Building a Personalized Audio Equalizer Interface with Transfer Learning and Active Learning

Bryan Pardo
EECS
Northwestern University
2133 Sheridan Rd, Evanston, IL, USA
+1 847 491 7184
pardo@northwestern.edu

David Little
Communication Sciences & Disorders
Northwestern University
2240 Campus Dr, Evanston, IL USA
+1 847 714 4163
d-little@u.northwestern.edu

Darren Gergle
Communication Studies
Northwestern University
2240 Campus Dr, Evanston, IL USA.
+1 847 467 1221
dgergle@northwestern.edu

ABSTRACT
Potential users of audio production software, such as audio equalizers, may be discouraged by the complexity of the interface and a lack of clear affordances in typical interfaces. In this work, we create a personalized on-screen slider that lets the user manipulate the audio with an equalizer in terms of a descriptive term (e.g. "warm"). The system learns mappings by presenting a sequence of sounds to the user and correlating the gain in each frequency band with the user's preference rating. This method is extended and improved on by incorporating knowledge from a database of prior concepts taught to the system by prior users. This is done with a combination of active learning and simple transfer learning. Results on a study of 35 participants show personalized audio manipulation tool can be built with 10 times fewer interactions than is possible with the baseline approach.

Categories and Subject Descriptors
H.5.5 [**Sound and Music Computing**]: *Signal analysis, synthesis, and processing*

General Terms
Algorithms, Human Factors.

Keywords
Active Learning, Transfer Learning, Interface, Equalizer, Audio, Music.

1. INTRODUCTION
We seek to simplify interfaces in software for media production and align them with the user's conceptual model. In this work, we focus on quickly and automatically personalizing the interface of an audio production tool (the equalizer). Equalizers affect the timbre of a sound by boosting or cutting the amplitude in restricted regions of the frequency spectrum. They are widely used for mixing and mastering audio recordings.

Many equalizers have complex interfaces that lack clear affordances and are daunting to inexperienced users. This is because controls typically reflect either the design of preexisting analog tools or the parameters of the algorithm used to manipulate the sound, rather than how sound is perceived. Figure 1 shows the interface to a typical parametric equalizer plugin. It has 20 knobs, 9 push buttons and 18 radio buttons. The relationship between this interface and analog hardware is clear. The relationship between this interface and an acoustic musician's conceptual goal (e.g. "I want to make the sound more 'bright'") is not.

Figure 1.A parametric audio equalizer.

Typically, musicians who lack the technical knowledge to achieve a desired acoustic effect hire a professional recording engineer and verbally describe the effect. This process can be expensive, since it requires paying a human expert by the hour. It is also limited by the musician's ability to convey their meaning with language and the engineer's ability to translate that language into parametric changes. The well-known audio engineer John Burton describes the problem as follows [1]:

'[It is] a problem that has arisen in studios ever since the beginning of the recording age: how can you best describe a sound when you have no technical vocabulary to do so? It's a situation all engineers have been in, where a musician is frustratedly trying to explain to you the sound he or she is after, but lacking your ability to describe it in terms that relate to technology, can only abstract. I have been asked to make things more "pinky blue", "Castrol GTX'y" and … "buttery".'

We therefore, favor a new alternative that lets the artist directly control the device in terms of the desired perceptual effect. For example, the tool would learn what "buttery" means to the artist, and then create a slider to let her make the recording more or less "buttery," bypassing the bottleneck of technical knowledge. This approach has been adopted in work to dynamically individualize the mappings between human language descriptors and parameters for equalization and reverberation [2]. It has been commercialized in the equalization plug-in *iQ* [3], which has been positively reviewed in the audio engineering press [1]. This indicates this approach to building a controller (Section 3) is a useful paradigm that complements existing tools.

While this approach has been successful in creating a new interface paradigm, the current method [2] requires a relatively large number of user ratings (on the order of 25) to achieve high-quality results. This may discourage users. In this work we improve on that approach by guiding the learning with selective information requests (active learning), informed by previously learned concepts (transfer learning). This lets us create a useful personalized tool using only two or three user ratings.

