Will we let computers determine what music we listen? Exploring user acceptance of music recommenders

Mark de la Court
University of Twente
PO Box 217, 7500 AE Enschede
the Netherlands

m.a.delacourt@utwente.utwente.nl

ABSTRACT

Music recommendation systems are becoming increasingly popular, but little is known as to how users come to accept and use such systems. In this research, a framework to explain user acceptance of music recommender systems is proposed, by means of a theoretical and empirical study. The results show that the accuracy of a recommender system alone cannot explain acceptance. In particular, it was found that we should also acknowledge the importance of Compatibility, Trust, Perceived Usefulness and Perceived Ease of Use in order to innovate music recommendation systems.

Keywords

Technology Acceptance Model, Music Streaming, Recommender systems, Recommender Acceptance.

1. INTRODUCTION

For most people, music is an important aspect of everyday life [30]. But the discovery of music can be a challenge, due to enormous collection of music available to explore. While we may be able to filter for a certain genre of music, this often still leaves hundreds of thousands of songs to examine and we have to resort to popularity metrics or curated playlists to sort through this music. Even our own collections can grow up to thousands of songs, also making it a challenge to navigate the music we own. Altogether, finding the right music to listen to can be tedious, and chances are that we do not actually find the music best matching our taste [25]. Music recommender systems are a solution to this problem, and recommendation systems are becoming increasingly important to discover music [6]. Through personalized recommendations, users can more easily discover music that matches their preferences. The industry is eager to offer such personalized recommendations, as it is an opportunity for them to attract and retain users [32]. An understanding of what influences user acceptance of music recommender systems is crucial to achieve the maximum benefits of recommender systems. However, to date little is known on exactly how users come to accept (or reject) such a system.

A substantial amount of research has been conducted on the development and improvement of recommendation algorithms [2, 32]. However, as recommendation technologies improve, we do not only want to focus on the accuracy of an algorithm, but also any other factors that may play a role in the use of music recommendation technology [1, 20]. Yet, predictors of recommendation technology acceptance other than accuracy have seldom been addressed. Research has been conducted on the user acceptance of recommender systems in general, a popular example being the research conducted by Pu et al [26]. However, rarely has user acceptance been explored specifically for the domain of music recommendations.

29th Twente Student Conference on IT, July. 6th, 2018, Enschede, The Netherlands.

To our knowledge, only one previous study has been conducted on user acceptance of music recommendation systems [16]. But this research was limited to an application of the Technology Acceptance Model, or TAM [9], and it did not extend or revise this model. This is a problem, because TAM as-is, is not very detailed and thus cannot always provide useful insights on why a technology is accepted or not. Consequently, the research only points out that *'Perceived Usefulness'* [9] is the most significant predictor, without actually defining when recommendations are perceived as useful. In addition, this research (published in 2009) is heavily outdated as music services have caused a shift from owning music to streaming music *[22],* which has major impacts on the way recommendations are made and used.

Therefore, there remains a gap in knowledge on the basis of which users accept music recommendation technology. Insights in this area will help researchers and developers to research and build better music recommendation systems that are truly innovative and fulfill users' needs. Thus, there is a need for additional research into the modeling of user acceptance of music recommender systems. This yields the following research question: *How do users decide whether or not to use a music recommendation technology?*

This paper is structured as follows: First, an acceptance model is selected. Second, this acceptance model is adapted to the domain of music recommendation technology. Then, the model is empirically validated with a user research and finally, the results will be presented and discussed.

2. THEORY

2.1 Background

Recommender systems play an increasingly important role in how we discover content every day. We rely on such systems to explore the overwhelming amount of information that is available to us. This is also true for music discovery. Music streaming services are no longer just listening platforms, but also the leading method for users to discover music [7, 15].

One of the ways through which streaming services can help users find music is through personal recommendations. However, this poses a challenge, since user involvement is typically low in this domain [16]. Interactions are limited, explicit feedback is seldom provided, and decisions, like the decision to play a song, are low-cost. The latter is in clear contrast with, for example, online webstores where users leave ratings and make weighted decisions on whereon to spend their money, as wrong decisions here involve higher costs. This leaves streaming providers to learn about their users based on only a limited set of information, like for example the songs that are skipped, and the playlists that are created.

Additionally, streaming services are pleasure or 'hedonic' oriented. This means that they are designed to provide self-fulfilling value, rather than an instrumental value and that they are used for fun rather than productivity [14]. And finally, music streaming services are often privately consumed. Therefore, outside pressures that normally exert great influence on use or purchase behavior may be less relevant for music streaming services.

This research will use a user acceptance model to predict the user acceptance of music recommender systems. A user acceptance model aims to predict the behavioral intention to use a system as well as the actual use behavior, based on several constructs. An example of a construct would be Perceived Usefulness; the more useful the application is perceived to be, the larger is the intention to use the system. The constructs vary per acceptance model. It is thus a challenge to find a model with constructs relevant for music recommender systems. A relevant model would need to take into account the hedonic, low-involvement and private consumption nature of music recommender systems.

