Simultaneous Naming and Configuration in Natural Language Interfaces

Qualifying Exam Proposal

Ana Ramírez Chang
University of California, Berkeley
anar@cs.berkeley.edu

April 23, 2007

1 Introduction

Most visions of ubiquitous computing including Weiser's argue that it should be "calm": almost invisible except during direct (focal) interaction. The interaction should also be well "fit" to the environment it supports. This is a particular challenge when a user wants to control an unfamiliar environment: whatever modality is used, user and computer must share conventions about the effects of the user's action. There is often no "direct manipulation" scheme and no obvious affordance for room control. In the worst case, this may lead to eccentric command conventions ("restaurant 2.0," "kitchen 3.1" etc.). We argue that an ideal environment should respond to users' own natural forms of expression. For open-plan room control, we believe speech interfaces are a very good option [1]. With distributed microphone technology the physical interface all but disappears, but jumps fluidly to the foreground when the system responds to spoken input. Speech is also the most natural form of human expression. The challenge remains how to interpret natural speech, and how to align users' *situated* understanding of the world with the system's.

We are exploring natural interaction for a room lighting application. The target environment is a 2,300 sq ft workspace which is semi-structured: it includes both dedicated space (cubicles, sofas, kitchen area, workshop) and a large amount of multi-use space for seminars, group and individual meetings etc. The room's lighting is very flexible: there are 79 independently-dimmable compact fluorescent downlights. The lights can be individually adjusted using a simple web interface. But this is difficult in the multi-use space (which contains no fixed furniture). As a practical matter, users often find it more convenient to turn all the room lights on rather than a small subset. Like many open-plan office spaces, the result is that many lights are left on for hours and energy consumption is

[1] Weiser explicitly critiqued speech-based agent interfaces in one of his well-known papers. However, his specific criticisms related to many aspects of design that we avoid: agent personality, knowledge, an "identity" you interact with and impart human-like traits to. By contrast, our use of speech is highly situated, in a shared context, and in short focal interactions followed by return to invisibility.

needlessly high. We have experimented with various types of remote control but ultimately, speech input seems the best option for calm lighting control in this space. The space is equipped with a large array of microphones in the ceiling to support speech input from any location, although this functionality is not supported yet. The present paper therefore concentrates on (natural language) *text* input from users.

The lighting control problem although simple, nevertheless raises two of the generic challenges for natural human-machine interaction: (i) discovering how users give commands and (ii) discovering the configurations *named* by those commands. That is, it is not just a problem of building a language model or grammar for user input, but in discovering the semantics of those inputs. Users name particular parts of the room ("kitchen area," "soft space" etc.), but these names must be matched to settings of groups of lights. We believe this theme occurs in many ubicomp environments: e.g. mapping the "semantic gap" from a configuration command like "presentation" to control of lights, projector, and sound devices in a workplace, or from a request like "high priority calls only" on a mobile device to a specific subset of the user's contacts who are important enough to trigger a ring response. We will call this problem the "SNAC problem," for "Simultaneous Naming And Configuration". Since users may have their own preferences, SNAC should be personalized when individual user history is available. On the other hand, SNAC systems should still do something reasonable for irregular users of a space (visitors), such as choosing a most popular configuration from previous user data.

For my thesis, I plan to propose and evaluate an approach to SNAC for the lighting control application. The approach uses a WoZ (Wizard of Oz) who receives natural language text requests from users, and who manually adjusts the lights according to his understanding of the user's command. Users occasionally enter further text if they are not happy with the Wizard's configuration, and the wizard tries to correct it. The final configuration, along with the user's input command, is recorded. The command-response pairs are then used to train a pattern recognizer which attempts to find the most common lighting patterns with command keywords used to invoke them.

We have run an initial wizard-of-Oz study, collecting a few hundred command-response pairs over several weeks and explored two different pattern recognizers (SVD and NMF) on two different feature sets. We found promising results from the NMF pattern recognizer on a feature set with difference data (differences in the intensity of the lights) and frequency vectors (unigrams) of the messages that resulted in the change in intensities.

Building on the initial results, I plan to address four main challenges in my thesis. 1) I will iterate on the pattern recognizer for lighting patterns. 2) In addition to a pattern-recognizer for names sets of objects, I will also train a pattern recognizer for the *actions* performed on the named sets of objects. 3) Given a message, I will need to figure out how to use the pattern-recognizers to interpret the message as to what action should be taken. 4) I will build an instant message bot that will allow members of the lab to manipulate the stat of the lights through natural language instant messages.

I will evaluate the approach before implementing the instant message bot with a wizard-of-Oz study similar to the initial study. Instead of having the wizard interpret the messages, the wizard will rely on the interpretation from the pattern-recognizers to decide what action to take. This will allow an *in situ* evaluation of the approach to SNAC for the lighting control application based on users' reactions and feedback.

The rest of the proposal is organized as follows. Section 2 situates the thesis contribution in the related work. Section 3 describes the proposed method. Section 4 describes the initial wizard-of-Oz study. Section 5 describes the methods (SVD and non-negative matrix factorization) for discovering text/configurations. I present the initial results in Sec. 6. I describe the four challenges I plan to address in my thesis in Sec. 7, I describe the proposed evaluation in Sec. 8 and conclude in Sec. 9.

2 Related Work

Below I describe the work related to designing natural speech interfaces, previous applications of factorization algorithms such as SVD and NMF, and existing smart home and home IT projects.

2.1 Natural Speech Interface Design

Whether a speech interface is: (a) grammar-based as many successful commercial systems are (HeyAnita now Kirusa [Kir07], BeVocal [Nua07], Tellme [Net07]); (b) uses a robust method such as a statistical language model [BCDP^{+}90], (c) or uses a combination of both [RBH^{+}04], most speech interfaces map a term to one concept, action or object. For example, in the RoomLine room reservation system [Boh07] , the user says the time of the day, the date, and the size of the room they would like to reserve, and the system replies with which rooms are available. A tool like Regulus [RHB04] can help figure out a collection of terms that map to one concept, such as the size of a room or a date, but the designer has to know what the objects or actions are that a user might refer to. We propose a method to figure out what collection of objects, actions or concepts the user may want to refer to with one phrase. For example, the user may want to turn on four lights together, and refer to those four lights as "the lights over the public PCs."

It is common practice to use a wizard-of-Oz study [DJ93,FG02] to collect a corpus of speech data on which to train a statistical language model for a speech interface. You might also run a wizard-of-Oz study with a simulated automatic speech recognition channel to get more realistic data about the interaction with the user [SWY04]. Our approach uses these practices, but also collects command response to build a SNAC model, which is novel.

