

Multimodal Analysis of Online Webinars Conducted in Zoom

Pasala Krishna Chaitanya

ITMO University, Russia,  <https://orcid.org/0000-0002-1982-8858>

Sofia Sorokina

ITMO University, Russia,  <https://orcid.org/0000-0001-9159-5203>

Oleg Basov

ITMO University, Russia,  <https://orcid.org/0000-0001-5788-4845>

Abstract: Student satisfaction is one of the important factors to be considered in the education system. Due to the COVID19 pandemic and to make quality education available for a wide range of people, the current higher education system has undergone several changes. The classes and webinars being conducted through video conferencing systems such as zoom have gained unprecedented popularity. So, according to statistics, more than 97% of students switched to online education in the United States as it is a safe and feasible option to continue their education during the pandemic situation. At the same time, only 31% of students consider this format to be of the same quality as full-time, and only 5% of students find online learning to be of higher quality than offline. That is, 63% of students receiving online education are not satisfied with its current quality. According to the many types of research on student-teacher communication, non-verbal communication plays a crucial role in student satisfaction. Therefore, the study is conducted to find the direct correlation between time characteristics such as loudness, silence duration, change in intonations, tempo, and hand gestures of instructors during online classes and student satisfaction levels. As part of the study, recent online classes and webinars from MIT, Harvard, Stanford, IIT's, ITMO, and few prestigious universities are analyzed. The quality of the classes is rated by the students as high or low based on their satisfaction level. Based on the analysis, successful and unsuccessful models are compiled. Based on the obtained models and the time parameters illustrated, assessment and recommendation systems are offered to improve the quality of both individual characteristics and the general concept of classes

Keywords: Non-verbal communication, Zoom, Online classes, Webinars, Correlations

Introduction

The COVID-19 pandemic forced most of the universities and schools worldwide to remain closed for students

and working staff safety. So, to continue with the curriculum and to provide uninterrupted education, most of the universities have switched to synchronous E-learning mode. Development of the technology has provided different options for distance education are quickly adopted to promote additional values in the educational area (Malinovski T., 2018). Synchronous E-learning is a method of education where teacher and students are separated in space by utilizing software-based services to bridge the gap between them (Singh S., 2020). It is the all-inclusive blending of ICT gadgets and modern telecommunication equipment into the education system (Eze S., 2018). The E-learning platform is a softwarebased module that integrates diverse management devices, communication, evaluation, checking, etc. with an objective that could provide technological help to teachers as well as the targeted students to improve the variety of phases of the teaching-learning process, also the educational procedure in its nature being distant, classroom-based or having a combined nature where they merge the two modes in various extents (Malinovski T., 2018).

Synchronous E-learning can happen with the help of video conferencing platforms. Videoconferencing-based educational systems can enable students in one or many locations, to watch and listen to an instructor in real-time, and maintain interactivity with those remote locations. It is a mature technology that can be used in the distance learning process to provide instructor-student or student-student interaction (D. Clawson, 2016). For most of the educational systems during the pandemic, It is the best way to deliver a better experience to the students. Most of the video conferencing applications are user-friendly and easy to operate while establishing the connection between users. Virtual classrooms can be created using video conferencing systems.

The advancement of digital technology and Internet services made people connect through video conferencing in the modern world. It has increased the flexibility of geographical locations and also increased cost and time efficiency. But still, users face few constraints due to the bandwidth limitations, high development, and maintenance costs, and lack of skills to use (H. Kimiloglu, 2017). Among the popular video conferencing applications Zoom video communications is considered to be more helpful in the teaching-learning process. It solves most of the connectivity and user issues by its unique architecture and userfriendly interface. Even in the low bandwidth, zoom provides hassles communication to its users (Z. V. C. Inc, 2020).

