English Language Teaching
Vol. 2, No. 2, pp. 1-21, 2015

Exploring the Role of Visual Representation Signals for Interactional Action in Conversation

Arash Gholamy Saleh Abady *
The Chairman of English Department, AJA University of Medical Sciences, Tehran, Iran
Sayyed Mohammad Alavi
Associate Professor, University of Tehran

Abstract

The main approach to conversation analysis is multimodal analysis, which can be explained by the distinction between the non-verbal and verbal expression in the communicative functions (Haddington & Kääntä, 2011; Streeck et al., 2011). The purpose of this study was to investigate whether there was a significant difference between non-verbal or verbal signals in conveying information in conversation. The participants of this study were 37 male Iranian B.S. Paramedic students at medical university for the Islamic Republic of Iran's Army. Two video talk show interviews were shown in order to determine the descriptive features for exchanging information. ELAN video annotation instrument was utilized for analyzing the interviews of this study. To find out which of verbal or non-verbal resources was effective in conveying information, a questionnaire was also developed by the researchers consisting of 19 items on the verbal and non-verbal signals. The results of ELAN analysis for both interviews showed that the descriptive visual cues such as hand movement, gaze, eyebrow motions, and torso were more frequent than the other non-verbal resources. Additionally, the analysis of the questionnaire data showed that there was a significant difference between the visual and verbal elements in the transmission of information from the students' viewpoints. Moreover, there was a significant difference between the non-verbal descriptive resources in conveying information. The findings of this study revealed that non-verbal cues were more effective in the transmission of information than the verbal cues. In addition, hand movements and laughing were found to be more effective than the other visual signals in conveying information.

Keywords: non-verbal signals; verbal signals; ELAN video annotation; interactional actions; multimodal signals.

---

* The Chairman of English Department, AJA University of Medical Sciences, Tehran
Received on: 03/11/2015 Accepted on: 20/02/2016
Email: *planetarash@gmail.com*

1. Introduction

The importance of conversation has now become crystal clear for people, and the main concern of individuals, particularly deaf or Hearing-Impairment (HI) people, is how to recognize the speech of another person completely. As Pajo (2013) points out that both interlocutors can watch together and utilize visual elements while talking. Thus, speech recognition is simpler for deaf individuals and if trouble occurs, the interlocutor can advert it.

A number of studies (e.g., Boersma & Weenink, 2007; Esfandiari & Ágnes, 2013; Foster & Oberlander, 2007; Kendon, 2004; Knight, 2009; Massaro, 1987; McNeil, 1992; Pastra & Wilks, 2004; Saferstein, 2004; Vilhjalmsson, 2009) have been performed in the field of multimodal analysis to find out the constituents of utterances. The results of these investigations revealed that there were two fundamental factors: (a) non-verbal or visual signals (b) verbal resources. Some researchers (e.g., Allwood, Cerrato, Jokinen, Navarretta & Paggio, 2008; Carrol & Bandura, 1982; Jokinen, 2009a; Jokinen & Vanhasalo, 2009; Kendon, 2004; McNeil, 2005; Pajo, 2013; Skelt, 2006) have attempted to show that non-verbal elements were more practical than verbal signals in conveying information through speech.

In addition, some researchers (e.g., Musgrave, 2012; O'Halloran, Smith, Tan & Podlasov, 2010) have examined the transmission of information of non-verbal elements on animals such as apes. Musgrave (2012) notes that the studies on non-verbal ape language were encountered with complex assessments. Similarly, O'Halloran et al. (2010) state that in multimodal text analysis a variety of techniques have been considered for such analysis as linguists argue that text analysis is related to the interrelation of meaning within texts or via interaction with language such as gaze, gesture or the visual art. The previous studies, however, paid less attention to the impression of visual representation cues in conveying information through speech about humans. Therefore, the lack of this issue is obvious, and more investigations are needed on the effect of non-verbal resources on the interactional action in humans` conversation. Given the previous studies, the purpose of this study was to determine which of the non-verbal or verbal signals was more significant in the transmission of information. Additionally, it aimed at understanding which of the non-verbal elements are more effective in the transmission of information.

2. Literature Review

2.1. Multimodal Analysis

To date, a number of researchers (e.g., Jewitt, 2006, 2009; Koutsombogera & Papageorgiou, 2009; O`Halloran, 2012) have carried out some experiments

about the impact of visual resources on the recognition of different sentences and even unknown words by means of video recording of various TV talk show interviews and even conversations. Koutsombogera and Papageorgiou argue that "the relations between distinct modalities in natural interaction have been thoroughly studied in order to deeply understand related to multimodal communication areas such as TV face-to-face talk show interviews occurring in various settings, human-avatar interaction etc." (p. 1). Multimodality in natural interaction can be explained by the distinction between the non-verbal and verbal expression in the communicative functions. The main approach to conversation analysis is multimodal analysis (Haddington & Kääntä, 2011; Streeck et al., 2011).

In line with visualization or in other words non-verbal signals, there are verbal signals. The analysis of both parts simultaneously leads to multimodal analysis. Some researchers (e.g., Boersma & Weenink, 2007; Esfandiari & Ágnes, 2013; Foster & Oberlander, 2007; Kendon, 2004; Knight, 2009; Massaro, 1987; McNeil, 1992; Pastra & Wilks, 2004; Saferstein, 2004; Vilhjalmsson, 2009) have concentrated more on the multimodal analysis. Multimodal analysis has two fundamental subcategories: (a) Visual or Non-Verbal signals (NV) (b) speech or verbal signals (V).