2. BACKGROUND AND RELATED WORK

In audiology, there has been prior work on learning a listener's preference to setting the equalization curve of a small number of frequency bands in a hearing aid [4, 5] or cochlear implant [6]. The most common interaction procedure for doing this is known as the modified simplex procedure [7] but it is slow and is not informed by prior user data.

In audio engineering work applied to equalization, there has been work directly mapping equalizer parameters to commonly used descriptive words using a fixed mapping [8, 9]. A problem with these approaches is that mappings are extremely time-consuming to develop, the former requires a painstaking training process for each user, and the latter is brittle with respect to the variability that exists across users. Our work addresses both these issues.

A recent movement in HCI has sought to integrate algorithmic advances from the machine learning community [10]. For example, Simon et al. make use of machine learning techniques to automatically add chords to a user-provided melody [11].

The most closely related work to ours in domains outside of audio production tools is the CueFlik system [12], which correlates natural-language concepts to technical parameters. There is an important distinction between learning natural-language concepts in order to *select* existing objects (e.g., the CueFlik approach) versus learning concepts in order to *manipulate* the degree to which an object conforms to a given concept. Our work changes the artifact itself, based on an understanding of the user concept.

In terms of artistic creation for music, one stream of work uses new interaction techniques, often serving as new musical instruments or audio control surfaces [13]. Our work is related, but the use of transfer and active learning to use prior learned concepts is distinct. The Wekinator [14] is a music performance control mapping tool that lets informed users interactively control machine learning algorithms of their choice by choosing inputs, features, learning algorithm and parameters, and creating training example feature/parameter pairs. This work is complementary to ours, since it is for technically knowledgeable users, does not use language as a paradigm, and does not use active learning.

We use transfer learning [18, 19] when we use prior data to inform learning a new equalization concept. Our approach is akin to collaborative filtering techniques [24] since correlation to prior user responses guides selection of data to use. It is distinct in that we do not merely filter a database of existing objects, but rather create a new personalized tool.

Active Learning [15-17] refers to case where the machine selects the examples to learn from, rather than passively receiving examples chosen by the teacher. While we apply active and transfer learning to a new problem, we do not claim algorithmic advances in active learning, transfer learning or collaborative filtering. Therefore, the references are foundational and address the approaches we use in this work, rather than a review of the cutting edge in these areas.

Transfer learning and active learning have both been applied to user interfaces outside the audio domain. Previous approaches range from customizing user interface layout for a given environment to interactively personalizing results in content discovery or search and retrieval [20-22].

The only work [23] we are aware of that applies transfer learning of audio concepts to create an audio production tool is our own preliminary work. There, we used an approach that can only apply transfer learning in the limited subset of cases where prior user-concepts share a label with the current concept. We also did not use active learning. In the current work, we extend transfer learning to cases where prior concepts do not share a label with the new concept. We also apply active learning for query selection. We also present experimental results showing user ratings of learned controller effectiveness. The sum of all this is a significant advance over [23].

3. THE BASELINE SYSTEM

Before continuing with the focus of this paper: speeding the creation of a personalized controller by applying transfer learning and active learning, we outline the baseline controller learning method [1]. Since space is limited and details of the baseline learner are not the focus, we give an overview of the process here. We refer the reader to the prior work for more detail.

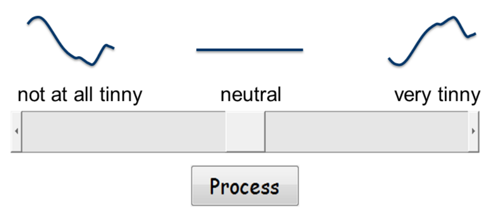

Figure 2. A learned controller for "tinny"

1. The user selects an audio file and a descriptor (e.g. "warm" or "tinny").
2. We process the audio file once with each of N 40-band probe equalization curves, making N examples.
3. The user rates how well each example sound exemplifies the descriptor.
4. We then build a model of the descriptor, estimating the effect of each frequency band on user response. This is done by correlating user ratings with the variation in gain of each band over the set of examples. The slope of the resulting regression line for each frequency band indicates the relative boost or cut for that frequency.
5. The system presents to the user a personalized controller (Figure 2) that controls filtering of the audio based on the learned model of how to manipulate audio.