2.2 Applications of Matrix Factorization

Singular value decomposition (SVD) is a commonly used tool for extracting patterns in various signals. While the SVD approach assumes the patterns are orthogonal, it is often applied with reasonable outcomes even when this does not hold: e.g. LSA (Latent Semantic Analysis) is an SVD of the document-term frequency matrix for a corpus, and has been widely used for text analysis. SVD analysis is closely related to eigenvalue analysis, and in fact the "eigen" prefix is often used to describe an SVD as in "eigen-faces" (SVD for face recognition), "eigen-taste" (SVD for collaborative filtering) or "eigen-behaviors" work [EP06] for space-time patterns of user movement.

In recent years, alternative factorization methods such as least-squares NMF (Non-negative Matrix Factorization) have found favor over SVD in cases where patterns are not orthogonal. This is especially true when the patterns appear "non-negatively." This is indeed the case for both light intensities and for word frequencies the user commands. So NMF seems like a natural candidate for SNAC analysis. Furthermore, NMF has been shown to be superior to SVD for pure text clustering [LS99], or for image segmentation [LS99] which is similar to our light grouping problem. Barnard et al. match words and pictures [BDdF$^+$03] using an "aspect" probabilistic model, which is another type of factor model. Other candidate factor models include Latent Dirichlet Analysis (LDA) [BNJ02] and GaP [Can04]. These methods add prior probabilities to the factors and use likelihood measures (rather than least squares) to fit the original data. We did not use these methods because: (i) when there is enough training data, the factor priors have little or no effect and (ii) least-squares NMF has been shown to produce better (more independent) factors compared to KL-divergence (likelihood) fitting methods [LS99]

2.3 Smart Homes and Home IT

Predictive light automation has been studied in [Moz05]. The authors of that work acknowledge that conventional light control interfaces have drawbacks even in the home (e.g. users often don't turn things off) and use a learning system to automate the lights. Quesada et.al. address the interface challenge in the home machine environment [QGS$^+$01] with a speech interface for lighting control in the home. Juster and Roy use situated speech and gestures to control a robotic chandelier, Elvis [JR04]. Their work focuses on how to move the chandelier arms to achieve a given lighting scene. They use keyword spotting and hand selected keywords to analyze the semantic content of the speech. Instead of focusing on how to achieve a given lighting scene, we focus on analyzing the semantic content of the messages.

3 Proposed Method

We would like our natural interface to allow users to refer to and manipulate the set of lights of their choice, with a vocabulary that makes the most sense to them. To that end we want to find the sets of lights commonly manipulated by our participants, and the terms they use to refer to those sets of lights.

As is common when designing a speech interface, we ran a wizard-of-Oz study to better understand the interaction with the user as well as collecting vocabulary data so that the system will understand the vocabulary the users tend to use. In addition to the traditionally collected data, we also collected data about the state of the lights. We recorded all of the changes to the state of the lights with timestamps, and we recorded all of the messages the users sent asking the state of some lights to be changed. At the end of the study we had co-occurrence state change data and messages requesting the state change. In the next two sections, we describe the study and the data analysis in detail.

4 Wizard-of-Oz Study

4.1 Description of the Workspace

The initial study took place in the 2300 square foot semi-structured workspace that currently has sixteen residents (researchers). Six graded-awareness cubicles occupy half the room, the other half is a multi-use space for meetings, presentations, ad-hoc teaming or individual work. The multi-use space has a presentation screen, a "soft space" with a couch and chairs, and a set of four computers for visitors to the lab to use. There is also a tool shop in one corner of the room. Figure 1 shows the floor plan of the room. The room has 79 individually-controllable compact fluorescent lights mounted overhead. The intensity of each of the lights can be controlled over the network via a web interface. All occupants of the lab have access to the web interface. They can access it from their personal computers, or they can access it on one of the public machines. One of the public machines always had the web interface open.

4.2 Study Details

We ran a wizard-of-Oz study with ten researchers who regularly work in the room diagrammed in Fig 1. Nine of the participants have desks in the room, and one of the participants sits in the adjacent room, but makes heavy use of the multi-use area and the tool shop. There were six researchers who have desks in the room that did not participate in the study. Two of them were out of town while the study was run, three did not volunteer to participate, and the last one was the wizard. The researchers not participating in the study used the web interface to control their own lights.

The study participants were asked to send the wizard an instant message each time they wanted to change the state of the lights in the room. For example, the participant might write *"turn on the two lights over my desk"* or *"turn off the lights over the public machines."* The wizard asked any necessary clarification questions (via instant messages), and then used the web interface to make the requested change to the state of the lights.

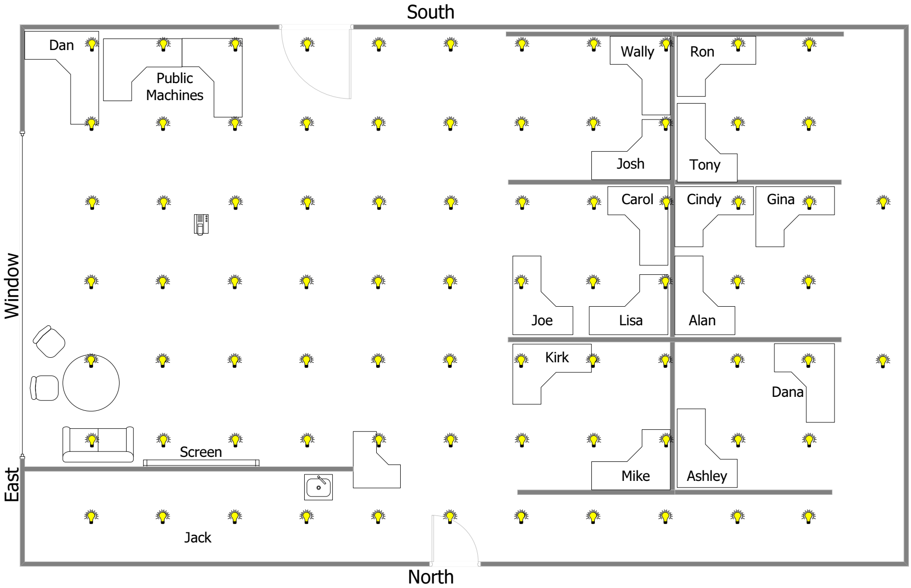

Fig. 1. Map of the room that was used for the wizard-of-Oz study. Each of the 79 individually-controllable lights are shown as light bulbs. Each of the 16 researchers' desks are shown, as well as the desks for the public computers. (names are anonymized)

If the wizard was not available to change the state of the lights, the participants were asked to type the message they would have sent to the wizard into the web interface, and then make the changes themselves with the web interface. This way data was still able to be collected if the wizard was not at his desk.

4.3 Data Collected

We collected twenty three days of data, including over 230 light state changes, and 306 messages, with 175 different words (not including stop words). Figure 2 shows the number of messages each person sent (not including messages confirming the change in the lights was correct).