In the education industry, the way instructors communicate with the students plays a crucial role in student satisfaction and academic performance. According to the research conducted it is assumed that 65% of the meaning of conversational messages comes from non-verbal communication and only 35% through verbal communication (Ahmad, 2009). Therefore, verbal and nonverbal communication is an aspect that needs to be emphasized for every teacher. The teacher's ability in applying both can help improve the quality of learning in the class. It has an impact on the academic and non-academic performance of students. Instead, if the teachers are not able to communicate effectively in transferring knowledge, skills, and attitudes, it will be hard to achieve the expected objectives. This shows that students who understand and can do the tasks very well also depend on how the teachers' style of verbal and nonverbal communication. Verbal communication includes a short talk, discussion, ideas, and so forth (Wahyuni, 2018). Nonverbal communication skills include body language, the use of eye contact, hand gestures, tone, change in intonations, and speed of speech which cues for creating

emotional connections and understanding the meaning of the words (Fan, 2021), (Scott-Parker, 2017). Then the two communication skills must always be honed by teachers, to create quality learning.

Non-Verbal Communication

Nonverbal communication is the process of sharing meaning interpretive and multi-media communication type. It is natural, reliable, and spontaneous (Wahyuni, 2018). In simple it is a process of conveying messages without using written or spoken words. Nonverbal communications help to maintain learner's focus throughout the class via a continuum of instruction to achieve an equilibrium between communication and rapport (Scott-Parker, 2017). Using nonverbal signals, the following information can be encoded such as physical attributes of the speaker, nonarbitrary referential meaning to objects and environmental states (S. Fruhholz, 2020). The sender of non-verbal communication may be unaware that he or she is sending a non-verbal message. Consequently, he or she may not be aware of the impact that the non-verbal message may have. Since the non-verbal message may be transmitted unconsciously, the sender will not have planned it. Therefore, a non-verbal message can be more reliable than an oral or written message. The types of nonverbal communication associated with student satisfaction are eye contact, paralinguistics, gestures, physical appearance, and proxemics (Fan, 2021).

According to (Fan, 2021), the instructor physical appearances are not associated with student satisfaction levels. As the study deals with digital education eye contact and proxemics are not considered. So, the relation between gestures and paralinguistics is considered in the study to find their association with student satisfaction during online webinars and classes.

Gestures

Gestures are movements and signals using body parts used to communicate messages with or without the conjunction of words. Gestures include movements of hands, face, head nods, and other body parts. Gestures refer to the interlocutor rather than to the topic of conversation, and they help maintain the conversation as a social system (J. Bavelas, 1992). Gestures are made predominantly when the person is talking, and they are temporally synchronized with the speaker's verbal syntax occurring at precisely the same moment as the relevant part of speech (S. Kita, 2017). Gestures are not treated as physical movements but rather as referential acts; they convey meaning, depict events, and represent ideas. They specify and often clarify verbal references, and they can denote meanings that may not be in the accompanying words. Gestures work together to convey semantic information in a highly integrated manner. Gestures are particularly suited for solving the speaker's problem of involving the listener without disrupting the topical verbal narrative. Sometimes, interactive gestures are specialized to serve their function efficiently without requiring words (J. Bavelas, 1992).

Interactive gestures would be particularly affected by the loss of visual availability (i.e., fewer in the partition than in the face-to-face condition) because unwitnessed gestures could not serve the function of including and involving the other person in the conversational system. Individuals talk and gesture to convey information to

others, so the absence of another person physically would be expected to decrease the overall rate of gesturing. The rate of interactive gestures was higher for partners interacting face-to-face than for those who could not see each other (J. Bavelas, 1992).

Paralinguistics

Paralinguistics is part of vocal communication that refers to every element and nuance of the speech that does not involve the verbal message of the speech. Paralinguistics is defined as the discipline dealing with those phenomena that are modulated onto or embedded into the verbal message, be this in acoustics (vocal, non-verbal phenomena) or linguistics (connotations of single units or bunches of units) (B. Schuller, 2013). Besides the linguistic (verbal) information conveyed by speech, the paralinguistic (non-verbal) information, such as intentions, attitudes, and emotions expressed by the speaker, also convey important meanings in communication. Therefore, to realize smooth communication between humans and spoken dialogue systems (such as robots), it becomes important to consider both linguistic and para-linguistic information. Para-linguistic communication can be much more subtle than other forms of nonverbal communication.