While the use-case presented here is equalization, this basic approach has been used to build personalized equalizers, compressors and reverberators. All these cases requires rating roughly 25 to 30 examples to generate a good controller [2].

Although 25 interactions may often be acceptable, many may not have the patience for this. We will now describe how to speed learning through application of knowledge from a database learned from prior users. This will allow learning a good controller from just two or three user-rated audio examples. While we apply it to equalization here, the approach is not specific to equalization and applies to all of the domains where this controller personalization approach has been used.

4. APPLYING TRANSFER LEARNING

Define a *user-concept* as a concept (in this work, concepts are sound adjectives) taught to the machine by a particular user (i.e. Bob's concept for "warm" sound). If two users teach the system the same word, then there are two user-concepts (Bob's "warm" and Tolga's "warm"). In the case of equalization, this results in two equalization curves. Figure 3 shows learned equalization curves for three user-concepts.

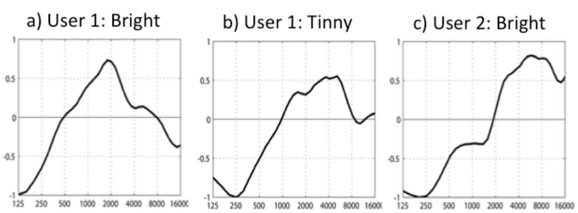

Figure 3. Three learned equalization curves for user-concepts.

While each user is unique, user-concepts may be related, even when they do not share a label. Figure 3 shows the equalization curve for User 2's "Bright" is more similar to User 1's "Tinny" than it is to User 1's "Bright." If a prior user-concept is similar to the current user-concept, then user responses to training examples for this prior concept may help in learning the current user-concept, even if the two do not share a label.

To determine which prior data is most informative, we measure similarity of user responses to the training examples in the course of teaching the system user-concepts. This is akin to collaborative filtering techniques [24] since correlation to prior user responses guides selection of data to use. It is distinct since we do not merely filter a database of existing objects. Instead we use prior user responses to guide creation of a new controller.

Note that our focus is on creating a personalized controller, rather than on new approaches to transfer learning or collaborative filtering. Therefore, the method we use was chosen for clarity and appropriateness. Experimental results (Section 7) show it is effective. A comparison of different transfer learning approaches on this problem is outside the scope of this workshop paper.

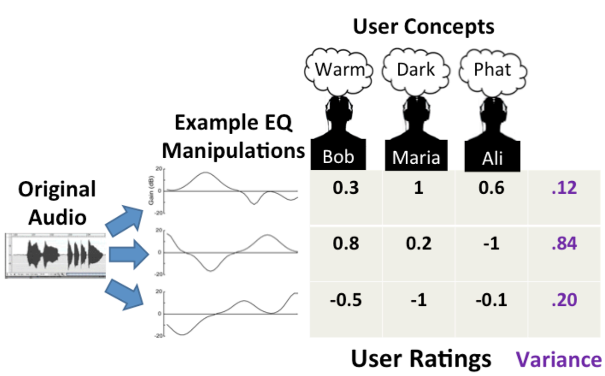

Figure 4. An audio file is manipulated with *m* equalization curves to create *m* examples. Each user rates all examples in terms of their own concept (e.g. "How 'warm' is this example). Ratings range from 1 to -1.

We create a fixed question set by manipulating a standard audio file (e.g. a 5 second passage from Delibes' *Flower Duet*) using a tool, such as an equalizer. Do this ***m*** times (on the order of 50), creating a set of examples ***M*** to be rated by users. Selection of this set is important, but outside the scope of this work. In this study, we use a set of EQ curves found to be effective in a previous study [2] and applied them to audio with significant energy across the frequency spectrum (see Section 6.2). All users rate the same set of examples for every user-concept.

