Investigating Audience Satisfaction Based on the Nonverbal Language Analysis of the Audience in Photos

Yu-Yi Liao¹, Sheng-Fen Chien²

¹ National Cheng Kung University, Tainan, Taiwan
liao.yuyi@gmail.com
² National Cheng Kung University, Tainan, Taiwan
schien@mail.ncku.edu.tw

Abstract. When we want to know satisfaction of audiences in a speech, the questionnaire survey is the most affordable method to collect quantitative data. However, the reliability and validity of the questionnaire is closely related to the design of the questions and the attitude of the respondents. Photos can capture nonverbal behaviors of the audience while listening to a speech. At that time, the audience did not hide their emotions. In this paper, we organize the photos of the three speeches hosted by the NCKU Art Center in 2016 as the data set. We found that the results of the behavioral analysis in photos are similar to the statistical results of the questionnaire. Finally, we made some suggestions for using the photos to analyze the nonverbal behavior of audiences and for establishing a database of nonverbal behaviors.

Keywords: Speech, nonverbal language analysis, audience satisfaction.

1 Introduction

The arts are at the heart of a well-rounded education. The NCKU¹ Art Center organizes exhibitions, performances, lectures, educational programs and related events to motivate audiences to be in touch and develop understandings of art. Speech (or lecture) is an oral presentation in which a speaker addresses the audience. Speech involves interplays of not only recitation and body language, but also on conversation, delivery and feedback. The Center hosts free lectures on topics related to music, drama, movie, literature, exhibition, art, and so on. In order to understand the audiences’ view and to use it as a reference for improving activity planning in the future, questionnaire surveys are conducted after lectures. The overall response rate of paper surveys is usually around 30% to 40%. The questionnaires cannot be assured that participant’s replies are seriously answered. Therefore, we try to analyze the audiences’ nonverbal language in activity photos to investigate their feelings and satisfaction.

¹ NCKU stands for National Cheng Kung University in the city of Tainan, Taiwan.
Audiences do not hide their emotions while listening to speeches, and consequently their nonverbal language is relatively easy to read. Psychologist Albert Mehrabian [1, 2] provides a personal communication equation: “total like = 7% verbal liking + 38% vocal liking + 55% facial liking”. He points out “the non-verbal elements are particularly important for communicating feelings and attitude, especially when they are incongruent: if words and body language disagree, one tends to believe the body language” [1]. Many public speaking training websites indicate that paying attention to your audience’s nonverbal language can help you connect with and persuade them. However, it is a top-level skill for speakers to read their audience’s nonverbal language. All nonverbal behavior must be taken in the context to determine the true mood because the same nonverbal behavior may have different meanings. Therefore, we endeavor to understand the nonverbal language of audiences by analyzing the activity photos. Then, we compare the results of the analysis with questionnaire to explore the satisfaction of the audience.

2 Nonverbal Languages

MacLean [3] formulates the brain function as a “triune brain”, containing a primitive brain (reptilian complex), a limbic brain (paleo-mammalian complex), and a new cortex (Neo-mammalian Complex), to explain many major aspects of human behavior. The limbic brain is the “emotional brain”, that gives off a true response to danger for our survival. The limbic behaviors are honest and reliable behaviors revealing our thoughts, feelings, and intentions. The new cortex is a “thinking brain”, that analyzes critically the limbic reactions of those around us to decode what other people are thinking, feeling, or intending.

Navarro and Karlins [4] postulate that the new cortex is a “lying brain” because it is responsible for higher-order conscious activity such as language, abstract thought, imagination, and creativity. The limbic brain comes to revealing honest nonverbal behaviors that help us to read people. To ensure our survival, the brain’s response to distress or threats has taken three forms: freeze, flight, and fight. Freeze is the first defense of the limbic brain to ensure our survival and it can avoid detection by dangerous predator of in dangerous situations. Flight is the second limbic response. The goal of flight is to escape the threat or to distance oneself from dangers. Fight is the strategy of turning fear into rage in order to fight off attackers. The forms of fight are argument, puffing out your chest, or violating another’s personal space.