Para-language, sometimes called vocalic refers to any cues created in the process of verbalizing other than the words themselves. This includes sniff and sneezes, pitch, rate of speech, nasality, pauses, and even silence. With spoken language, loudness, rate of speaking, tone, pitch variation, and uses of pauses can have a major influence on whether and how one can react to the words (Ruben, 2020).

Research reveals that when a listener or receiver attempts to judge where the speaker or source is on a contempt-affection continuum, they rely more on the tone of the voice than on the content of the message. Paralinguistic features of the human voice are used to identify a person's age, gender, emotions, and health condition (H. Perez-Espinosa, 2017).

Change in intonations: The term intonation refers to a means for conveying information in a speech that is independent of the words. Central to intonation is variation in speaking pitch, and intonation is often thought of as the use of pitch over the domain of the utterance (B. Aarts, 2016). Variation of pitch within speakers, whether within or across utterances, is known as intonation, and largely reflects the syntactic and semantic structure of the utterance. Intonational contours are more abstract than absolute pitches and are usually described as sequences of relative units of highs, and Lows. Languages vary in the intonational contours they use for particular types of utterances. Intonation can also, for instance, help to regulate turntaking in conversation, since there are intonational mechanisms speakers can use to indicate that they have had their say, or, conversely, that they are in full flow and do not want to be interrupted.

Tempo: The tempo is defined as the speed of speaking, that is speaking rate, which is best measured by the rate of syllable succession (Wood, 1973). The speech tempo of a person may vary based on contextual and emotional factors, and the tempo is also dependent on the language is spoken and the dialects of the language.

People may interpret speech tempo as an indicator of the importance of what is being said (Quene, 2007).

Loudness: Loudness is a psychological term used to describe the magnitude of an auditory sensation (Munson, 1933). It is evident that loudness is dependent not alone upon the intensity of the sound but also upon their physical composition. Loudness is the perceptual attribute of sound intensity and is that attribute of auditory sensation in terms of which sounds can be ordered on a scale from quiet to loud. The perception of loudness is a psychological attribute that depends on intensity, but also on several other parameters, like frequency selectivity, bandwidth, and duration of the sound (Gidla, 2020).

Experiment and Results

The experiment is conducted to find the direct correlation between time characteristics such as gestures, and paralinguistic features of the instructors during online classes and student satisfaction levels. Based on the obtained results, assessment and recommendation systems are offered to improve the quality of both individual characteristics and the general concept of classes.

Sample Data

As part of the experiment, recent online classes and webinars from MIT, Harvard, Stanford, Oxford, IIT's, ITMO, and few prestigious universities during the period 2019-2020 are collected. The study included 215 webinars. Out of the sample, 192 units are of English medium and 23 units are in Russian medium. The data is extracted from free sources such as YouTube and university websites available on the internet. From the obtained data, characteristics such as loudness, tempo, change in intonations, instructor visual presence, and duration of hand gestures performed are extracted and analyzed.

Data Extraction

From the collected video sample data, selected non-verbal features are extracted using modules build in python language. All the paralinguistic features are extracted using librosa and pyloudnorm libraries in python. All the frames from the video are gathered and Pre-trained modules are used to identify and extract information about gestures and host visual presence using opencv-python.

Student Satisfaction Level

The quality of the classes/ webinars is rated by the viewers as positive or negative. Final rating to the webinar is represented on the scale of 1 to 5. Based on the rating system, '1' implies all the participants are dissatisfied, whereas '5' implies all the participants are satisfied with the class or webinar. As the data is collected from the public domain, most of the students, who rated the video may be interested in the topic or subject. From Fig. 1,

the frequency of the rating is aligned right toward '5'. Therefore the distribution of rating is leftskewed. According to the collected sample data, the median is considered a central tendency due to its skewness. The median satisfaction level is 4.82.