To do transfer learning we put an existing set of user-concepts into a vector space defined by user ratings. Let ***Q*** be a subset drawn from set ***M*** of examples rated by users. The set ***Q*** may be drawn randomly or with a smarter selection criterion (Section 5). Each user–concept's location is determined by that user's ratings of the examples in ***Q*** when training the system on the concept. Figure 5 shows the user-concepts from Figure 4 in a space defined by user ratings of examples 2 and 3 from Figure 4.

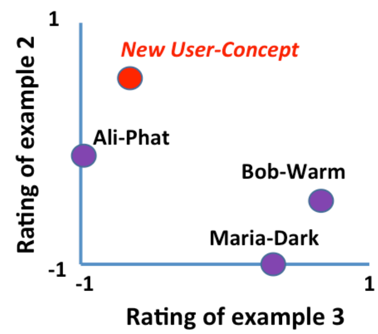

Figure 5. Concepts from Figure 4 placed in a 2-d space based on user ratings of the 2nd and 3rd examples from Figure 4.

When training the system on a new user-concept, rather than asking the user to rate the full set of ***M*** examples, we ask them to rate only the subset ***Q***, placing the new user-concept in the user rating space. Then, we estimate the current user's ratings for the remaining ***M-Q*** examples by a weighted combination of user responses for user-concepts learned previously. Call the set of existing user-concepts ***U***. We use these estimated ratings in the concept training procedure for the new user's concept.

4.1 Distance and Weighting

Given a new user-concept (Maria's "dark"), the user (Maria) rates the examples in the query set ***Q***. We estimate the user's ratings for the remaining ***M-Q*** examples by using a weighted combination of past user ratings for previous user-concepts.

Let there be a set of prior user-concepts, for which users have each rated all the examples in ***M***. Call this set ***U***. The weight for a prior user-concept u should go down as the distance been u and the new user-concept v increases. In a pilot study, we tried a variety of p-norms and found Manhattan distance (Equation 1) performed well. Here, $r_u(q)$ and $r_v(q)$ are the ratings given to example q for user-concepts u and v.

$$d(u,v) = \left(\sum_{q \in Q} \left| r_u(q) - r_v(q) \right| \right) \qquad (1)$$

The weight of user-concept u is determined by the distance as shown in Equation 2.

$$w(u) = \frac{\exp(-2d(u,v)^2)}{\sum_{k \in U} \exp(-2d(k,v)^2)} \qquad (2)$$

We can then estimate the rating the user will give to un-rated example ***q*** with a weighted sum of prior user-concept ratings for that example (Equation 3).

$$\tilde{r}_v(q) = \sum_{u \in U} w(u) \cdot r_u(q) \tag{3}$$

Our default approach to transfer learning does not restrict the pool of prior user-concept data (set ***U***). All prior learned data from all users and all concepts is employed. In our experimental results, this is referred to as the ***Pooled Transfer Learning*** approach. The intuition here is that a previous user-concept (User 1's "tinny") may be similar to the current concept (User 2's "bright"), even if they have different labels (e.g. Figure 3). A second approach, which we called ***Same-word Transfer Learning***, applies transfer learning only to data collected from other users training the system on the same concept word that the current user is teaching the system.

5. APPLYING ACTIVE LEARNING

We now address the question of which subset (i.e. ***Q***) of the examples in ***M*** can best locate the current user-concept in the space of prior learned user-concepts. This can effectively be done by asking questions that best let us differentiate between prior user-concepts. Selecting examples that caused greatest disagreement is a clear way to do that. In this case, we apply a query-by-committee active learning method [15]. This approach was selected for simplicity and appropriateness. A comparison to other approaches is outside the scope of this workshop paper.