Our body expresses the comfort being experienced by our brain. When we feel discomfort, the limbic brain expresses nonverbal behavior that mirrors our negative state of being. Pacifying behaviors serve to calm us down after we experience something unpleasant or downright nasty. Pacifying behavior is an important signal to tell us that they are discomfort. In general, when the limbic brain is in a state of comfort, this mental and physiological well-being is reflected in nonverbal displays of contentment and high confidence [3, 4]. However, when the limbic brain is experiencing discomfort, the corresponding body language is characterized by behaviors emblematic of stress or low confidence. Navarro points out those universal
nonverbal behaviors constitute one group of body cues: those that are relatively the same for everyone \[4\].

There are many websites to train people to be better speakers. They all point out that audience’s body language is a vital clue that can help speakers connect with and persuade audiences. Speaker Hub [5], Club Solutions [6], the Genard Method [7], and Speak Like Pro [8] summarize some vital clues of audiences’ nonverbal behaviors. According to Navarro “When it comes to honesty, truthfulness decreases as we move from the feet to the head. The face is the one part of the body that most often is used to bluff and conceal true sentiments” \[4\]. To summarize, the audience’s nonverbal languages are articulated through 4 groups of human bodies: feet and legs, torso and arms, hands and fingers, head and face.

2.1 Feet and Legs

Morris [9] states that our feet are the most honest parts of the whole body as they have uncanny abilities to leak inner moods. Feet carried us to move, escape, and survive immediately, without the need for conscious thought. Navarro [4] observed that we tend to turn toward things we like or are agreeable to us. On the contrary, we stay away from things that we do not like or that are disagreeable to us. He also indicates that “Clasping of the knees and shifting of weight on the feet is an intention cue that the person wants to get up and leave. A sudden interlocking of the legs may suggest discomfort or insecurity.” \[4\].

Speaker Hub writes “when the audience is impatient, they will start getting very active with their feet and legs by circling or swaying their dangling foot, alternating their legs, or tapping one of their feet.” \[5\]. The website also points out if the audience disagrees, is disapproving or hostile, they try to escape by “pointing their feet towards an exit” \[5\]. Club Solutions website mentions “if audiences are frustrated, impatient or ready to go, they may suddenly cross their feet and move the dangling foot quickly up and down, jiggle the crossed-over knee, or tapping the foot.” \[6\].

2.2 Torso and Arms

Navarro states that “the torso houses many vital internal organs, so the brain will seek to diligently protect this area when threatened or challenged.” \[4\]. We demonstrate comfort by using our torsos and shoulders to lean in the direction of that which we favor. We usually use our arms to protect us, keep people away, and mark territory. For example, our arms will rise up to defend us. When we are happy and content, our arms move freely, even joyfully. When we are upset or fearful, we withdraw our arms. Arm-distancing behavior occurs not only when we encounter object we do not like, but also when we are around people we do not enjoy. “Arms akimbo is a powerful territorial display that can be used to establish dominance or to communicate that there are ‘issues’.” \[4\].

Speaker Hub website summarizes “if audience is open and interested, they will lean slightly forward or will be fairly upright and attentive.” \[5\]. The audience may turn away from speaker slightly, slump in their chair, or cross their arms over their body if
they are bored [5, 6]. If audiences disagree or hostile, their arms are folded defiantly coupled with legs crossed [5]. Club Solutions indicates that “Audiences are usually forward or bouncing in the seat or a rocking motion while they are excited and happy.” [6].

2.3 Hands and Fingers

Navarro considers “human hands can reflect very subtle nuances within the brain because everything your hands do that is consciously or subconsciously directed by your brain.” [4]. Interlacing fingers signifies feeling of low confidence or stress. Steepling hand (touching the spread fingertips of both hands) indicates confident of thoughts [4, 5]. Stroking fingers across the palm or rubbing hands together signifies pacify anxiety or nervousness [4].

“When an audience is considering what speaker’s said, they will clasp their hands together, stroke the chin, put their fingers or hands close to their mouth in a very relaxed fashion.” [5]. If the audiences are bored, they may whisper to each other, or use their hand to support their head [5, 8]. When audiences are uncomfortable, they may be massaging their neck, touching something in the area around their neck or face [4, 5, 8].