2.2 Choosing an acceptance model

To find suitable acceptance models, a limited literature review is conducted. The terms "user acceptance information technology" and "user acceptance hedonic technology" were used in Scopus to search for both generic and hedonic technology acceptance models. Results were sorted based on relevancy and the first 100 results for each query were inspected. From the results, all articles that did not propose a generic or a hedonic acceptance model were rejected. This also implies that all articles presenting an acceptance model for a specific domain (i.e. ERP systems, health, social networks) were rejected. Finally, all articles with less than 50 citations were excluded, as a generally well-known, well-accepted and well-used model is preferred. This literature review yielded eight models [9, 10, 14, 18, 35–37, 41] that met the stated criteria.

To select the most suitable model from the eight candidate models, the models are compared based on the following criteria:

- The expected *relevancy* of the constructs: This is intuitively meaningful. For example, 'Voluntariness of use' (from the UTAUT model discussed later) would not be very relevant for the acceptance of hedonic information systems.
- The *number* of constructs: This is meaningful because more relevant constructs provide a more detailed insight into the user perceptions of a system.

The TAM [9] and the Van der Heijden [14] model both offer a very limited set of constructs, with respectively two and three constructs. While suitable for application to the music recommendation domain, these constructs provide only a very high-level overview on what contributes to acceptance. The models, for example, do not provide more detail than pointing out that "Perceived Ease of Use" or "Perceived enjoyment" are important. For more detailed and informative insights into what exactly drives acceptance, these models would need to be extended with more constructs or external variables.

The Unified Theory of Acceptance and Use of Technology, or the UTAUT model [36], and the UTAUT 2 [37] model are not ideally suited for a hedonic purpose, as they are designed to measure acceptance of utilitarian information systems. Utilitarian systems are, in contrary to hedonic systems, primarily focused on increasing productivity and use in a work environment. This is reflected in, for example, the construct 'Voluntariness of Use', which is not particularly relevant for hedonic purposes. The UTAUT model has been used for the hedonic domain. For example, Wang et al. [39] adapted the UTAUT for product recommender systems and found no difference in the model for hedonic compared to utilitarian products. However, the use of UTAUT models for hedonic technology is not common and acceptance models designed specifically for a hedonic domain are more likely to have a higher predictive power.

The model by Wixom [41] has the most constructs, with 17 constructs total. While a large number of constructs is generally preferable to get more detailed insights, 17 constructs may be a disproportionate number considering the limited timeframe and the limited resources available for this research. Additionally, similar to the UTAUT model, its focus is predominantly on utilitarian information systems.

The Hedonic Motivation System Adoption Model (HMSAM) [18] and the model by Turel [35] are both focused on the acceptance of hedonic information systems and they both have around 8 constructs. Both models have a construct of immersion (or 'Escapism' in Turel's model), which may not be very relevant for music recommendation and consumption. This is because research showed that people mostly listen to music as a secondary activity while performing another task [17, 23]. The model by Turel also has a focus on monetization as becomes apparent from the construct "Value for money". Since music recommendation technology is often a complimentary service provided with a music streaming service this is not a very relevant construct as well.

The Automation Acceptance Model (AAM) [10], extends the TAM with two constructs 'Compatibility' and 'Trust'. Both of these constructs are very suitable for (music) recommender systems, as will be explained in more detail further on in this paper. Since the TAM has already been applied with success to music recommender systems [16], a model extending the TAM would be a very plausible option and the two extensions to the TAM proposed by the AAM provide the model with a level of explanatory power that the TAM by itself cannot provide.

Concluding, the AAM appears to have the most suitable number of constructs with a high relevancy. The alternative models, while suitable, seem to focus more on utilitarian information systems, lack in explanatory power or are to complex compared to the AAM.

2.3 The Automation Acceptance Model

The Automation Acceptance Model (AAM) as proposed by Ghazizadeh is pictured in Figure 1.

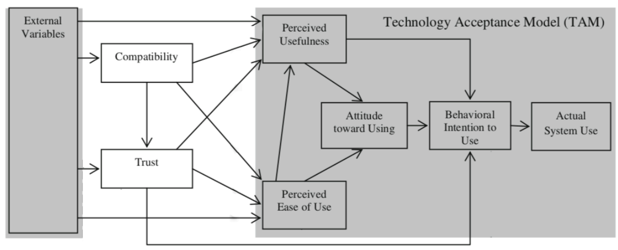

Figure 1: The Automation Acceptance Model

The AAM integrates the constructs of Compatibility and Trust with the constructs of the TAM. Thus, the core constructs of the AAM are:

- Compatibility: “The technology’s consistency with users’ values, past experience, and needs” [29].
- Trust: “The willingness to rely on a specific other, based on confidence that one’s trust will lead to positive outcomes” [8]. Trust is typically defined in relation to the provider of technology and the channel of communication, more than the technology itself [10].
- Perceived Usefulness (PU): “The degree to which a person believes that using a particular system would enhance his or her job performance” [9].
- Perceived Ease of Use (PEU): “The degree to which a person believes that using a particular system would be free from effort” [9]

All these constructs predict the Attitude towards Using, which in turn predicts Behavioural Intention to Use, which finally predicts Actual System Use. Attitude towards Using will however not be considered in this research, as it was omitted by Davis et Al. in their final model [36]. Instead, Perceived Ease of Use and Perceived Usefulness are considered as direct predictors of Behavioural Intention to Use.