Light State Changes A light state is a set of intensity values, one for each of the 79 lights. A light state change is the difference between two light states, where at least one of the values is non zero. Each message is tagged with the participant's name. From here on, we will refer to a light state change vector simply as a change vector.

Messages The users used a variety of different ways to request changes in the lighting scene. The different kinds of messages collected include: messages that refer to two sets of lights; messages that use direction phrases such as southeast;

messages referring to the amount of light in the room; messages asking all the lights to be turned on or off; messages that refer to the person speaking; messages with timing information as to when the change should take place; and messages that excluded lights in the description of which lights to change. Figure 3 shows the different kinds of messages collected and some examples of each kind of message.

In terms of Speech Acts [Sea5c], the messages are *directives*, or illocutionary acts where the goal is to get the addressee to do something. Austin introduced *perlocutionary acts* and *illocutionary acts* [Aus62] and Searle described five types of illocutionary acts. Perlocutionary acts are acts which result in an action. Illocutionary acts are the act of getting the audience to recognize the speaker's meaning. Searle's five illocutinary acts are: *assertives* — the act to get the addressee to form or attend to a belief; *directives* — the act to get the addressee to do something; *commissives* — the act to commit the speaker to doing something; *expressives* — the act to express a feeling toward the addressee; *declarations* — these rely on organized convention of institutions. [2] Directive illocutionary acts can be requests for action as with most commands and suggestions, or requests for information as with most questions. Some of the messages are phrased as commands, such as *"Turn up the lights near the sink."* The other messages are phrased as questions, such as *"Can I get the lights on in the design research cubicle?"* The latter message is literally asking a yes or no question, not giving a command. Searle calls the literal illocutionary act the *secondary illocutionary act* [Sea69] and the intended illocutionary act the *primary illocutionary act*. From the messages themselves its hard to tell what the primary illocutionary act is, but from the context in which the message is received (the interface to the lighting control system), we can safely assume the messages are commands to the system (or the wizard in the initial data).

5 Methods: SVD and NMF

Given a set of state change vectors and corresponding messages that "caused" the state change, we would like to find groups of lights that commonly change together, and a set of terms that refer to those changes. We assume the 79 lights fall into k (possibly overlapping) groups each of which corresponds to a set of lights that are commonly manipulated at the same time. Each state change (a set of 79 values representing the change in intensity of each light) either completely belongs to a particular group, or covers multiple groups. We project the set of state changes into a k-dimensional semantic space in which each axis corresponds to a particular group of lights. Each state change can be represented as a linear combination of the k group. Our approach is a *factorization* rather than a *clustering* approach. While similar, factorization is more general because it allows overlapping groups, and graded membership in each group.

The values in the state change vector range between -255 and 255 because the range of possible intensities range between 0 and 255 and the change in

[2] The descriptions of Searle's illocutionary acts are from [Cla96].

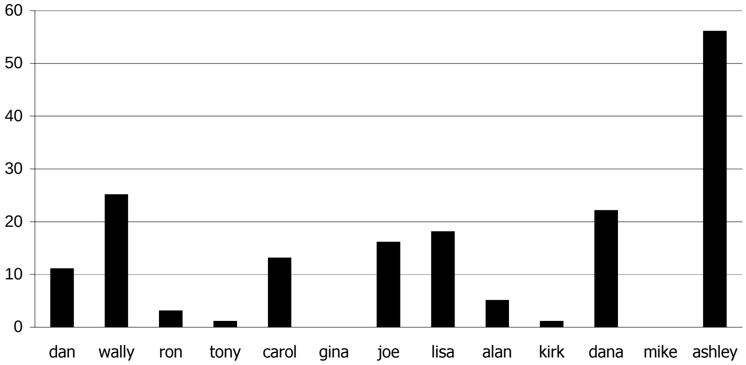

Fig. 2. The number of message sent per person (not including messages confirming a change in the state of the lights was correct). Gina and Matt did not send any messages because they did not participate in the study, but their names still show up in the results because participants in the study referred to lights in their cubicles by using their names.

intensity can be positive or negative. At first glance, the range of the values in the state change vector seems to lend itself to an eigenvector solution using singular value decomposition (SVD), but SVD assumes the factors are orthogonal and this often turns out not to be the case. For example, if two researchers share a cubicle, both of them may turn on all the lights in the cubicle, and each researcher may sometimes only use the lights over their desk, depending on the amount of ambient light in the room from the window. This would result in a set of factors that are not orthogonal.

Values in the state change vector range from -255 to 255, but NMF only deals with positive factors. On the other hand, when looking for sets of lights that are manipulated together, it doesn't matter if the intensity of the lights is increased or decreased, it only matters what has changed. So we use absolute values of change vectors for NMF analysis. For SVD, we tried both signed or absolute value of changes. It did not seem to matter much to the resulting factors, and the data shown are for absolute values for easier comparison to NMF. As discussed earlier, both SVD and NMF have been applied to text analysis and image segmentation tasks. So it is natural to apply them to combined datasets comprising text terms and the light state changes.

5.1 Data Representation

Since we want to find the groups of lights that are commonly manipulated at the same time, we want to represent the changes to the state of the lights in our matrix. We would also like to find the terms that are most commonly used to refer to a group of lights, so we represent the messages that lead to a state change in our matrix. For each state change, we have a change vector of intensity difference values, one for each of the p lights. Each state change vector has

- **Messages referring to two sets of lights.**
  - Joe: *Can I get my desk light on pretty bright and the other cubicle lights, light but a little dimmer than normal.*
  - Joe: *Design research lights up please, and my light a little more than the others.*
- **Messages that use direction phrases to refer to lights.**
  - Wally: *Middle two lights in the southeast cubicle on full.*
  - Wally: *Four westmost lights in the southeast cubicle off.*
  - Wally: *All east side lights off.*
- **Messages referring to the amount of light in the room.**
  - Dana: *hey ana it's dark.*
  - Lisa: *Could you brighten up my cube this morning?*
- **Messages referring to object in the room**
  - Dana: *can you brighten the lights in the aisle by the cabinets?*
  - Joe: *can i get the lights up a little just over the strip of tables in the meeting area?*
  - Jack: *Turn up the lights near the sink.*
  - Wally: *Could I get the row of lights next to the window on about 2/3 of the way?*
  - Wally: *Window lights off please!*
  - Lisa: *Could you turn on the lights overhead the couch/chair in the soft space?*
- **Messages that use names for spaces in the room.**
  - Jack: *can you turn up the lights in the tool shop - we'll be going back and forth*
  - Joe: *Can i get the lights on in the design research cubicle.*
  - Carol: *Turn lights on in the MechE cube.*
  - Dana: *could you turn on the lights in Alan's cube?*
  - Lisa: *Hey, could you turn up the lights over here in the soft space corner?*
- **All on or all off messages.**
  - Lisa: *could you turn off all the lights?*
  - Ron: *all off.*
  - Wally: *all lights off.*
  - Joe: *I'm the only one here. can i get all the lights off now. I'm leaving.*
- **Messages that make reference to the person speaking.**
  - Wally: *Could I please have the four lights overhead on to about half brightness?*
  - Dan: *lights over my head.*
  - Jack: *turn up the light nearest (and in front, so that i don't cast a shadow over my work) a bit.*
- **Messages with timing information.**
  - Joe: *Hi ana can i get my cubicle lights out in 5 minutes.*
  - Joe: *Can i get my lights on and the rest of the BID lights off please. in that order, so I'm not left in the dark. thanks!*
- **Messages that exclude lights.**
  - Wally: *All lights except the westmost four in the southeast cubicle off.*
  - Joe: *Can i get my lights on and the rest of the BID lights off please. in that order, so I'm not left in the dark. thanks!*