2.4 Head and Face

Ekman and his team developed the Facial Action Coding System (FACS) to measure human facial expression and found that the appearance of the face for each of the primary emotions is common to all peoples but facial expressions do vary across cultures [10]. Navarro stresses that “while our faces can be very honest in displaying how we feel, they do not always necessarily represent our true sentiments.” [4]. By assessing facial behaviors in context and comparing them to other nonverbal behaviors, we can use them to help reveal what the brain is processing, feeling, or intending. Comfortable or positive behaviors include the loosening of the furrowed lines on the forehead, relaxation of muscles around the mouth emergence of full lips, widening of the eye area as surrounding muscles relax, head tilt to the side, pupillary dilation, eyebrow arching, flashbulb eyes. Discomfort or negative behaviors include squinting, eye blocking with hand, a brief touch of the eyes, a delay in opening of the eyelids, the lids compress tightly, eye flash.

We must rely on other facial displays that accompany eye-gaze behavior to determine liking (a relaxed smile) or dislike (tightened jaws, compressed lips) [4, 5, 6, 10]. When audiences are confused, they may furrow their brows, rub their eyes or face, blink their eyes, tilt their head to the side, or lean to one side [6, 10]. If audiences are bored, they may put their head to the side or down, break eye contact, fix their eyes into space, or close their eyes for brief or even long periods [5, 6].
2.5 Emotion Expression

Nonverbal languages express emotions. Research on emotion perception and recognition has primarily emphasized on facial expressions (e.g., [10, 11, 12]). Wood et al. [13] reviewed literature of human facial perception and summarized that people are sensitive to faces, especially those with expressions, and can extract the emotional meaning of those faces instantly even when expressions are fleeting or subtle. Our ability to quickly perceive facial expressions of emotion are rooted in our perceptual expectations [14]. Jack, Garrod, and Schyns [14] postulate the perception expectation model and show that ‘facial expression signals follow a “biologically basic to socially specific” hierarchical signal evolution’. They identify four (as opposed to six [10]) basic emotion categories: happy, sad, fear/surprise, and disgust/anger [14]. These categories are later evolved into complex set of signals, for example some idiosyncratic facial gestures shaped by learning and culture [13]. Emotion expressions are multimodal [13, 15]. Schirmer and Adolphs [15] examine emotion perception by comparing facial with vocal and tactile processing; they observe human processing of signals of different modalities converges to enable holistic emotion judgments.

In summary, to perceive emotion expressions of the audiences in a lecture or performance, facial expressions, and gestures, as well as contextual information, are fundamental. Contextual information ranges from the immediate environment to the social status and cultural background.

3 Research Design

With the advances of AI, in particular image classifications using deep learning technologies, NCKU Art Center has been contemplating the possibility of developing a system to evaluate customer satisfaction for art events including gallery exhibitions, concert or opera hall performances, workshops and lectures. As a preliminary exploration, we focus on art lectures. The lecture hall settings (i.e., clearly defined space and conditions of lighting) and the form of performance (i.e., speech) make the observations of audiences much easier than other types of events.

NCKU Art Center events have been using questionnaire surveys to understand students’ attitudes in arts occasionally. Survey is a common method to evaluate thoughts, opinions, and feelings of a particular group of people in research of human subjects. The reliability of survey data may depend on the questionnaire and respondent’s manner. Questionnaire is one of the most affordable ways to gather quantitative data, but it is hard to design the questions which can capture emotional responses or the feelings of the respondents.

Every event has been documented through photography, and sometimes as well as video-taping. Audiences of each event have been informed about the photo and video recording that may be used for research purposes through posters at the entrance gates and event tickets. In 2018, the procedure to requests for data removal has been added to the posters. Since photos capture the moment of the audiences’ nonverbal language in the lecture, in this preliminary exploration, we try to discover the feelings of the
audiences by analyzing activity photos, and then to compare the results with questionnaire for validation.

We have identified three art lectures, with post-lecture surveys and photo records, in the years of 2016. All three lectures took place in the same auditorium. The auditorium has a total of 580 seats (Fig. 1). The number of participants in the three lectures was 414, 525, and 414 respectively. There are 174 seats marked with dash-line in the middle back of the auditorium. Usually the first two rows marked with dotted-line, a total of 22 seats, are set as the VIP seats. Our image analysis focuses on photos of these two areas. We analyze the activity photos manually with the observation techniques described in the previous section to identify the behavior of the audience. For the purpose of this paper, we have blurred all photos shown as figures for audience identity protection.

