUP 2

Today we are going to ask you to test out our Yea-Nay radio system by using it in the car while “driving.” In the interest of safety, however, we can’t actually have you drive the car while testing the interface, so we’re going to have you perform an attention task on a laptop while sitting in a parked car. The attention task will involve a slide show showing a number of pictures of the Georgia Tech Campus, and your task will be to identify when Buzz is present in the pictures. Please keep in mind that the attention task is your primary task. That is the task we want you to concentrate on the most, trying to be as correct as possible. The secondary task will be to listen to songs that we play for you and vote on those songs using the Yea-Nay interface. As a song plays, we will tell you whether or not you like the song and whether or not you want to hear it more or less often. After we tell you your “preferences,” we ask that you vote or abstain accordingly.

Process:
Experimenter:
You've just tuned to the Yea-Nay radio station. This is the welcome screen. Choose the Road Trip profile

User: selects the Road Trip profile with the mouse.

System: starts playing "American Woman"

Experimenter: "You like this song and would like to hear it more often. What would you do?"

User: points to the interface and then explains what he would do and what he thinks would happen. Wait for the user to explain what he would do and then do it.

Experimenter: "Normally, voting Yea on a song would result in the song playing all the way through, but for the sake of time, we're going to move on to the next song."

System: starts playing "We belong to the Night"

Experimenter: "You do not like this song. You don't want to hear this song now, and you want it to play less often in the future. What would you do?"

User: points to the interface and then explains what he would do and what he thinks would happen. Wait for the user to explain what he would do and then do it.

System: starts playing "Stairway to Heaven"

Experimenter: "You like this song and want to listen to it right now, but you don't want it to play more or less often in the future. What would you do?"

User: points to the interface and then explains what he would do and what he thinks would happen. Wait for the user to explain what he would do and then do it.

Experimenter: "Again, for the sake of time, we're going to move on to the next song."

System: starts playing "Living on a Prayer."

Experimenter: "You like this song, and you don't want it to play more or less often in the future. However, you don't want to hear this song right now. What would you do?"

User: points to the interface and then explains what he would do and what he thinks would happen. Wait for the user to explain what he would do and then do it.

Experimenter: "Based on what you see here, can you tell how you've voted on this song in the past?" (Wait for participant to answer.)

Experimenter: "You like this song and would like to hear it more often. What would you do?"

User: points to the interface and then explains what he would do and what he thinks would happen. Wait for the user to explain what he would do and then do it.

Experimenter: "Now you want to change to the Susie profile. What would you do?"

Experimenter: "You think this song has been playing a little too often, so you want to hear it less often in the future. What would you do?

User: points to the interface and then explains what he would do and what he thinks would happen. Wait for the user to explain what he would do and then do it.

Post Experiment Interview

1. What did you think was the easiest thing to do on the interface?

2. What did you like best about the interface?

3. What did you think was the hardest thing to do on the interface? How would you make it easier?

4. What did you like least about the interface?

5. What do you think “happens” to the system when you vote “Yea” on a song?

6. How would you describe how the system works to someone who is thinking of purchasing it?

7. Do you have any suggestions for improvements?