Fig. 3. Sample messages from Wizard-of-Oz study.

a corresponding term-frequency vector that represents the terms used in the message that resulted in the state change. We represent our data set in an $m \times n$ matrix where m is the number of state change vectors, and n is the number of lights (p) plus the length of the term-frequency vector (q). Each row of the matrix has a state change vector, followed by the term-frequency vector for the message that resulted in the state change. Not all state changes have a message associated with them because not all of the researchers in the lab participated in the study. Let $L = \{l_1, l_2, \ldots, l_p\}$ be the set of lights in the room and $W = \{w_1, w_2, \ldots, w_q\}$ be the complete vocabulary used in the messages after stop words are removed, and the stemming operation is performed. The vector X_i of state change c_i is defined as

$$X_i = [d_{1i}, d_{2i}, \ldots, d_{pi}, t_{1i}, t_{2i}, \ldots, t_{qi}]$$

where d_{ji} is the difference in intensity of light l_j after state change c_i, t_{ji} is the number of times vocabulary word w_j appeared in the message that triggered state change c_i. (Since each message is tagged with the user's name, the user's name is included in the vector X_i.) With the vector X_i as the i^{th} row, we construct the $n \times m$ matrix $\mathbf{X}$. We use non-negative matrix factorization with this matrix, and the light group weights and group terms are directly obtained from the results. See Fig. 4.

5.2 Computing a Non-Negative Matrix Factorization

Non-negative matrix factorization (NMF) is an algorithm that finds a positive factorization of the given matrix [LS00,LS99]. NMF requires we know the number of factors in the data, which we will call k. We explain shortly how we chose k for the experiment. We want to factorize the matrix $\mathbf{X}$ into the non-negative $m \times k$ matrix $\mathbf{U}$ and the non-negative $k \times n$ matrix $\mathbf{V}$ such that $\mathbf{X} \approx \mathbf{U}\mathbf{V}^T$. Each element u_{ij} of matrix $\mathbf{U}$ represents the degree to which the state change c_i is associated with factor j. Each element v_{ij} of matrix $\mathbf{V}$ represents the degree to which light l_j belongs to factor j if $0 < i < p$, and v_{ij} represents the degree to which term w_i belongs to factor j if $p < i < q$ where p is the number of lights and q is the number of terms in the complete vocabulary. See Figure 4. Our visualization of the results (on the far right side of the figure) has three parts. On the left, we show the degree to which each of the lights is part of the group with the size of the square over the light on the map of the lab. These values are scaled with respect to the strength of the group. In the middle, we show the strength of the group with a single black bar. The strength of two groups can be compared by comparing the middle black bar on the visualizations of two groups. On the very right hand side we plot each term along the y-axis based on the degree to which the term belongs to this group.

We used the NMF algorithm described in [XLG03] for document clustering. The algorithm minimizes the following objective function:

$$J = \frac{1}{2} \left\| \mathbf{X} - \mathbf{U}\mathbf{V}^T \right\|$$

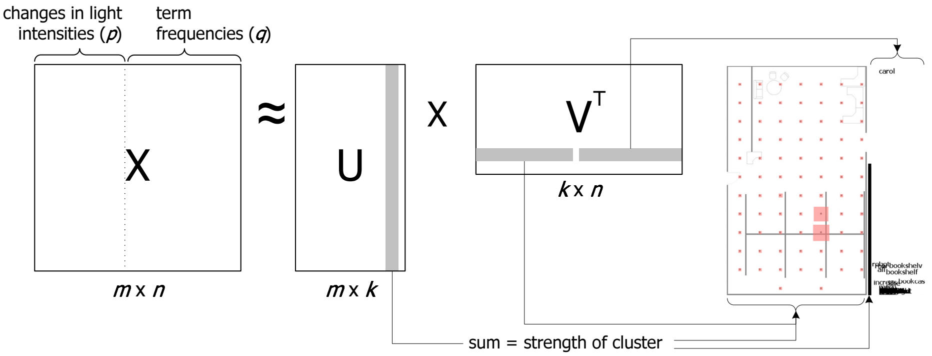

Fig. 4. We factorize matrix $\mathbf{X}$ into matrix $\mathbf{U}$ and matrix $\mathbf{V}$ such that $\mathbf{X} \approx \mathbf{U}\mathbf{V}^T$.

where $\|\cdot\|$ denotes the squared sum of all the elements in the matrix. The algorithm starts with random values in the matrices $\mathbf{U}$ and $\mathbf{V}$. In each iteration, it updates $\mathbf{U}$ and $\mathbf{V}$ based on the update rules below. It iterates until the result of the objective function J stops decreasing.

$$u_{ij} \leftarrow u_{ij} \frac{(\mathbf{X}\mathbf{V})_{ij}}{(\mathbf{U}\mathbf{B}^T\mathbf{U})_{ij}} \qquad v_{ij} \leftarrow v_{ij} \frac{(\mathbf{X}^T\mathbf{U})_{ij}}{(\mathbf{V}\mathbf{U}^T\mathbf{U})_{ij}}$$

Normalization The matrices $\mathbf{U}$ and $\mathbf{V}$ are not unique: multiplying the i^{th} column of U by s and the i^{th} column of V by $1/s$ produces the same product $\mathbf{U}\mathbf{V}^T$. To get a unique solution, we need to normalize one and adjust the other. In our case, we normalize matrix $\mathbf{V}$ by scaling column j so that the maximum value of the column $v_{1j}, v_{2j}, \ldots, v_{nj}$ is 255, i.e. we assume that some light in group j has "full" membership in the group, and the others full or less. Thus all elements of V are in the range $[0, 255]$. We then scale the corresponding columns of U so that the product $\mathbf{U}\mathbf{V}^T$ is preserved. The normalization has two roles: one is to make the factorization unique, the other is to allow us to measure the "strength" of a particular factor. The strength of factor i is the sum of the i^{th} column of U after normalization. This measures roughly the number of times that light group contributes to a state change, times the contribution of the factor to that change.