For a user-concept, the system presents example manipulations of an audio file to be rated by the user. Given a pool of prior user-concepts ***U***, where all users rate the same set ***M*** of audio examples, one can measure the variance of responses for each example across all prior users. An audio manipulation with high variance among user responses is a promising query. The wide spread of responses makes it easier to distinguish which existing user-concepts are closest to the new concept the system is attempting to learn.

Consider Figure 4. The example in the top row generated responses that were in broad agreement: all positive, with a low variance among ratings. The second row shows a set of responses that range from positive to negative, with a large variance. Asking the user to rate the example that generates disagreement between concepts will give much more information. Our system ranks the examples in ***M*** by the variance of prior user ratings. The query set ***Q*** is formed by selecting the top |***Q***| examples, as measured by variance of user ratings.

6. EXPERIMENTAL DESIGN

We have argued that our approach should let us build personalized relevant controllers after many fewer ratings than the baseline and will do so even when prior user-concepts do not share a label with the user-concept being learned. We now establish whether this is true by answering the following questions:

1) Does the baseline system create controllers that manipulate the audio as desired?
2) How much (if at all) does transfer learning speed (or improve) learning for our problem?
3) Is combining active learning with transfer learning superior to transfer learning alone?
4) What is the appropriate pool of prior experience to draw from when applying transfer learning?

To answer these questions, we had a set of 35 users train the system. We now explain these experiments in detail

6.1 Descriptive Terms

In deployment of a system, we expect people to use a multitude of terms we cannot predict. However, to evaluate the effectiveness of transfer learning in a controlled environment we required participants use the same set of terms, so that we could build a sufficiently large pool of same-word user-concepts.

We selected five adjectives for study participants to teach the system: "muffled", "tinny", "broad", "bright" and "warm". The terms selected were chosen in consultation with a Ph. D. in Speech and Hearing Science with experience as a recording engineer. The words "bright", "tinny" and "muffled" were selected because we felt they would map well onto the manipulations an equalization tool performs. We expected "warm" to be a border case requiring both equalization and reverberation to be truly captured. We included "broad" because we intuitively felt this word would map onto spatialization parameters (e.g. panning) better than to equalization.

6.2 Stimuli

The stimuli were always manipulations of a short (5 second) musical passage from Delibes' Flower Duet at the compact disc standard bit depth and rate (16 bits at 44.1 kHz). The Flower Duet was chosen for its broad spectral coverage and ease of repeated listening. We used a query set of 50 equalization curves found to be effective in previous work [2]. The excerpt from the Flower Duet was manipulated once by each of the 50 curves, creating 50 manipulated examples (the set ***M***). The same 50 examples were presented in randomized order for every word concept taught to the machine by all study participants.

6.3 Participants

35 people (hereafter called *users*) participated in the experiment. Thirteen were female. Average age was 27.8 years. All reported normal hearing and were native English speakers. 24 users reported at least 2 years of experience with a musical instrument. 10 reported 2 or more years of audio equipment experience.

6.4 Data Collection Procedures

Users were seated in a quiet room with a computer that controlled the experiment and recorded user responses. The stimuli were presented binaurally over headphones. Each user took part in a single one-hour *session*. Each session was grouped into five *runs*. In a run, the user was presented with a single word (e.g. bright) and asked to teach the system their concept for that word by rating a set of example audio files on how well each example embodied the concept. Word order was randomized for each user.

For each run, there were 70 *trials*. One trial is the rating of a single audio example on how well it embodies the concept word. In each run, 50 audio examples were unique. 20 repeated prior examples. This resulted in 20 matched-pairs of responses to the same stimuli for each run. The repeats were selected randomly on each run. Presentation order was randomized for each run. Ratings were given by moving a slider that ranged from -1 (the opposite of the concept) to 1 (perfectly embodiment of the concept).

At the end of a run, the system learned an equalization (EQ) controller (Figure 2) for the concept word (e.g. "bright"). To serve as a baseline, the system also learned an EQ controller from 50 randomly-generated synthetic ratings distributed uniformly on the interval (-1,1). For each word, the user was asked to rate the two control sliders. Rating was on a scale from -1 (learned the opposite concept), to 1 (learned concept perfectly). The presentation order of the controllers was randomized.