2.4 Identifying additional dimensions

The constructs of the model should not be seen as fixed entities, but rather as categories that can consist of several dimensions. The original TAM proposal also structured question items for a single construct into various dimensions. For example, Davis used, among others, the dimensions ‘Flexible’ and ‘Easy to Learn’ for the construct Ease of Use [9]. Other researchers continue to identify domain-specific dimensions in applications of the TAM [1, 16, 26] to allow for better explanations of user perceptions, beliefs, and attitudes.

In this section, domain-specific dimensions for each of the constructs of the model will be proposed. These dimensions will help to provide better explanations of the constructs of the model and they will also help to capture domain-specific attributes of the construct in the measurement.

2.4.1 Dimensions of Perceived Usefulness

It is commonly known in recommendation research that accuracy, novelty, and diversity are important properties of recommendations [4, 5, 43]. Research by Pu et al. showed that accuracy, novelty, and diversity are significant predictors of the Perceived Usefulness of product recommender systems [26]. These predictors are intuitively relevant in a (music) discovery process. What we find should match our preference (accuracy), it should not only be music we already know and listen to (novelty) and we expect the songs to be different from each other (diversity). Thus, it is hypothesized that accuracy, novelty, and diversity are important dimensions of Perceived Usefulness in the music recommendation domain.

2.4.2 Dimensions of Ease of Use

The Perceived Usefulness as well as Use Intentions of a recommender system depend directly on the perceived Ease of Use of the system. After all, if finding new content manually would be easier than finding new content using the recommender system, it would not be a very useful. Armentano uses four dimensions for the construct Ease of Use in his research into recommender system acceptance [1]. One of these dimensions, the ability to evaluate, is unique for recommender systems, while the other dimensions match the dimensions posed by Davis. The ability to evaluate shows how well we are able to assess the recommendations provided to us. For example, Spotify provides the recommendations as a regular finite playlist while Google Play Music provides the songs as an endless radio station; the first provides a way to see all the songs and artists at glance, while the second only allows evaluation by skipping through songs one by one. Thus, the dimension ‘ability to evaluate’ can help to determine which form factor for presenting recommendations is best.

2.4.3 Dimensions of Trust

Music recommendations, like many recommendations, may seem to come from a black box to users. Therefore, we may be sceptical about a recommender system before we trust it to provide us recommendations. Users will rely on a system more and use a system more if they trust it [24] and Wang showed in an extension and modification of the UTAUT model, that trust is a significant predictor of the intention to use a recommender system [40]. The Automation Acceptance Model [10] proposes that Trust is a relevant predictor of both Perceived Ease of Use and Perceived Usefulness of information systems. Trust is thus an important value in recommender systems. We may trust a recommender system more if we know how it works (transparency) and if we can exert influence over it (control). Research has confirms this, and it is shown that trust is influenced by control over recommendations [42] and transparency about why recommendations are made [31, 34]. Thus, for Trust the dimensions of transparency and control are proposed.

2.4.4 Dimensions of Compatibility

Compatibility describes how well a system supports a given task. The music that we want to play depends on our motivation, which is in turn determined by our context [38]. For example, if we are working out, we want music supporting this activity and we are consequently likely to be seeking for high tempo music. The perceived compatibility between what we want and what the system provides thus influences Perceived Usefulness. That motivation is important for overall acceptance follows from the inclusion of motivation in popular and well supported high-level acceptance models, for example, in the UTAUT 2 model (hedonic motivation) [37]. Lowry also stresses the importance of including motivations in the limitations of his HMSAM model [18]. Contexts are likewise frequently considered in research into music consumption [23, 38]. Therefore, both the dimensions of motivation and context are hypothesized to be relevant for the construct of Compatibility.

Motivation, however, is a rather complex dimension to capture in an acceptance model. Where all other dimensions previously identified are continuous, the dimension of motivation is more categorical. I.e. some motivations may be better supported than others. To be able to measure motivation as a continuous construct in the model, the mean of different kinds of motivations will be taken. This will result in a continuous measure of motivation, reflecting the degree to which various motivations are supported. The more and the better individual motivations are supported, the higher the mean of all motivations will be and vice versa.

To find the motivations to include, literature on the psychological determinants for listening to music has been consulted. Lonsdale and North identified six factors, or ‘reasons’, for listening to music [17]. Compared to other categorizations of reasons for listening to music [3, 23, 30, 33, 38] the research of Lonsdale and North has the highest amount of support (150+ citations) given the high amount of variance in music preference (64%) it accounts for. The six factors, as identified by Lonsdale and North, are:

- Positive mood management (i.e. to achieve and optimize a positive mood)
- Diversion (i.e. to pass time or distract)

- Negative mood management (i.e. to alleviate negative mood)
- Interpersonal relationships (i.e. to have something to talk about)
- Personal identity (i.e. to create an image of oneself)
- Surveillance (i.e. to learn how other people think).

The first three of those factors are the primary reasons for listening to music, the last three factors are the secondary, or social, reasons for listening to music. For the dimension of motivation, as to be included in this research, only the primary reasons will be considered.

3. METHODOLOGY

To verify the Automation Acceptance Model, as well as the proposed dimensions for the music recommendation domain, participants will be questioned about their beliefs about, and use of, music recommendation systems.