The normalization does not affect the visualization of of the vector V_j in Fig. 8 because the values in the vector V_j are scaled by the strength of the group. The subjective evaluation of each of the groups shown by the color of each of the bars in Fig. 7 shows the normalization we selected makes sense for our data since the unexpected groups (white in the figure) have the lowest strength measures. The rate at which the strengths of light groups decays in this plot is a measure of how "good" the factorization is. A good factorization should "explain" the data with the smallest number of factors. From the plot it can be seen that group strength drops almost to zero for 40 light groups. The first 20 groups "explain" more than 90 % of the data variation. This plot further shows that increasing the number of groups beyond 40 will have little or no effect on the factorization that results (it will produce only further almost-zero-strength factors).

Intuitively, we would expect 15-30 strong groups. There are 10 users most of whom set personal light configurations over their desks. There are about 5 clearly distinct areas in the public space, and then various ad-hoc areas that people use. The NMF factorization is right in the ball park with its own analysis of the strong light groups.

6 Results

We found light groups with matching terms both over participants' desks and in the public space. The groups over participants' desks are not necessarily labeled with the participant's name. If the participant used unique words to refer to his lights, those terms are the strongest (because users' names will also be present in their commands to public light groups). If participants tended to use more generic terms to refer to their light, the strongest associated terms include their name.

In Fig. 5 we show two of the groups in Wally's cubicle. He tends to refer to his lights by describing where they are in the room, and which lights he would like on within his cubicle. Wally tends to refer to his lights by describing their location in the room, and by describing which lights he would like within his cubicle. His cubicle is the southeast cubicle, and he likes to turn on the westmost lights in his cubicle. See Fig. 4 for a description of the visualization. On the very right hand side of the visualization, we provide a zoomed in version of the plot with the terms.

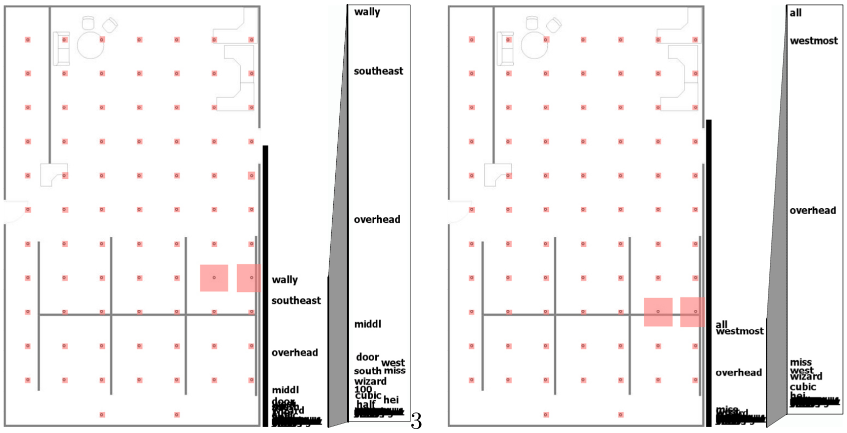

Fig. 5. Two light groups in user Wally's cubicle.

Figure 6 shows two examples of groups in the public space. The first is the row of lights next to the windows. The second is the set of lights over the fridge in the kitchen area. Multiple participants referred to these lights so the strongest associated terms are not participant's names, rather they are terms used to refer to the set of lights. The lights next to the window are on the east wall of the room, and are in a row. Notice the terms *window, row and east* are among the strongest terms in this group. The strongest terms in the group for the lights over the fridge include *area, kitchen, all, near, ron, dan, jack*. Ron, Dan and Jack often turn these lights on, and refer to them as the lights near the kitchen area, or some subset of those terms.

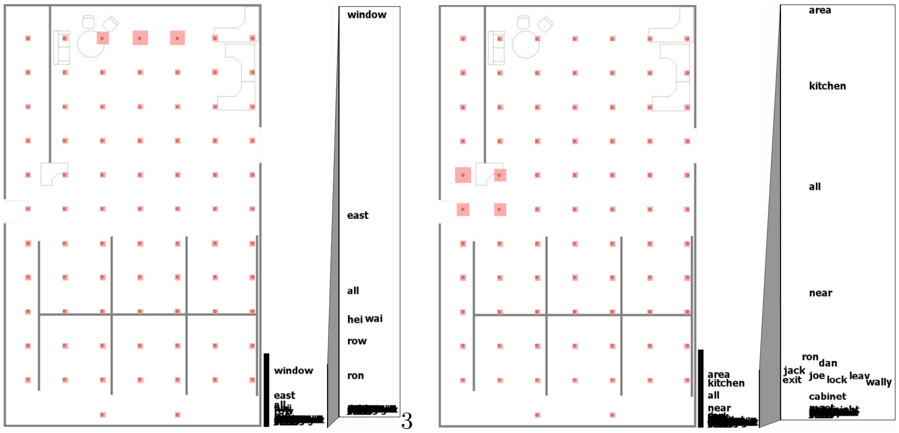

Fig. 6. Two examples of groups in the public space. The left visualization shows the group of lights commonly used next to the window once it gets dark outside. Notice among the strongest terms describe are descriptive words for the light group (*window, east, row*). The right visualization shows the group over the fridge. Here the descriptive words among the strongest terms include *area, kitchen, near*, and Ron, Dan and Jack all tend to turn on these lights.

In the next two sections we present the results from the NMF analysis. We also present results from the SVD analysis as a contrast to the NMF results.

6.1 Non-negative Matrix Factorization

For NMF, we selected $k = 40$ factors so that we would make sure to see all of the meaningful groups of lights. We also tried smaller values of k such as $k = 10$. This resulted in some factors with lights in just one cubicle, and other factors having lights in two cubicles, which we interpret as the selected value for k being too small. Figure 8 shows a visualization for each of the twelve strongest factors from the NMF analysis with $k = 40$. Each visualization has three parts: on the left, there is a small map of the room; in the middle there is a single black bar; on the right there is a collection of words. The map shows the degree to which each of the lights belongs to that group, and is scaled by the strength of that group.

The single black bar shows the strength of the group. The strongest group is the top left-most visualization. The collection of words on the right side shows all of the terms in the complete vocabulary. They are arranged along the x-axis based on the degree to which they belong to the group (scaled by the strength of the group). Figure 4 shows how the visualization is constructed from the matrices **U** and **V**. Figure 1 shows where each of the researchers sits in the room. Figure 7 shows the strengths of the forty groups found using NMF and our subjective evaluation of each of the groups.