Fig. 1. The auditorium layout.
Areas enclosed by dash-line and dotted-line are focuses of data analysis.

4 Data Analysis

4.1 Photo Analysis

The first set of photos are from the Cai, Shi-Ping lecture on April 26, 2016. Mr. Cai is a famous writer; his talk was devoted to three families and love. This lecture started at 19:00. There were three types of activity: the speech, a question-and-answer session, and a book signing session. In total, the event was recorded with 128 photos, out of which, only 48 photos aimed toward the audience. These photos are divided into three parts: before-speech (15 photos), during-speech (20 photos), and after-speech (13 photos).

People demonstrate comfort or love by leaning body toward the direction of favor [4-6]. When the audience likes the speaker, they will choose the seat closest to the speaker. Fig. 2 was taken before the speech began, and it reveals that most audiences
like this speaker because they choose to sit in the front rows and feel the comfortable to do their own things.

Fig. 2. Before speech, many audients, in the rectangle enclosed area, chosen to sit in front rows.

On the other hand, when the audience do not like the speaker or the topic of speech, they may choose seats that are far from the podium [4]. Fig. 3 shows the nonverbal language of audiences who sat at the back of the auditorium during the speech. We can see that many audiences bent down their heads and did their own things except the audiences marked by rectangles. Audiences, marked by dark-line rectangles, kept the direction of their head on the speaker and their body was still upright as they were feeling neutral to speaker. Some audiences, marked by light-line rectangles, put their fingers close to their mouth, but in a very relaxed fashion.

Fig. 3. During the speech, many audiences in back rows bent down their heads and did their own things except the audiences marked by rectangles.

TED curator Chris Anderson emphasizes “18-minutes is long enough to be serious and short enough to hold people’s attention.” [16]. King and Behnke [17] found that listener suffer from “cognitive backlog” which means listening is an exhausting activity because the learner is continually adding material to be remembered—retrieved—later. They considered that a 60-minute presentation produces so much backlog that the speaker risks seriously upsetting the audiences unless the speaker creates a very engaging presentation with “soft breaks”—stories, videos, demonstrations, or other speakers [17].

Photos in were taken about an hour into the speech. In Fig. 4-left, most audiences who set in front rows were still attentive. Some of them, marked with a rectangle in
right, were taking notes. Some of them, marked with light-line rectangle, may be thinking about the speaker’s words as they stroked the chin, clasped hand together, or put their fingers close to their mouth with. Some of them marked with shaded circles, still paid attention to the speaker as they still kept their body upright and attentive or lean forward. In Fig. 4-right, two audiences, marked with bright-line rectangle, were bored as they bowed their heads to do their own things. Some audience, marked with circles, stay focused as they tilted head, leaned forward and had a serious face. When audiences, marked by light-line rectangle, were thinking as they closed finger resting on the chin.

**Fig. 4.** About an hour into the speech, many audiences in front rows are still paying attentions.

Question-and-answer (Q&A) session allows the audience to ask various questions pertaining to the topic of the lecture. **Fig. 5** was taken at Q&A session; all audiences were doing their own things except three people. Although the audience marked by pink rectangle looked at questioner, she may want to leave because she placed a purse on her lap and crossed her arms over her body. Two audiences marked by blue rectangle may be feeling neutral. One of them leaned forward and supported the head with hands but in relaxed fashion.

**Fig. 5.** In Q&A session, three audiences, marked with rectangle, turned their face to the questioner while others were doing their own things.

Book signing provides readers with a chance to see and meet a favorite author and to obtain the author’s signature that can increase the value of books. In addition, it also allows authors to connect with their fans. Navarro [4] considers that cheek or face touching is a way to pacify when nervous, irritated, or concerned. **Fig. 6** show that
famous writer Mr. Cai, Shi-Ping (left) enjoys signing book for his fan. The reader (right) who touches his cheek may be a little nervous.

Fig. 6. An interaction of the speaker and an audience during the book-signing session.