For this user study, the focus will be on the most common form of music recommender systems, which are recommended playlists. Such recommended playlists are part of most popular music streaming services today, like Spotify, Apple Music, and Google Play Music. Most services offer various kinds of recommended playlists, for example, Spotify offers both Discover Weekly playlists to discover previously unknown music, as well as Daily Mixes, to play both known and unknown music from the various music styles users listen to.

This survey will be conducted with Dutch residents, where Spotify is with distance the largest music streaming service with 6,2 million users, Apple music comes second with 1,1 million monthly users and Deezer is third with 0.8 million [44]. The survey will be designed with this in mind, providing relevant examples for these three services when needed. Since the survey will address the perceptions of music recommender systems, both users as well as non-users of these systems will be invited to join the survey.

3.1 Measures

The constructs of the model will be measured by means of several statements per construct for the respondent to rate. Questions for Perceived Ease of Use and Perceived Usefulness are adapted from Davis [9], questions to measure Compatibility (as defined by Rogers [29]) are adapted from Moore [21] and questions for Trust are adapted from an earlier study of Ghazizadeh extending the TAM with Trust [11]. For conciseness, no more than three questions from literature are adapted for each construct. The questions to include were selected based on how well they could be adapted to the recommender domain and, if available, their rating (accuracy) in the literature.

To measure the dimensions, users will also be asked to rate a statement for each dimension. The question for transparency was adapted from [31]. The questions for each kind of motivation are adapted from [17]. The questions to access accuracy, novelty and diversity are adapted from [26]. The questions to access the other dimensions have been carefully created by the Author based on the literature on these dimensions discussed earlier. Since context is a very broad concept, only the elements 'where' and 'what' are considered in the survey, this because these elements provide insight into the activity context, which is one of the most effective contexts to consider [38].

Use, use intentions and satisfaction are also measured. Two questions are included to measure usage intention. One question is included to measure actual use by asking about the number of hours/minutes that the respondent listens to recommended playlists. And two questions have been included to measure satisfaction. One as a direct measure of satisfaction, by asking for satisfaction directly, the other as an indirect measure of satisfaction, by asking the number of songs the respondent saved from the recommended playlists to their own playlists. Finally, questions for the demographics are included inquiring about gender, age, education, streaming services used, recommended playlists used and duration of use of streaming services.

To reduce the chance of any misunderstandings among respondents, all the questions in the survey were translated into Dutch. At the start of the survey, a definition and small explanation of recommended playlists were provided in addition to some examples. This was done to ensure that all participants, even the participants unfamiliar with streaming services, understood what was meant with 'recommended playlist'.

3.2 Survey Instrument

An online survey instrument was used to distribute the survey and collect responses. Respondents were asked to indicate to which degree they agreed to the statements. A five-point Likert scale was used for this to ensure conciseness, as well as to enhance data quality [27].

The instrument was first tested with five participants to ensure that the questions were understandable, and that the information provided was sufficient to answer the questions. Upon this first test, several small improvements were made before the survey was distributed.

3.3 Participants

Over one hundred people were invited to respond to the survey, of which 35 people responded and fully completed the survey. The age of the respondents varied from 16 years old to 62 years old with a mean age of 26. About 30% of all respondents were female. 60% of all respondents had a university degree or were studying at a university.

Only four of the respondents never used a streaming service. The most common streaming service among respondents was Spotify, with a 95% share. Some also used Apple Music (13%), Google Play Music (10%) or Deezer (7%). About 80% of our respondents listened to a recommended playlist at least once a week. Spotify's Discover Weekly, Daily Mixes, and Release radar being the most popular playlists. Recommended playlists from other music services were barely used, with only less than 8% of our respondents using them (1 person for Apple Music and 1 for Google Play music).

3.4 Data analysis

The data will be analyzed using a structural equation model. This is a popular method to assess the constructs in an acceptance model and the relations between them. This method has been frequently applied in fundamental user acceptance research [14, 18, 35, 36]. A structural equation model (SEM) consists of a measurement model and a structural model. The *measurement model* contains the constructs, or 'latent variables', as well as the question used to measure these constructs, the 'observed variables'. The relations between the constructs are defined in the *structural model*, these relations are validated through a statistical method. A structural equation model can be used to find the strength the paths between two constructs (or between constructs and their indicators) and significance of these paths.

There are many statistical methods available to compute structural equation models, the most popular being CB-SEM (Covariance Based) and PLS-SEM (Partial Least Squares). Primarily considering the small sample size, but also considering the exploratory nature of this research and the uncertainty about the distribution of the data, PLS-SEM is the most suitable statistical method to compute the SEM for this dataset [13, 28].

The SEM will be computed for both the original AAM and the adapted AAM, which includes the proposed new dimensions. Both models will be computed with the SmartPLS 3.0 software suite. Other computations, like Cronbach's Alpha, the one-way ANOVA and correlations, will be computed using IBM SPSS 25.

4. FINDINGS

4.1 Cronbach's Alpha

Each construct has two or more indicators, which are the questions from the survey. To ensure that the indicators are reliable and that all indicators for one construct measure the same, the Cronbach's Alpha has been computed. A value above .7 is considered acceptable [12]. As can be seen in the table below, all constructs meet this criterium.