To help the reader interpret the visualizations shown in Fig. 8 we describe each of the twelve visualizations shown. The first group shows Dana's desk, and although weak, the term dana is the strongest term. The second group shows Ashley's desk. Ashley was the wizard, so not only did her messages about her lights have her name on them, many of the other messages sent to her to change other lights also had her name in them, making it difficult to separate out her desk with her name. The third group shows a light over the bookshelf in Carol, Lisa and Joe's cube. Since each of them often manipulated that light, it is not labeled with any of their names. The fourth group shows the the lights around Joe's desk, and is labeled with Joe's name. He is the primary person to manipulate the light directly over his desk, but the other two are not as strongly part of this group since he is not the only one who manipulates them. The fifth group shows the two lights over Carol's desk, and is labeled with her name. The sixth group shows the two westmost lights in Wally's cubicle. He often refers to the lights in his cubicle based on directions. The seventh group shows the two lights over Wally's desk, which is in the southeast cubicle, and is labeled as such. The eighth group shows the lights Lisa likes to use. She refers to her cubicle as the "team design" cubicle. Only the term "team" came out in the NMF analysis for her lights. The ninth group shows the light over the sink in the toolshop that Jack often turns on. The three lights over Gina's desk are strongest in the tenth group, and the term Gina is the strongest term. The two lights over Dan's desk are strongest in the eleventh group, and again, the strongest term in the group matches. The twelfth group shows the lights Dana likes to turn on around her desk. She often turns on the lights over her desk and then decides she would also like the light on around her desk, including the two closest to her in Alan's cubicle.

6.2 Singular Value Decomposition

Figure 10 shows the strengths of the factors from the SVD analysis. Note that there is substantial strength in the 40th factor and beyond, indicating that many more factors are needed to explain the data as compared with NMF.

Figure 9 shows the strongest factors and corresponding terms from the SVD factorization. It can be shown that the first SVD factor is an average of all of the light intensities. Multiple groups of light from the NMF analysis show up in one factor from the SVD analysis. The second through fifth SVD factors in Fig 9 demonstrate this. The second SVD factor appears to include the nineteen strongest NMF group. The third SVD factor appears to include the strongest

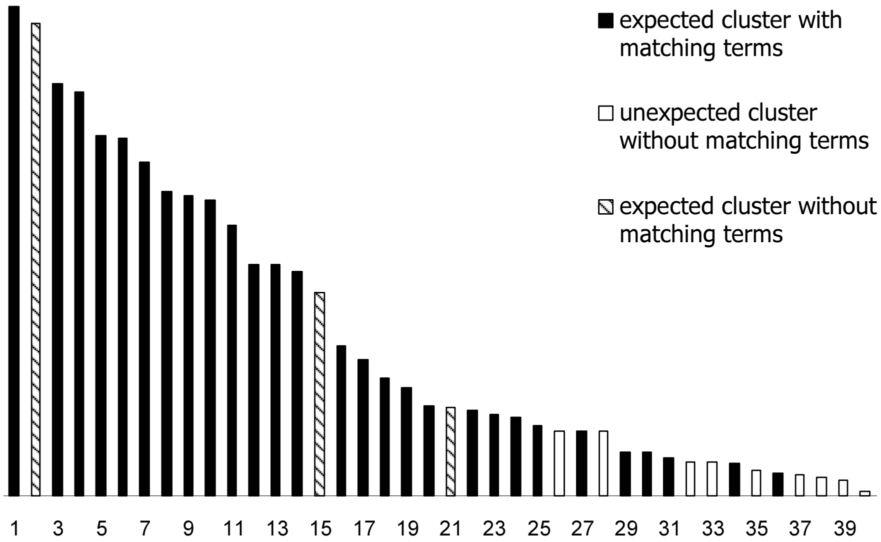

Fig. 7. The strengths and our subjective evaluation of the forty groups found using NMF.

ten NMF groups and the seventeenth NMF group (not show in the figure). These results suggest in order to effectively separate the light groups using SVD analysis, we need to apply a second clustering analysis, such as a K-means analysis. We leave this to future work.

7 Proposed Challenges to Address in the Thesis

In my thesis I plan to address four main challenges to the SNAC problem for lighting control. 1) I will iterate on the pattern recognizer for lighting patterns. 2) In addition to a pattern-recognizer for names sets of objects, I will also train a pattern recognizer for the *actions* performed on the named sets of objects. 3) Given a message, I will need to figure out how to use the pattern-recognizers to interpret the message as to what action should be taken. 4) I will build an instant message bot that will allow members of the lab to manipulate the state of the lights through natural language via instant messages.

The initial pattern-recognizer appears to do a good job of finding meaningful groups of lights for our lighting application. It also produces keywords which in many cases are good names for the groups. However, the names of many other groups are less effective (e.g. "space"). In most of these cases, ordered groups of terms (bigrams and trigrams) would provide much better names (e.g. "kitchen space, "public PCs"). I plan to extend the initial pattern-recognizer to compute factors from n-grams instead of only using unigrams. The resulting factors will contain n-grams of various orders which can act as fully-functional language