The second set of photos are from the Wang, Hao-Yi lecture on May 6, 2016. Mr. Wang is a famous writer of Tainan’s literature and history. The lecture topic is about Tainan’s tourism and food. This lecture started at 19:00. In total, the event was recorded with 127 photos, out of which, 57 photos aimed toward the audience. These photos are divided into three parts: before-speech (34 photos), during-speech (16 photos), and after-speech (7 photos).

Navarro [4] stated that establishing a person’s baseline behavior (call starting position) is critical because it allows you to determine when he or she deviates from it. Therefore, we set the nonverbal behaviors of the audiences at the opening as their baseline behavior. Fig. 7 show behavior changes of two groups of audiences. The first group has three people (Fig. 7-left marked with light-line rectangle) where one man put hand close to cheek. The second group has two people (Fig. 7-left marked with dark-line rectangle); they were whispering. A second photo taken 37 minutes later (Fig. 7-right) showed the second group was still whispering while the first group was still listening to the speech. Fig. 7-left was recorded 39 minutes into the speech session, two people, marked by bright-line rectangle, were bored to take a nap; and next to them, another person marked with light-line rectangle was thinking. Two people marked with dark-line rectangle were still interested; one was taking a note, and another was bright smile coupled with eye contact. At 44-minute into the speech, Fig. 8 reveals that some audiences, marked with bright-line rectangle, were tired, some marked with light-line rectangle were thinking, and some marked with circles were still listening.

Fig. 7. An example of audience behavior changes during the speech.
Fig. 8. Audience behaviors at the 44th minute of the speech.

The third set of photos are from the Neal Wu lecture on June 26, 2016. Mr. Wu is a director and famous writer. His talk was devoted to his film “At Cafe 6”. This lecture started at 19:00. In total, the event was recorded with 99 photos, out of which, 83 photos aimed toward the audience. These photos are divided into three parts: before-speech (24 photos), during-speech (37 photos), and after-speech (22 photos).

The speaker likes to get close the audience and interact with them. Fig. 9-left was recorded shows at 62-minute into the speech, the audience, marked in circles, was still eager to interact with speaker. In the Q&A session, the speaker took the microphone to the questioner is the most intimate interaction. Fig. 9-right displays some audience, marked with light-line rectangles, were thinking, some, marked with circles, were feeling comfortable, or the person marked with dark-line rectangle was doing own thing when speaker handed the microphone to the questioner.

74 minutes into the speech, Fig. 10-left shows that most audiences were still listening carefully and some, marked with light-line rectangle, were thinking about what the speaker were saying. Fig. 10-right demonstrates that most audience were smiling when they take a group photo with speaker although we can still spot some audiences’ facial expression, marked by dark-line rectangles, were neutral.
This was a long lecture but audiences stayed interested and were excited to take photos with the speaker at the end of the lecture.

4.2 Questionnaires

In this section, we will discuss the results of questionnaire survey and nonverbal language analysis. The number of questionnaire responses for the three lectures was 170, 183, and 140 respectively. The response rate was about 41%, 35%, and 34% respectively (Table 1). The questionnaires consisted of three sections, the first of which was intended to demographic questions. The second section asked respondents favor about lecture topic and promotion method. The last section was designed to know the rate of their satisfaction. Our questionnaire analysis focuses on the last section, where three questions were related to the speaker and the content of the speech.

Q1: The speaker’s popularity is attractive to me
Q2: The content of the speech appeals to me
Q3: The interactions and interests in the lecture are good