Construct	**Alpha**
Perceived Usefulness	.848
Perceived Ease of Use	.737
Perceived Compatibility	.726
Perceived Trust	.899
Use Intention	.948

4.2 PLS Analysis of the AAM

The table below shows the path coefficients and P-values (two-tailed) for the AAM computed with a PLS Structural Equation Model based on the survey data.

Path from	Path to	Path Coefficient	P-value
Compatibility	Trust	0.367*	0.024
	PU	0.305*	0.022
	PEOU	0.227	0.204
Trust	PU	0.356**	0.006
	PEOU	-0.011	0.954
	Use Intention	0.036	0.830
PEOU	PU	0.465**	0.002
	Use Intention	-0.062	0.732
PU	Use Intention	0.668**	0.000
Use Intention	Use	0.614**	0.000

** P < 0.05 ** P < 0.01 (Bootstrapping: 5000 subsamples, PLS: 300 iterations)*

The path model additionally shows that the indicators all have communalities above .80 with p-values of 0.00. The indicator communalities are essentially regression coefficients and they reflect the strength of the relationship between the indicator and the construct. Since an average of .7 and a minimum of .6 has been found to be the threshold for communalities for small sample sizes [19], the communalities of .8 and the p-values confirm the reliability of the indicators established earlier with the Cronbach Alpha values.

4.3 Correlations of Dimensions

Correlations are used to establish whether there is a relationship between the additional dimensions identified for each construct and the indicators of the construct itself, and whether this relationship is statistically significant.

Perceived Usefulness	PU1	PU2	PU3
Diversity	-.034	-.044	-.091
Accuracy	.498**	.602**	.578**
Novelty	.433**	.446**	.355**

Compatibility	COMP1	COMP2
Context	.446**	.277
Motivation	.473**	.258

Trust	TRUST1	TRUST2
Control	-.069	-.004
Transparency	-.046	0.045

Perceived Ease of Use	PEOU1	PEOU2
Ability to Evaluate	.012	.356*

** P < 0.05 ** P < 0.01 Both P-levels are two-tailed.*

4.4 PLS Analysis of the adapted AAM

To confirm the results of the correlation analysis, all potentially significant dimensions have been added to the structural equation model and the model was updated *(Bootstrapping: 5000 subsamples, PLS: 300 iterations).* All non-significant factors (P > 0.05 and communality < .6) were then removed. This left the dimensions of context, motivation, accuracy, and novelty. The threshold of an average communality of .7 and a minimum of .6 for all indicators is still met in the updated model. Also, the Cronbach's Alpha have been re-computed to include the new dimensions, and all alpha values are still above .7.

The table with the path coefficients and P-values (two-tailed) for the *adapted* AAM with the new dimensions, is shown below.

Path from	Path to	Path Coefficient	P-value
Compatibility	Trust	0.350*	0.025
	PU	0.487**	0.001
	PEOU	0.271	0.146
Trust	PU	0.242*	0.040
	PEOU	-0.022	0.910
	Intention to use	0.047	0.784
PEOU	PU	0.401**	0.008
	Use Intention	-0.092	0.518
PU	Use Intention	0.738**	0.000
Use Intention	Use	0.614**	0.000

** P < 0.05 ** P < 0.01 (Bootstrapping: 5000 subsamples, PLS: 300 iterations)*

The path model for the AAM and the path model for the adapted AAM, with the new dimensions, are very similar. Small differences are that the model with the new dimensions appears to have a stronger and more significant connection between

Compatibility and PU, but it also has a weaker connection between Trust and PU. Both models predict the intention to use with path coefficients around .7, the model with the new dimensions is slightly more accurate with a path coefficient of .738 compared .668, but these differences are minimal especially considering the sample size. Since the predictive power of the model with or without the new dimensions could be considered the same, the inclusion of the additional dimensions is preferred since it provides the model with more explanatory power.

4.5 Application

The primary purpose of the survey is to collect data to verify the acceptance model. However, when applied to the acceptance model, the data gathered with the survey also provides an early insight into how recommender systems are currently perceived, for example, what is liked and disliked about these systems.

In the demographics, it was already discussed that over 80 percent of the respondents use recommended playlists. This may be explained by the high level of satisfaction with these playlists; 82% people using recommended playlists were overall satisfied with the system and 88% of the people using recommended playlists occasionally save a song from a recommended playlist to their own playlists. People indicate they usually listen to recommended playlists for about half an hour per week.

Rated on a scale of one to five, recommended playlists are perceived to be most suitable to get into a positive mood or maintain a positive mood (μ=3.91, σ=.66) but also to distract oneself (μ=3.54, σ=.89), recommended playlists are perceived as slightly less compatible to elevate negative moods (μ=3.32, σ=1.03). Regarding context compatibility, it is perceived that recommended playlists can be used across locations (μ=3.63, σ=1.190) and is perceived as somewhat less compatible across activities (μ=3.11, σ=1.183). This may suggest that additional attention is needed in making and optimizing recommended playlists to elevate negative moods, also further research could be conducted to investigate the types of activities that are supported by recommended playlists and the ones that are not.

The validation showed that only novelty and accuracy were dimensions of perceived usefulness. The means of the results also showed that recommendations are respectively most perceived as novel (μ=4.20, σ=.677), then accurate (μ=3.57, σ=.739) and finally as diverse (μ=3.26, σ=.886). The fact that diversity is rated lower, and is not significant for perceived usefulness, may suggest that people still like to use recommended playlists primarily to discover music, and not for other purposes. Further research could look into why recommender systems would not be used as the primary way to listen to music, but instead only for discovery of new music. It could be hypothesized that the limited diversity is one of the reasons.