We examined user consistency for each user-concept by finding the Pearson's correlation coefficient between a user's first and second rating on the 20 examples that were presented twice during the run. We removed lowest 5% of user-concepts (ones with a consistency of less than 0.19), leaving 162 (at least somewhat) consistent user-concepts, out of 175.

7. SYSTEM EVALUATION

7.1 Users' Ratings of Controllers

Figure 6 shows boxplots for user satisfaction with controllers taught to the system, broken down by word. The red line inside each box indicates the median value. The text on the Y-axis was provided to users when evaluating controllers as a guide to scoring. As expected, "broad" was the least satisfactory term, although the median satisfaction with even this word was high enough to provide a useful controller.

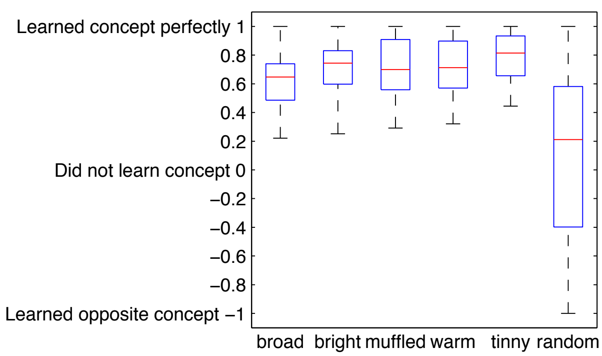

Figure 6. User satisfaction with equalization controllers learned for each of five words. "Random" indicates user satisfaction with controllers generated from random data.

User ratings for controllers built from user responses were consistently better than the scores for the controller learned from random data (according to a pair-wise Student's t-test, $p < 0.001$).The mean rating given by users for the controllers learned using our baseline method was 0.70 (SE=0.017), on a scale of 1 to -1. The mean user rating of controllers learned from random data was 0.12 (SE=0.043). This indicates that the baseline method generates controllers that users find relevant. This answers our first experimental question: The baseline system creates controllers that manipulate the audio as desired.

7.2 Learning Method Evaluation

For the 162 consistent user-concepts, we collected user ratings of the 50 unique audio examples all users rated. We could then simulate the effectiveness of the baseline learner (without transfer or active learning) by selecting each user-concept and building a concept model from a randomly-selected set of n rated examples. Using the learned concept, we predicted user responses on the remaining rated examples and measured the correlation between machine predictions and actual user ratings.

We generate a prediction of the user's rating of a new example by comparing the learned concept EQ curve to the probe EQ curve applied to a new example. The more similar they are, the higher the predicted rating of that example. Similarity is measured by their cross-correlation, which generates a value between -1 and 1. This can be compared to the user's actual rating for the new example. Given a set of rated examples, $\boldsymbol{M}$, the *machine-user correlation* is the Pearson correlation coefficient of user ratings to machine-generated ratings for the entire set of rated examples.

For a given value of n (number of rated examples) this gave 162 machine-correlation values (one per user-concept). We did this for each value of n from 1 (a single rated example) to 50 (all the rated examples for that user-concept), calculating the mean machine-user correlation across the 162 user-concepts. This formed our baseline learning method.

We then repeated this process for each learning method. *Pooled transfer learning* utilized data from n randomly selected ratings by the user, augmented by all of the example ratings from the 161 other learned user-concepts. The data from the remaining concepts was weighted using the n-dimensional Manhattan distance measure described earlier, where n is the number of rated examples for the current concept. *Same-word Transfer learning* was the same as Pooled transfer learning, except only those user-concepts with the same word descriptor were used.

Pooled Transfer + Active learning was identical to Pooled Transfer learning, except that the n examples were not randomly selected. Instead the n with the greatest variance in user responses across the other 161 user-concepts were used. *Same-word Transfer + Active learning* was identical to Pooled Active + Transfer learning, except that the pool was restricted to user-concepts with the same word label.