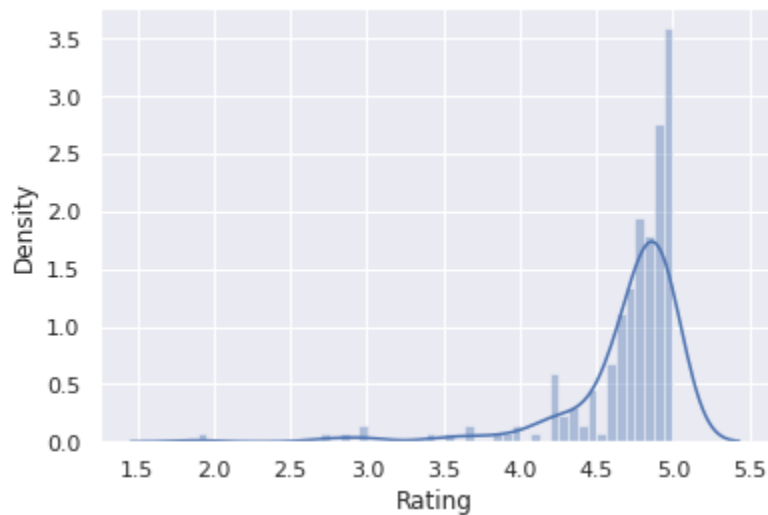


Figure 1. Frequency Distribution of Students Rating

Correlation and Dependence

Correlation is used to find the association between two variables and estimate the strength of their relationship. Correlation provides information about not only the strength but also the direction of a relationship. The most important fact is that correlation does not imply causation. The correlation between two variable is quantified with a number which varies between -1 to 1, negative values mean that variables are inversely related. The study is conducted to find the association between individual nonverbal characteristics and student satisfaction levels, using the collected sample data.

In the experiment, the Pearson correlation coefficient is used to determine the strength of association between the time characters tics and the student’s satisfaction level. Pearson correlation coefficient just efficiently characterizes the linear correlation between the attributes of the normal distribution. A Pearson coefficient also ranges from -1 to +1. It can be interpreted as describing anything between no association ($\rho = 0$) to a perfect monotonic relationship ($\rho = \pm 1$). Any relationship or association between two variables should be assessed not just for the strength and direction, but also by whether the relationship is “significant”, based on the p-value. The significance level of $\alpha = 0.05$ is chosen for the experiment to check the reliability of correlation analysis.

Gestures: According to the sample data collected, the correlation between gestures performed during online lectures and student satisfaction level is 0.036, which is weak. And the corresponding p-value for the sample is 0.6, which is greater than the significance level 0.05. Therefore the relation is not significant enough. So, it is assumed that there is no significant relation between gestures performed during online classes and student

satisfaction

Table 1. Pearson Correlation Matrix between Variables

Index	Gestures	Change in intonations	Silence	Tempo	Loudness	Visual presence	Rating
Gestures	1	-0.156	-0.148	-0.01	-0.083	0.75	0.04
Change in intonations	-0.156	1	0.75	0.186	-0.112	-0.02	0.15
Silence	-0.148	0.75	1	-0.08	-0.07	-0.03	0.1
Tempo	-0.01	0.186	-0.08	1	-0.09	0	-0.02
Loudness	-0.083	-0.112	-0.07	-0.09	1	-0.06	0
Visual presence	0.75	-0.02	-0.03	0	-0.06	1	0.05
Rating	0.04	0.15	0.1	-0.02	0	0.05	1