Perceptions of people that do not use recommended playlists are different from those that do. In the table below, the mean ratings for the constructs (and their dimensions) for both users and non-users of recommender systems have been given. Significance was computed using a One-Way ANOVA test.

	Non-users (μ)	Users (μ)	Sig.
Perceived Usefulness	3.67	4.12	.139
Perceived Ease of Use	3.75	4.41	.007**
Accuracy	3.13	3.70	.050*
Novelty	3.63	4.37	.005**
Diversity	3.22	3.38	.675
Trust	3.31	3.31	.995
Compatibility	3.25	3.50	.436

** Significant at the 0.05 level ** Significant at the 0.01 level*

Non-users rate Perceived Usefulness lower than people that do use recommended playlists. This may be explained by the fact Perceived Ease of Use is rated significantly lower. Non-users also rated Accuracy and Novelty significantly lower than users did. Diversity, as expected is rated similarly among both groups and remarkably, both groups also rate Trust similarly. Compatibility is rated only slightly lower though. Thus, in order to increase acceptance of recommender systems for non-users, it could be important to improve perceptions of Usefulness and Ease of Use. Perceived Usefulness could, in turn, be improved by promoting perceptions of high accuracy and novelty.

5. DISCUSSION

The results show that Compatibility, Trust, Perceived Ease of Use and Perceived Usefulness play a significant role in the acceptance of recommender systems. The confirmed relationships of the adapted Automation Acceptance model are depicted in Figure 2.

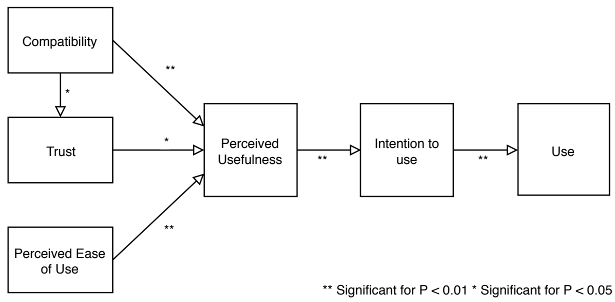

Figure 2: Adapted Automation Acceptance Model

Some of the relationships hypothesized by the AAM could not be confirmed. While not all unconfirmed relationships will be discussed, it is perhaps interesting to address why the relationship between Perceived Ease of Use and Intention to Use could not be confirmed, as one might have expected based on the TAM. Davis stated about this that Perceived Ease of Use should be seen as a predictor of Usefulness rather than a direct predictor of Intention to Use: “From causal perspective, [..] ease of use may be an antecedent to usefulness, rather than a parallel, direct determinant of usage.” [9].

Additional dimensions were also confirmed for the constructs of the Automation Acceptance Model: accuracy, novelty, context, and motivation. Some hypothesized dimensions could not be confirmed; the dimensions of control, transparency, diversity and the ability to evaluate were not relevant or significant for their respective constructs. Not all of the dimensions will be discussed individually, but it is perhaps interesting to hypothesize about why novelty and accuracy were found to be relevant for Perceived Usefulness, but diversity was not; If we are out to find new music to listen, we may be more interested in the fact that it is new and that it matches our taste. Diversity may be quite important in playlists when actually playing music, but it is less relevant for the usefulness of recommendations, when we are only seeking to find new music that we will like.

6. CONCLUSION

Whether we buy a recommended product on Amazon or read a news article on our Facebook timeline, we are relying on suggestions from intelligent recommender systems. The same holds for the music we play, we increasingly rely on music recommender systems to find new music to listen to. However, while on the rise, music recommender systems are still far away from being our primary way to listen to music. To change this, and to make recommender systems more suitable for everyday use we need to gain an understanding as to how users come to accept a music recommender system.

To explain user acceptance of music recommender systems, an acceptance model was proposed and empirically confirmed. The literature research shows that the Automation Acceptance Model is the most suitable acceptance model to predict user acceptance, as it provides a sufficient level of detail and fits the hedonic and low-involvement domain of music recommender systems. The empirical study validated this model and confirms that perceptions of Compatibility, Trust, Usefulness, and Ease of Use in fact play an important role in user acceptance of music recommenders. Compatibility, Trust and Perceived Ease of Use all contribute to our Perceptions of Usefulness, which in turn predicts the intention to use and actual use of recommender systems. More specifically, Compatibility and Perceived Ease of Use directly predict Perceptions of Usefulness, while Trust acts as a mediator between Compatibility and Perceived Usability.

To adapt to model to the domain of music recommender systems, additional dimensions for the constructs of the model have been identified. The results show that motivation and context are important dimensions of Compatibility and that accuracy and novelty are important dimensions of Perceived Usefulness. These additional constructs provide the acceptance model with a greater level of explanatory power.

The proposed framework can help developers and researchers of music recommender systems to understand user acceptance. The framework stresses the importance of focusing on a diverse set of user needs, rather than only focusing on accuracy. Through this, the framework aims to enhance the discovery experience for users of streaming services.