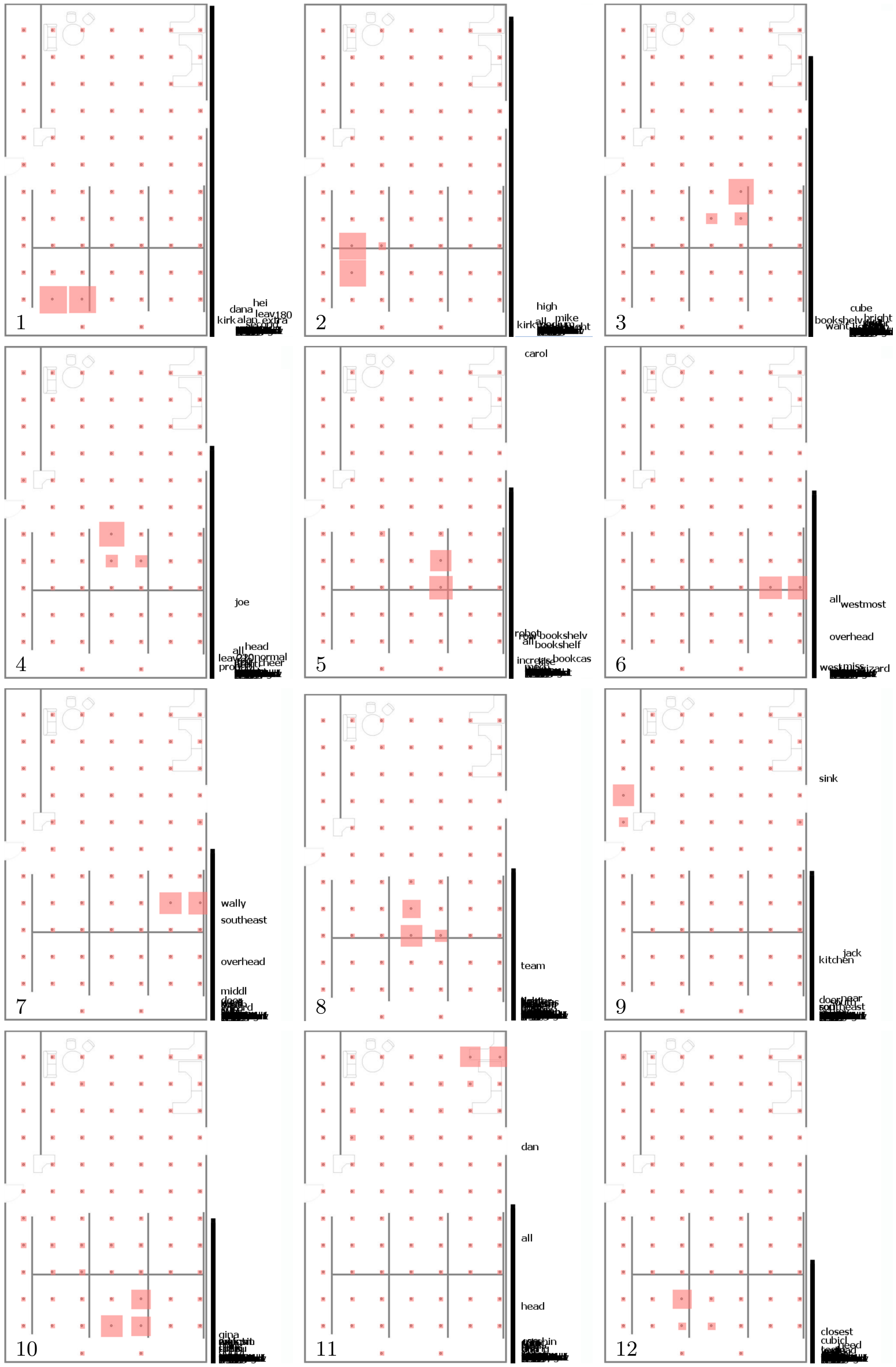

Fig. 8. The strongest twelve clusters from the NMF analysis. Figure 4 shows how to read this figure.

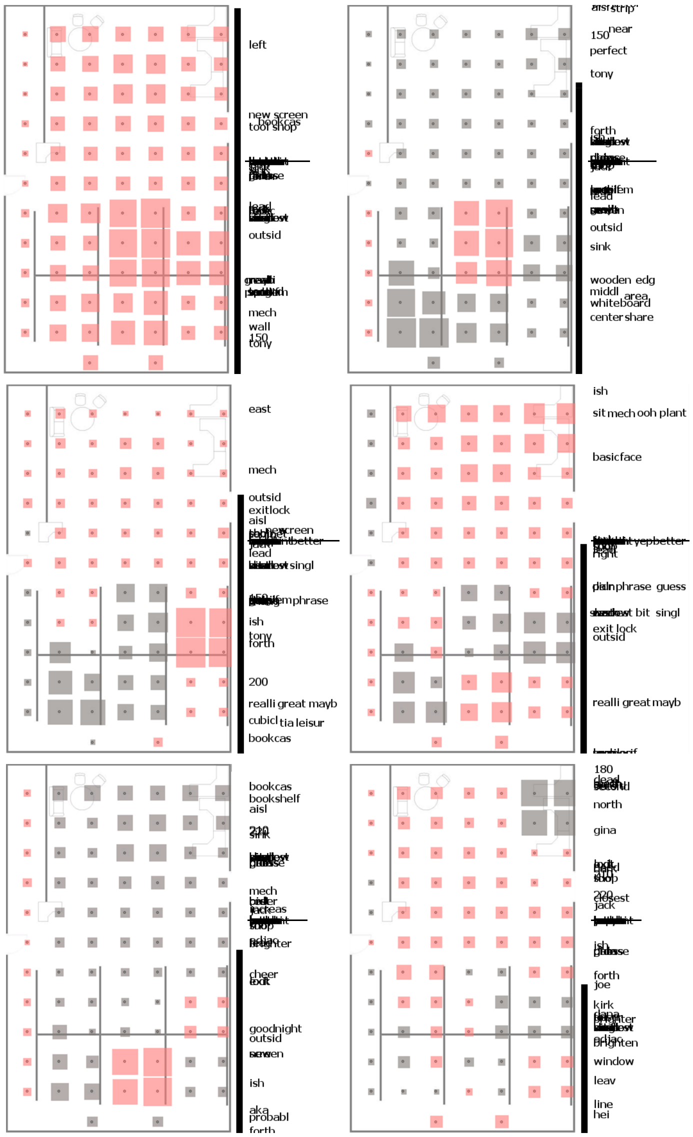

Fig. 9. The strongest six factors from the SVD analysis. The first one is the average of light intensities. Negative values are shown in red (the lighter color in gray scale).

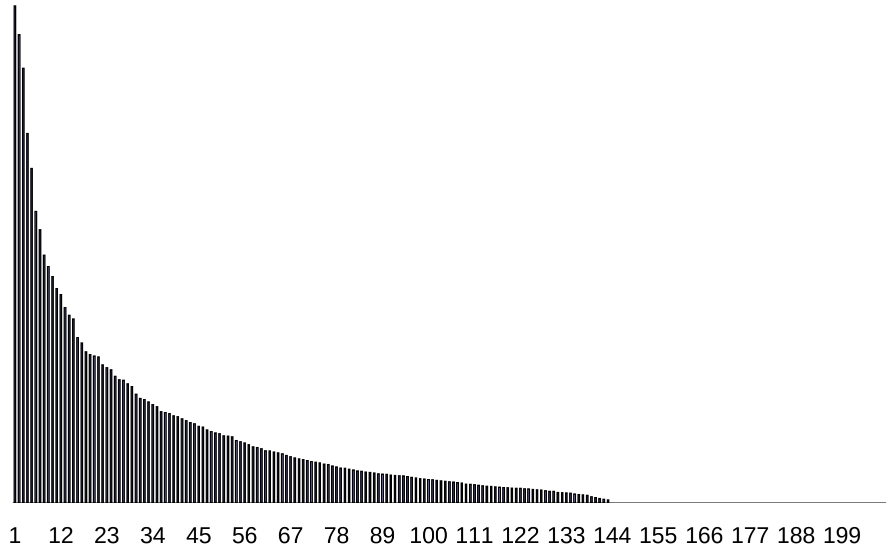

Fig. 10. The strengths of the factors from the SVD analysis.

models (n-grams with backoff and smoothing) for modeling the command utterances for each group.

In the design of the initial pattern-recognizer, we intentionally removed signs from the change data, arguing we would look at actions separately. I plan to train a pattern-recognizer on the kind of changes in the light data and the actions in the messages. In the initial pattern-recognizer, we did not explicitly extract only group names and participant names, they came out in the factors naturally. I expect to find a similar effect with the actions and the kinds of changes to the lights.

In order to interpret the semantic content of the messages, I will need to make use of both of the pattern-recognizers. The state of the lights at the time of the message together with the action should help resolve any ambiguity in the group of lights referred to in the message.