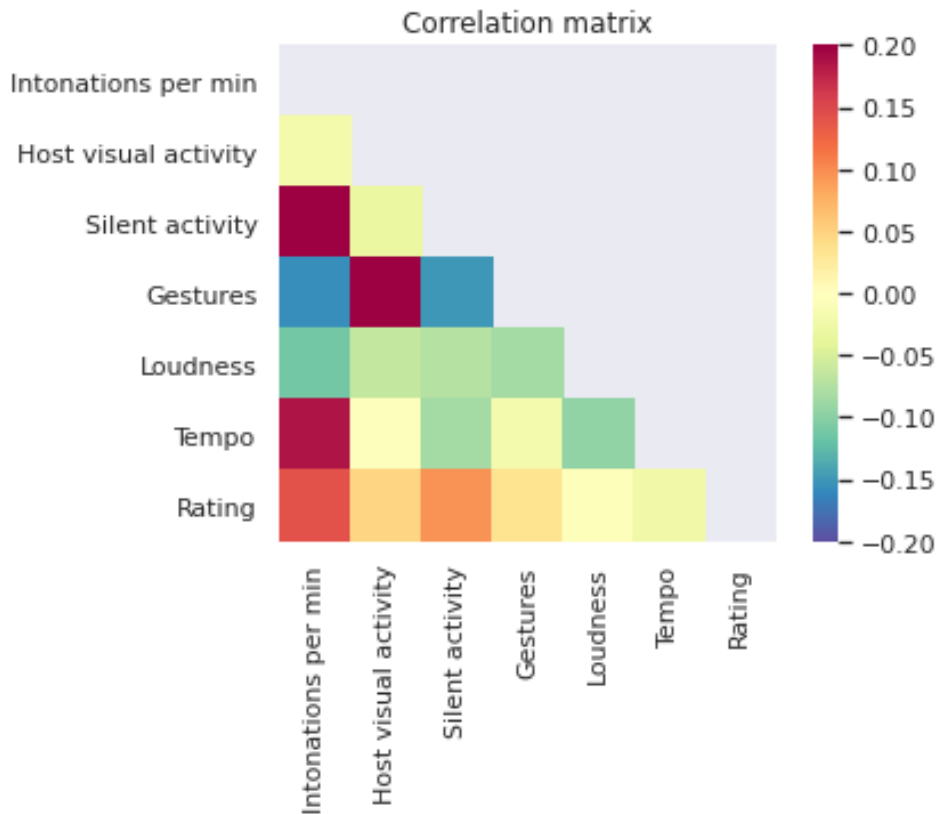


Figure 2. Heatmap of Correlations between the Variables

From the sample data it can be observed that the host is visually present in the online class or webinars about 30

percent of the duration only, as, during most of the lecture, the instructor used digital presentation in form of slides to interact with the audience. From the table I, the correlation between host visual presence and gestures performed is 0.75, which is strong. So, due to the loss of visual availability of the host, gestures are not showing much effect on students' satisfaction level.

Paralinguistics: Among the paralinguistic features, From the table II, it is clear that silence, tempo, and loudness are not significantly related to satisfaction level during online classes. But there is a significant relationship between change in intonations and students' satisfaction. As the Pearson coefficient between the variables is 0.14, and the p-value is 0.036, which is less than the significance level 0.05. So, it is assumed that there is a significant positive relationship between change in intonations and satisfaction.

Table 2. Reliability and Pearson Correlation of Paralinguistic Features

Paralinguistic feature	Correlation	p-value	Significance
Change in intonations	0.15	0.036	Significant
Silence	0.1	0.15	Not significant
Tempo	-0.02	0.75	Not significant
Loudness	0	0.94	Not significant

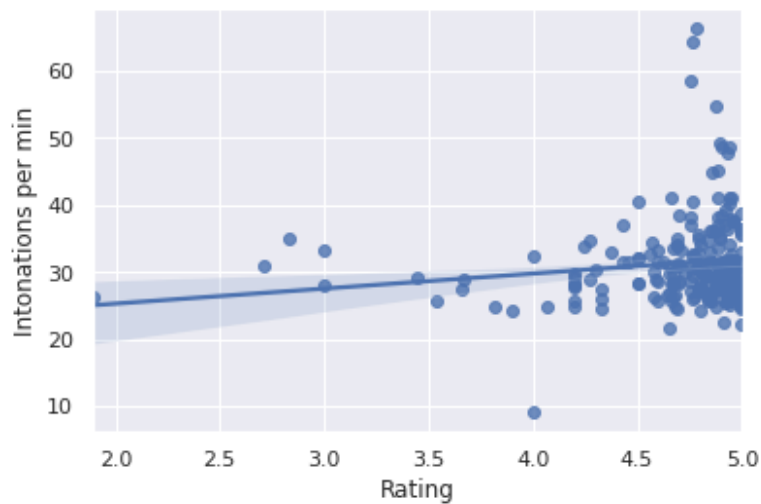


Figure 3. Regression between Change in Intonations and Students Rating

Limitations

One of the major limitations of the study conducted is that the majority of the data collected is from free sources like YouTube. So, the data may differ from the real-time online lectures conducted by the universities. And most of the people who participated in the lecture or webinar are interested in the topic or course. So, the ratings given by the participants may be biased. Due to the small sample size of 215 units, the results may not reflect

the whole population. There are few other factors such as internet connectivity, geographical locations, students' mood, instructors' behavior towards students which also affect the student's satisfaction level. But these all factors are neglected in this study case. In further research, these limitations will be avoided by collecting proper data from universities and by continuous student's feedback.