The data from the user study also provides an early insight into what the user perceptions and acceptance issues are. It is found that music recommenders are used for around 30 minutes a week and they are predominantly perceived as a method to discover new music. To make recommender platforms more useful, they need to evolve beyond a discover platforms centered around accuracy. Instead, recommender systems should focus on matching our needs throughout the day; taking into account our contexts, motivations and cognitive preferences for listening to music, like our preference for novelty. Additionally, the user data revealed that perceptions of Usefulness (accuracy and novelty) and Ease of Use are significantly lower for users than non-users. To attract new users, streaming services could thus focus on stressing to Ease of Use, novelty, and accuracy of their recommendation systems.

7. LIMITATIONS

Several limitations of this research have been identified which should be considered when using or interpreting the results:

(1) Due to the limited sample size, this research should be considered as preliminary and exploratory. The proposed framework will need additional research for further validation, in order to be relied upon in practice when developing music recommender systems.

(2) A translation bias may apply to the questions used. While the questions were carefully translated from English to Dutch, it is possible that some questions may have a slightly different meaning in Dutch and that this may have affected the measurements.

(3) The personal network of the author was used to find respondents. Hence the respondents may be younger, more affinitive with technology, and have a higher level of education than a truly heterogeneous population would have. As is valid for every research, a more heterogeneous and larger sample size would increase the external validity of this research.

(4) The user study only validates the framework for recommended playlists. The findings may thus not extend to other music recommender systems, such as personalized radio.

8. ACKNOWLEDGMENTS

The author would like to thank Dr. Ir. A.A.M. Spil for his supervision as well as the respondents of the survey for their participation in this research.

9. REFERENCES

[1] Armentano, M.G., Abalde, R., Schiaffino, S. and Amandi, A. 2014. User Acceptance of Recommender Systems: Influence of the Preference Elicitation Algorithm. *2014 9th International Workshop on Semantic and Social Media Adaptation and Personalization* (Nov. 2014), 72–76.

[2] Bobadilla, J., Ortega, F., Hernando, A. and Gutiérrez, A. 2013. Recommender systems survey. *Knowledge-Based Systems*. 46, (Jul. 2013), 109–132.

[3] Boer, D., Fischer, R., Tekman, H.G., Abubakar, A., Njenga, J. and Zenger, M. 2012. Young people's topography of musical functions: Personal, social and cultural experiences with music across genders and six societies. *International Journal of Psychology*. 47, 5 (Oct. 2012), 355–369.

[4] Castells, P., Hurley, N.J. and Vargas, S. 2015. Novelty and Diversity in Recommender Systems. *Recommender Systems Handbook*. F. Ricci, L. Rokach, and B. Shapira, eds. Springer US. 881–918.

[5] Castells, P., Vargas, S. and Wang, J. 2012. Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance. *In Proceedings of International Workshop on Diversity in Document Retrieval* (2012), 29–37.

[6] Celma, O. and Lamere, P. 2011. Music recommendation and discovery revisited. *Proceedings of the fifth ACM conference on Recommender systems* (2011), 7–8.

[7] Charles, L. 2016. An Exploration Into How The Rise of Curation Within Streaming Services Has Impacted How Music Fans in the UK Discover New Music. *Journal of Promotional Communications*. 4, 1 (2016), 115–141.

[8] Chopra, K. and Wallace, W.A. 2003. Trust in electronic environments. *Proceedings of the 36th Annual Hawaii International Conference on System Sciences* (2003), 10.

[9] Davis, F.D. 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*. 13, 3 (Sep. 1989), 319.

[10] Ghazizadeh, M., Lee, J.D. and Boyle, L.N. 2012. Extending the Technology Acceptance Model to assess automation. *Cognition, Technology & Work*. 14, 1 (Mar. 2012), 39–49.

[11] Ghazizadeh, M., Peng, Y., Lee, J.D. and Boyle, L.N. 2012. Augmenting the Technology Acceptance Model with Trust: Commercial Drivers' Attitudes towards Monitoring and Feedback. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 56, 1 (Sep. 2012), 2286–2290.

[12] Gliem, J.A. and Gliem, R.R. 2013. Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. *2003 Midwest Research to Practice Conference* (Oct. 2013), 82–88.

[13] Hair, J.F., Ringle, C.M. and Sarstedt, M. 2011. PLS-SEM: Indeed a Silver Bullet. *The Journal of Marketing Theory and Practice*. 19, 2 (Apr. 2011), 139–152.

[14] van der Heijden, H. 2004. User Acceptance of Hedonic Information Systems. *MIS Quarterly*. 28, 4 (2004), 695.

[15] Jacobson, K., Murali, V., Newett, E., Whitman, B. and Yon, R. 2016. Music Personalization at Spotify. *Proceedings of the 10th ACM Conference on Recommender Systems* (2016), 373–373.

[16] Jones, N. and Pu, P. 2009. User Acceptance Issues in Music Recommender Systems. *EPFL Technical Report*. (2009), 10.

[17] Lonsdale, A.J. and North, A.C. 2011. Why do we listen to music? A uses and gratifications analysis: Music uses and gratifications. *British Journal of Psychology*. 102, 1 (Feb. 2011), 108–134.