Finally, after evaluating and iterating on the approach, I will implement an instant message bot to handle the lighting control in our lab. The bot will allow residents and visitors of the lab to control the lights through natural language in text messages. In addition to users using instant message clients on their computers in the cubicle area, at least one computer in the public space will have an instant message client running to allow users to control the lights from the public space. Users will also be able to control the lights from laptops in the public space via an instant message client.

8 Evaluation

The proposed method aims to enable a more natural interface to the lighting system discussed, and more broadly to complex home IT systems. I plan to evaluate the proposed method with a wizard-of-Oz study. Similar to the initial wizard-of-Oz study, the wizard will change the state of the lights based on instant text messages from members of the lab, but the wizard will use the trained pattern-recognizers to decide how to change the lighting scene. This way the proposed method can be evaluated *in situ*. A lighting scene change will be evaluated based on the reaction of the user. In the initial wizard-of-Oz study, users wrote clarification messages when the change to the lights was not what he/she expected. If a user responds to a change in the lights with a clarification message, we will record the change in the lighting scene as incorrect.

9 Conclusions

For my thesis I plan to propose and evaluate an approach to SNAC for the lighting control application in our lab. We conducted an initial wizard-of-Oz study of the usage of the lighting in a semi-structured workspace. In the wizard-of-Oz study, participants controlled the lights by asking the wizard to change the state of a particular set of lights via instant messages. The text command, along with the final configuration produced by the wizard is used to derive the name/configuration pairs.

I plan to address four main challenges in my thesis. 1) I will iterate on the initial NMF based pattern recognizer for lighting patterns. 2) In addition to a pattern-recognizer for names sets of objects, I will also train a pattern recognizer for the *actions* performed on the named sets of objects. 3) Given a message, I will need to figure out how to use the pattern-recognizers to interpret the message as to what action should be taken. 4) I will build an instant message bot that will allow members of the lab to manipulate the stat of the lights through natural language instant messages.

I will evaluate the approach before implementing the instant message bot with a wizard-of-Oz study similar to the initial study. Instead of having the wizard interpret the messages, the wizard will rely on the interpretation from the pattern-recognizers to decide what action to take. This will allow an *in situ* evaluation of the approach to SNAC for the lighting control application based on users' reactions and feedback.

Designing natural language human-machine interfaces present many challenges including (i) discovering how users give commands and (ii) discovering the configurations *named* by those commands. It is not only a problem of building a language model or grammar for user input, but in discovering the semantics of those inputs. In an effort to align users' *situated* understanding of the world with the system's and interpret their natural speech, I present an approach to the problem of simultaneous naming and configuration for natural language interfaces.

References

[Aus62] John L. Austin. *How to Do Things With Words.* Oxford University Press, 1962.

[BCDP+90] Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. A statistical approach to machine translation. *Comput. Linguist.*, 16(2):79–85, June 1990.

[BDdF+03] Kobus Barnard, Pinar Duygulu, Nando de Freitas, David Forsyth, David Blei, and Michael I. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.

[BNJ02] David Blei, Andrew Ng, and Michael Jordan. Latent Dirichlet Allocation. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems 14.* MIT Press, 2002.

[Boh07] Dan Bohus. Roomline: a spoken dialog system that provides assistance for conference room reservation and scheduling, as of Feb 28 2007. http://www.ravenclaw-olympus.org/roomline.html.

[Can04] John Canny. GAP: a factor model for discrete data. In *ACM Conf. on Research and Development in Information Retrieval (SIGIR)*, pages 122–129. ACM Press, 2004.

[Cla96] Herbert H. Clark. *Using Language.* Cambridge University Press, Cambridge, MA, 1996.

[DJ93] Nils Dahlbǎck and Arne Jőnsson. Wizard of oz studies – why and how. In *Proceedings of International Workshop on Intelligent User Interfaces*, pages 193–200, 1993.

[EP06] Nathan Eagle and Alex Pentland. Eigenbehaviors: Identifying structure in routine. In *Proc. Roy. Soc. A (in submission)*, 2006.

[FG02] Armin Fiedler and Malte Gabsdil. Supporting progressive refinement of wizard-of-oz experiments. In *Proceedings of the ITS 2002 - Workshop on Empirical Methods for Tutorial Dialogue Systems*, pages 62–69, 2002.

[JR04] Joshua Juster and Deb Roy. Elvis: situated speech and gesture understanding for a robotic chandelier. In Rajeev Sharma, Trevor Darrell, Mary P. Harper, Gianni Lazzari, and Matthew Turk, editors, *ICMI*, pages 90–96. ACM, 2004.

[Kir07] Kirusa. Formerly heyanita, a voice technology and services company, as of Feb 28 2007. http://www.kirusa.com/.

[LS99] Daniel D. Lee and Sebastian H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, October 1999.

[LS00] Daniel D. Lee and Sebastian H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, volume 13, pages 556–562, 2000.

[Moz05] Michael C. Mozer. Lessons from an adaptive house. In D. Cook and R. Das, editors, *Smart environments: Technologies, protocols, and applications*, pages 273–294. Wiley & Sons, 2005.

[Net07] TellMe Networks. Fundamentally improving how people and business use the phone, as of Feb 28 2007. http://www.tellme.com.

[Nua07] Nuance. Formerly bevocal, a leading provider of hosted application systems for customer self-service, as of Feb 28 2007. http://www.bevocal.com.

[QGS+01] Jose F. Quesada, Federico Garcia, Ester Sena, Jose A. Bernal, and Gabriel Amores. Dialogue management in a home machine environment: Linguistic components over an agent architecture. In *In SEPLN*, volume 27, pages 89–98, September 2001.

[RBH$^+$04] Manny Rayner, Pierrette Bouillon, Beth A. Hockey, Nikos Chatzichrisafis, and Marianne Starlander. Comparing rule-based and statistical approaches to speech understanding in a limited domain speech translation system. In *Proceedings of the 10th International Conference on Theoretical Methodological Issues in Machine Translation*, 2004.

[RHB04] M. Rayner, B. A. Hockey, and P. Bouillon. Building linguistically motivated speech recognisers with regulus. In *Tutorial presented at the 42nd Annual Meeting of the Association for Computational Linguistics*, Barcelona, Spain, 2004.

[Sea69] John Searle. *Speech Acts.* Cambridge University Press, 1969.

[Sea5c] John Searle. A taxonomy of illocutionary acts. In K. Gunderson, editor, *Minnesota studies in the philosophy of language*, pages 334–369. University of Minnesota Press, Minneapolis, 1975c.

[SWY04] Matthew Stuttle, Jason D. Williams, and Steve Young. A framework for dialogue data collection with a simulated asr channel. In *In Proceedings of Interspeech 2004*, October 2004.

[XLG03] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 267–273, New York, NY, USA, 2003. ACM Press.