Conclusion

Based on the experiment conducted, it is assumed that there is a significant positive association between instructors' change in intonations in speech and students' satisfaction during online classes. This implies that any effort to enhance students' satisfaction should be preceded by improvement in paralinguistic features. This result is also supported by previous researches conducted based on nonverbal communications.

The study offers sufficient depth of insights into the specific areas that require attention and improvements during online classes. Despite many limitations and disadvantages, online education is growing every day. So, to provide proper education and satisfy students, educational departments should try to overcome these kinds of situations, the instructor should try to show more variations in pitch based on the emotion carried by the sentence and its meaning. Mostly, due to the unavailability of visual presence of the instructor during online classes or webinars, para-linguistic features should be prioritized.

References

- Ahmad, S. S. (2009). The relationship of non-verbal communication cues and student satisfaction.
- B. Aarts, A. M. (2016). *The handbook of English linguistics*.
- B. Schuller, S. S. (2013). Paralinguistics in speech and language—state-of-the-art and the challenge. *Computer Speech & Language*, 4-39.
- D. Clawson, M. K. (2016). Society for Information Technology & Teacher Education International Conference. *Cloud-based videoconferencing learning environments in higher education: The growing pains of synchronous learning in an asynchronous world*, (pp. 178-184).
- Eze S., C.-E. V. (2018). The utilisation of e-learning facilities in the educational delivery system of Nigeria: a study of M-University. *International Journal of Educational Technology in Higher Education*, 1-20.
- Fan, L. Q. (n.d.). Learning Triadic Belief Dynamics in Nonverbal Communication from Videos. 2021.
- Gidla, B. N. (2020). "Detection, thresholds of human echolocation in static situations for distance, pitch, loudness and sharpness. *Applied Acoustics*, 107214.
- H. Kimiloglu, M. O. (2017). Perceptions about and attitude toward the usage of e-learning in corporate training. *Computers in Human Behavior*.
- H. Perez-Espinosa, J. M.-M.-C.-J.-G. (2017). Using acoustic paralinguistic information to assess the interaction quality in speech-based systems for elderly users. *International Journal of Human-Computer Studies*, 1-13.

- Inc, Z. V. (2020). Zoom connection process.
- J. Bavelas, N. C. (1992). Interactive gestures. *Discourse Processes*, 469-489.
- Malinovski T., T. V.-S. (2018). Impact of Different Quality of Service Mechanisms on Students' Quality of Experience in Videoconferencing Learning Environment. *Turkish Online Journal of Distance Education* (19(3)), 24-37.
- Munson, H. F. (1933). "Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 377-430.
- Quene, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics*, 353–362.
- Ruben, L. S. (2020). Communication human behavior — higher education. *Kendall Hunt*.
- S. Fruhholz, S. R. (2020). Nonverbal auditory communication—evidence for integrated neural systems for voice signal. *Progress in Neurobiology*, 101948.
- S. Kita, M. W. (2017). How do gestures influence thinking and speaking? the gesture-for-conceptualization hypothesis. *Psychological review*, 245.
- Scott-Parker, B. (2017). Nonverbal communication during the learner lesson with a professional driving instructor: a novel investigation. *Transportation research part F: traffic psychology and behaviour*, 1-12.
- Singh S., H. S. (2020). Mechanising e-learning for equipping startup entrepreneurs. *Materials Today: Proceedings*.
- Wahyuni, A. (2018). 1st International Conference on Intellectuals' Global Responsibility (ICIGR 2017). *The power of verbal and nonverbal communication in learning*. Atlantis Press.
- Wood, S. (1973). Speech tempo (introduction). *Dept. of Linguistics, Lund University*.