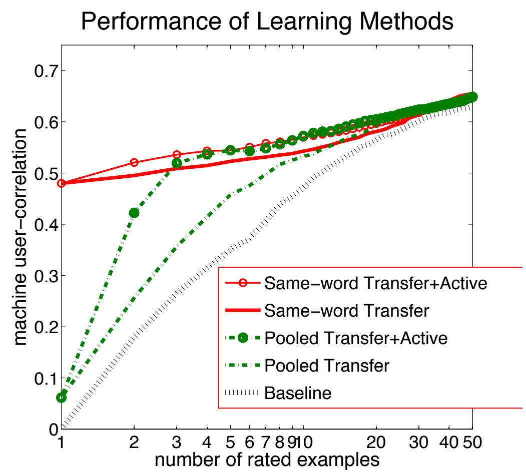

Figure 7 Mean machine-user correlation for each learning method, averaged over all words. The relative performance of these methods remained identical for each of the five descriptive words taught to the system.

Figure 7 shows the learning curve for all these methods. As the figure shows, all methods that use transfer learning outperform the baseline method. Same-word learning outperforms Pooled for one or two examples. Once three examples are rated, the performance of both methods are indistinguishable. For both Pooled and Same-word learning, active selection of training examples shows a noticeable effect. The effect of active learning is significantly stronger for Pooled transfer learning. These results were consistent for each of the five words used in training. Space limits preclude showing the individual-word data.

This answers experimental question 2: learning user concepts with transfer learning is much more effective than learning user concepts using no prior knowledge. It also answers question 3: active selection of training examples significantly improves transfer learning of concepts in this domain.

These results were confirmed by an analysis of covariance. A pair-wise comparison of all learning methods revealed the observed differences between learning methods were statistically significant (by Tukey's HSD $p < 0.001$). Learning methods were ordered, from best to worst as follows: Pooled Transfer+Active, Pooled Transfer, Same-word Transfer+Active, Same-word Transfer, and Baseline learning.

The final experimental question is “What is the appropriate pool of prior experience to draw from when applying transfer learning?” Figure 7 indicates that, if a prior user has already taught the system the word currently being learned, same-word transfer learning is the best choice. That said, if the user can answer three or more questions, pooled transfer learning with active learning does as well or better. Pooled transfer learning has the added benefit that it does not require any user-concept in the database share a label with the current concept.

8. CONCLUSIONS

Using a simple approach, we have demonstrated significant improvements in the number of user-ratings needed adequately learn a desired equalization controller from user feedback. A previous method required roughly 25 user-rated examples to yield an effective controller. Through transfer learning and active learning, we can reduce this to between one and three ratings.

This work promises to enable useful on-the-fly tool building in the recording studio or for home-studio use (e.g. Apple’s Garage Band) A user unfamiliar with existing equalizer (or reveberator, or compressor) interfaces could quickly (after answering just two or three questions) create tools to manipulate audio within the terms defined by the user. Another interesting application area is to let hearing-aid users adjust their hearing aids in the field.

9. ACKNOWLEDGMENTS

This work was funded by USA National Science Foundation Grant 0757544.

10. REFERENCES

[1] Burton, J., "Ear Machine iQ: Intelligent Equaliser Plug-in for MacOS & Windows," *Sound on Sound Magazine*, 2011, Sound on Sound Ltd.: United Kingdon.

[2] Sabin, A., Rafii, Z. and **B. Pardo**, “Weighting function-based rapid mapping of descriptors to audio processing parameters,” *Journal of the Audio Engineering Society,* 2011, vol. 59(6): p. 419-430

[3] Ear Machine's iQ equalization plugin. Available from: http://www.ear-machine.com/iQ.html.

[4] Neuman, A.C., et al. , "An evaluation of three adaptive hearing aid selection strategies," *Journal of the Acoustical Society of America*, 1987. 82(6): p. 1967-1976.

[5] Durant, E.A., et al., "Efficient Perceptual Tuning of Hearing Aids with Genetic Algorithms," *IEEE Transactions on Speech and Audio Processing*, 2004. 12(2): p. 144-155.