<table>
<thead>
<tr>
<th>Lecture</th>
<th>N/A</th>
<th>0*</th>
<th>1*</th>
<th>2*</th>
<th>3*</th>
<th>4*</th>
<th>respondents</th>
<th>audiences</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Q1</td>
<td>2</td>
<td>1</td>
<td>.006</td>
<td>3</td>
<td>.00</td>
<td>19</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>Q2</td>
<td>2</td>
<td>0</td>
<td>.000</td>
<td>0</td>
<td>.00</td>
<td>12</td>
<td>.07</td>
</tr>
<tr>
<td></td>
<td>Q3</td>
<td>2</td>
<td>0</td>
<td>.000</td>
<td>0</td>
<td>.00</td>
<td>23</td>
<td>.14</td>
</tr>
<tr>
<td>#2</td>
<td>Q1</td>
<td>2</td>
<td>1</td>
<td>.005</td>
<td>0</td>
<td>.00</td>
<td>27</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td>Q2</td>
<td>2</td>
<td>0</td>
<td>.000</td>
<td>0</td>
<td>.00</td>
<td>7</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>Q3</td>
<td>6</td>
<td>0</td>
<td>.000</td>
<td>0</td>
<td>.00</td>
<td>21</td>
<td>.11</td>
</tr>
<tr>
<td>#3</td>
<td>Q1</td>
<td>1</td>
<td>1</td>
<td>.007</td>
<td>0</td>
<td>.00</td>
<td>5</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td>Q2</td>
<td>1</td>
<td>1</td>
<td>.007</td>
<td>0</td>
<td>.00</td>
<td>16</td>
<td>.11</td>
</tr>
<tr>
<td></td>
<td>Q3</td>
<td>1</td>
<td>1</td>
<td>.007</td>
<td>0</td>
<td>.00</td>
<td>10</td>
<td>.07</td>
</tr>
</tbody>
</table>

*: 0-strongly disagree; 1-disagree; 2-neutral; 4-agree; 5-strongly agree.
4.3 Comparison

In the first lecture, 85.3% of the respondents liked the speaker. 92% and 85% of the respondents agreed that the content and interaction of the speech were attractive. Only 48 shots of this speech picture were taken to the audience and most of the shots were concentrated in the front row of the audience. However, the comparison between the results of the behavioral analysis in the photos and the statistical results of the above questions are mostly consistent. More than half of the audience chose to sit in the front seat of the auditorium, representing their subconsciously desire to approach the speaker. The audiences are satisfied with the content and interaction of this speech, because they can overcome the “cognitive backlog” and continue to pay attention for more than an hour.

In the second lecture, 84% of the respondents liked the speaker. Up to 95% and 85% of the respondents liked the content and interaction of the speech. Among the three lectures, the content of the second one was most appreciated by the responded audience. In the last lecture, up to 95% of the respondents attended the speech due to the speaker’s popularity. He is the most popular speaker of the three lectures. 87% and 89% of the respondents agreed that the content, interaction and interest in the speech were sufficient. The interaction and interest of the third speech is the most adequate.

5 Discussions

The most realistic nonverbal behaviors of the audiences when they are listening in the speech were recorded by the photo. Analyzing the audience’s nonverbal behaviors in the photo provides another way to understand audience satisfaction beyond questionnaire surveys. In this paper, we use manual methods to analyze the photos. There are following difficulties:

1. The details of the body posture are too similar and it is not easy to analyze. For example, the audience uses his hand to support his head, which may represent thinking or being bored. If this posture is very relaxed, it means that he is thinking. If this posture makes the cheeks wrinkled, then he is bored.

2. Different people analysis may have different results.

3. There are too many audiences in the photo; human eyes are so tired that the analysis is incorrect.

Therefore, if we control the following essential factors, the computer can help us improve the accuracy of the analysis.

1. To control situation and subject of photo shooting
   a. The light in the lecture hall is as consistent and bright as possible.
   b. The shot audiences are as consistent as possible in order to analyze changes in their body posture.
   c. The audience’s body position in the photo is as clear as possible.
   d. The body behaviors of the audience sitting in the back row of the lecture hall are the key to analyzing audience satisfaction.

2. To establish nonverbal behavior database
(a) To establish a photo database of behaviors based on feet and legs, torso, arms, hands, figures, and face, respectively.
(b) Each behavioral database collects behavior from different angles as much as possible.
(c) In order to analyze facial expressions, the facial database is subdivided into eyebrows, eyes, nose, mouth, chin, cheeks, and ears.

At the time of writing, we have started the implementation process of using the manually annotated data to train our system for an initial version of image classification database. We envisage potential applications of our image classification database, beyond evaluations of audience satisfaction, are behaviors of occupants in interior environment.

Acknowledgements. This project is partially supported by NCKU Aim for the Top University Project 2016-2018.

References
7. The Genard Method: How to read body language – 5 ways your audience can make you a better speaker, URL https://www.genardmethod.com/blog/how-to-read-body-language-5-ways-your-audience-can-make-you-a-better-speaker