[18] Lowry, P., Gaskin, J., Twyman, N., Hammer, B. and Roberts, T. 2013. Taking "Fun and Games" Seriously: Proposing the Hedonic-Motivation System Adoption Model (HMSAM). *Journal of the Association for Information Systems*. 14, 11 (Nov. 2013), 617–671.

[19] MacCallum, R.C., Widaman, K.F., Zhang, S. and Hong, S. 1999. Sample size in factor analysis. *Psychological Methods*. 4, 1 (1999), 84–99.

[20] McNee, S.M., Riedl, J. and Konstan, J.A. 2006. Being accurate is not enough: how accuracy metrics have hurt recommender systems. *CHI '06 Extended Abstracts on Human Factors in Computing Systems* (2006), 1097–1101.

[21] Moore, G.C. and Benbasat, I. 1991. Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation. *Information Systems Research*. 2, 3 (Sep. 1991), 192–222.

[22] Morris, J.W. and Powers, D. 2015. Control, curation and musical experience in streaming music services. *Creative Industries Journal*. 8, 2 (Jul. 2015), 106–122.

[23] North, A.C. 2004. Uses of Music in Everyday Life. *Music Perception: An Interdisciplinary Journal*. 22, 1 (Sep. 2004), 41–77.

[24] Parasuraman, R., Sheridan, T.B. and Wickens, C.D. 2008. Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*. 2, 2 (Jun. 2008), 140–160.

[25] Park, H.-S., Yoo, J.-O. and Cho, S.-B. 2006. A Context-Aware Music Recommendation System Using Fuzzy Bayesian Networks with Utility Theory. *Fuzzy Systems and Knowledge Discovery*. L. Wang, L. Jiao, G. Shi, X. Li, and J. Liu, eds. Springer Berlin Heidelberg. 970–979.

[26] Pu, P., Chen, L. and Hu, R. 2011. A user-centric evaluation framework for recommender systems. *Proceedings of the fifth ACM conference on Recommender systems* (Oct. 2011), 157–164.

[27] Revilla, M.A., Saris, W.E. and Krosnick, J.A. 2014. Choosing the Number of Categories in Agree–Disagree Scales. *Sociological Methods & Research*. 43, 1 (Feb. 2014), 73–97.

[28] Richter, N.F., Cepeda, G., Roldán, J.L. and Ringle, C.M. 2016. European management research using partial least squares structural equation modeling (PLS-SEM). *European Management Journal*. 34, 6 (Dec. 2016), 589–597.

[29] Rogers, E.M. 1995. *Diffusion of innovations*. Free Press.

[30] Schäfer, T. and Sedlmeier, P. 2009. From the functions of music to music preference. *Psychology of Music*. 37, 3 (Jul. 2009), 279–300.

[31] Sinha, R. and Swearingen, K. 2002. The Role of Transparency in Recommender Systems. *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (2002), 830–831.

[32] Song, Y., Dixon, S. and Pearce, M. 2012. A Survey of Music Recommendation Systems and Future Perspectives. *9th International Symposium on Computer Music Modeling and Retrieval* (Jun. 2012), 17.

[33] Tarrant, M., North, A.C. and Hargreaves, D.J. 2000. English and American Adolescents' Reasons for Listening to Music. *Psychology of Music*. 28, 2 (Oct. 2000), 166–173.

[34] Tintarev, N. and Masthoff, J. 2007. A Survey of Explanations in Recommender Systems. *IEEE 23rd International Conference on Data Engineering Workshop* (Apr. 2007), 801–810.

[35] Turel, O., Serenko, A. and Bontis, N. 2010. User acceptance of hedonic digital artifacts: A theory of consumption values perspective. *Information & Management*. 47, 1 (Jan. 2010), 53–59.

[36] Venkatesh, V., Morris, M., Davis, G. and Davis, F. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*. 27, 3 (2003), 425.

[37] Venkatesh, V., Thong, J.Y.L. and Xu, X. 2012. Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*. 36, 1 (Mar. 2012), 157–178.

[38] Volokhin, S. and Agichtein, E. 2018. Understanding Music Listening Intents During Daily Activities with Implications for Contextual Music Recommendation. *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (2018), 313–316.

[39] Wang, Y.-Y., Luse, A., Townsend, A.M. and Mennecke, B.E. 2015. Understanding the moderating roles of types of recommender systems and products on customer behavioral intention to use recommender systems. *Information Systems and e-Business Management*. 13, 4 (Nov. 2015), 769–799.

[40] Wang, Y.-Y., Townsend, A., Luse, A. and Mennecke, B. 2012. The Determinants of Acceptance of Recommender Systems: Applying the UTAUT Model. *AMCIS 2012 Proceedings*. 3, (Jan. 2012), 2238–2246.

[41] Wixom, B.H. and Todd, P.A. 2005. A Theoretical Integration of User Satisfaction and Technology Acceptance. *Information Systems Research*. 16, 1 (Mar. 2005), 85–102.

[42] Xiao, B. and Benbasat, I. 2007. E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. *MIS Quarterly*. 31, 1 (2007), 137.

[43] Zhou, T., Kuscsik, Z., Liu, J.-G., Medo, M., Wakeling, J.R. and Zhang, Y.-C. 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*. 107, 10 (Mar. 2010), 4511–4515.

[44] 2018. Spotify groeit weer in Nederland, betalende klanten >50%. *Telecompaper*.