[6] Wakefield, G.H., et al., "Genetic Algorithms for Adaptive Psychophysical Procedures: Recipient-Directed Design of Speech-Processor MAPs," *Ear Hear*, 2005. 26 (4 Suppl.): p. 57S–72S

[7] Kuk, F.K. and N.M. Pape, "The reliability of a modified simplex procedure in hearing aid frequency-response selection," *Journal of Speech and Hearing Research*,1992. 35(2): p. 418-429.

[8] Mecklenburg, S. and J. Loviscach, "subjEQt: Controlling an equalizer through subjective terms," *Extended Abstracts of the Proc. of ACM CHI 2006.* NY: ACM Press. : Montreal, Canada. p. 1109-1114.

[9] Reed, D., "Capturing perceptual expertise: a sound equalization expert system," *Knowledge-Based Systems*, 2001. 14: p. 111-118.

[10] Morris, D. and J. Secretan, "Computational creativity support: Using algorithms and machine learning to help people be more creative," *Ext.Abs. of the Proc. of ACM CHI 2009*, Boston, USA. p. 4733-4736.

[11] Simon, I., Morris, D., and Basu, S. , "MySong: Automatic accompaniment generation for vocal melodies," *Proceedings of ACM CHI 2008*, Florence, Italy. p. 725-734.

[12] Fogarty, K., Tan, D., Kapoor, A., and Winder, S. , "CueFlik: Interactive Concept Learning in Image Search," *Proc. of ACM CHI 2008,* Florence, Italy. p. 29-38.

[13] Fiebrink, R., D. Morris, and M.R. Morris. "Dynamic mapping of physical controls for tabletop groupware," *Proceedings of ACM CHI 2009.* Boston, MA.

[14] Fiebrink, R., D. Trueman, and P.R. Cook, "A meta-instrument for interactive, on-the-fly machine learning," *Proceedings of New Interfaces for Musical Expression (NIME) 2009*, Pittsburgh, PA.

[15] Cohn, D., Atlas, L., and Ladner, R. , "Improving Generalization with Active Learning," *Machine Learning*, 1994 15: p. 201-221.

[16] Settles, B., Active Learning, *Synthesis Lectures on Artificial Intelligence and Machine Learning* , Morgan & Claypool, 2011. ISBN: 9781608457250

[17] Bloodgood, M.a.C.-B., C. Bucking the Trend: Large-Scale Cost-Focused Active Learning for Statistical Machine Translation. . in Proc. of the Association for Computational Linguistics(ACL 2010). Uppsala, Sweden

[18] Pan, S.a.Y., Q. , "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, 2010. 22(10): p. 1345-1359.

[19] Ando, R. and T. Zhang, "A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data," *Journal of Machine Learning Research*, 2005. 6(Nov):p. 1817-1853

[20] Gajos, K.a.W., D.S. , SUPPLE: Automatically Generating User Interfaces. , *Proc. of Intelligent User Interfaces (IUI 2004)*, 2004, Funchal, Portugal. p. 93-100.

[21] Lin, J. and J.A. Landay, Damask, "A Tool for Early-stage Design and Prototyping of Multi-device User Interfaces," *Proc. of the 8th International Workshop on Visual Computing*, 2002. p. 573-580.

[22] Nichols, J., Myers, B.A., Higgins, M., Hughes, J., Harris, T.K., Rosenfeld, R., and M. Pignol, "Generating remote control interfaces for complex appliances," *UIST '02 - Proceedings of the 15th annual ACM Symposium on User Interface Software and Technology*, 2002, Paris, France. p. 161-170.

[23] Pardo, B., Little, D. and D. Gergle, “Towards Speeding Audio EQ Interface Building with Transfer Learning,” *Proceedings of New Interfaces for Musical Expression (NIME) 2012,* Ann Arbor, MI, USA, May 21-23, 2012

[24] Su, X. and T.M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, 2009. Volume 2009 (2009)