2.2 Verbal and Non-verbal Elements

Although non-verbal resources play a significant role in conveying information to the listeners, there is a supportive and complementary element for the visual cues in order to transmit the gap spaces such as emotional signs, emphasis parts, pitches, intonation, rhythm, and this complementary part for the visual signals is verbal or speech cues. O`Halloran and smith (2012) pointed out that there were two semiotic sources or modes of interaction; one was related to the features of speech (i.e., verbal expression) such as intonation and vocal characteristics, and the other was related to the semiotic action (i.e., non-vocal expression) of body resources such as gestures of hand, face and the body.

As Ba and Odobez (2011) argued, the interrelationship between non-verbal or visual cues and verbal or speaking signals was one of the most remarkable features of the visual advertence in conversation; therefore, the significance of using both speech signals and visual signals for interactional action in the mutual conversation should be emphasized simultaneously.

2.3 The Role of Non-verbal Signals

Performing different studies in the field of multimodal resources, some researchers (e.g., Aran & perez, 2011; Chen, 2011) have considered both parts of multimodality (i.e., verbal and non-verbal cues), but some other researchers in the same filed (e.g., Carrol & bandura, 1982; Jokinen, 2009a; Jokinen &

Vanhasalo, 2009) have considered more to one of the branches or resources of multimodal resources, and it was visual or non-verbal signals. They argued that visual or non-verbal cues were more significant than the verbal or speech cues. They further stated that by means of only non-verbal signals, speaker could convey his/her message(s) to the listener. Koutsombogera and Papageorgiou (2009) noted that "multiple functions of non-verbal (NV) communication lies in that it provided information and shed light on the interplay between verbal and non-verbal signals" (p. 3). In other words, Koutsombogera and Papageorgiou pointed out that numerous messages could be transferred by use of the visual elements which might complement intonation, accent, and substitute the verbal messages. Furthermore, non-verbal tools of communication grant us the conversational aspect of the speakers because they can be considered as powerful tools for the self-expression.

As briefly presented in this study, by means of video talk show programs, individuals have been able to transmit the various forms of information in conversation through pitches, intonation, rhythm, and pauses and through gaze, hand gesture, lip reading and lip movement, finger movement, and body posture. But the question is that, which kind of verbal or non-verbal resources is more significant in conveying information than the other within a TV face-to-face conversation? In order to extract the answer to this question, the researchers eliminate each of these indicated verbal or non-verbal signals by means of ELAN video annotation in order to understand which of them can transmit more information to the listener.

By means of ELAN annotator tool, there will be an amazing and clear relationship among various facial displays, body postures, body gestures, and video extracted non-verbal signals annotation (Figure 1). In other words, the researchers have implied to hand movement which has been considered as one of the branches of visual displays for communication. According to McNeil (1992), the communicative hand motions were divided into five subcategories, which were adaptors, metaphoric, iconic, beats and deictic. At the time of annotating hand movements, it is significant to realize what these motions represent. If the movement did not have any proposed communicative meaning such as speaker itching his/her nose and after that hand movement has been considered and annotated as an 'Adaptor'. Metaphoric relates to the gestures that are so similar to iconic which clarifies what is being stated. In a conversation, iconic can be perceived when the interlocutor has made some sort of hand motions, which shows what was being said by producing some type of a thing or even a form. For instance, when a speaker is explaining a rectangle, and he draws a rectangle shape in the air with his hands to clarify what the rectangle looked like. Beats refer to the motions that have had melodious and transmitted utterance, which have relationship to the sense of what was being said along with intonation. Deictic implies to the motions that

speakers utilize to point to the special location, direction, item, and person or to point to the abstract imaginary thing or a concept (Kipp, 2004). They refer to the signals that have an implication entirely by themselves and can be utilized in order to transmit the meaning when the speech is absent. These signals are known as thumbs up or, in other words, the OK sign.

3. Method

In present study, the non-verbal resources used in the process of producing the visual patterns in the various video talk show programs e.g., Late-show with David Letterman (2013) were employed.

In this part, the participants of this study, the instruments and materials, the procedure and data analysis were investigated.

3.1 Participants

The participants of this study were 37 Iranian B.S. students of Paramedic at the Medical University for the Islamic Republic of Iran's Army. They were all male students with the age range between 19 to 22 years old. They participated in a General English course which offered them three hours of instruction per week.

3.2 Instruments and Materials

3.2.1 Video Talk Shows

There were two video talk show conversations, which occurred between Larry king as an interviewer and other participants such as George Clooney and Jim Carry as an interviewee. Both interviewer and interviewees were native speakers of English. The interview was extracted from the website www.YouTube.com, which was accessible online. The interview was recorded using Internet Download Manager Version 6.14, www.internetdownload manager.com, which was a specific software program used for recording straight from the computer screen. The structure of the video talk show interview was in the form of an interviewer organized in the question-answer order. Video talk show interviews were selected because by means of ELAN video annotation each non-verbal signal could be compared easily with other visual data. One of the verbal or non-verbal resources was more effective in conveying information, and also it could be considered as a user friendly software program at the time of annotation process. ELAN recommended various and numerous functions such as synthesizing or merging two different annotation files into a single, importing or copying annotations. Synthesizing two annotation files in order to control their agreement along with the ‘Undo’ functionality if a wrong action was occurred. Additionally, there is a perfect

engine for the purpose of searching and designing statistics from not only a single file but also multiple files, as appeared in the Figure 1.

Figure 1. A Statistical Summary of Video Data Multimodal Annotation Provided by ELAN.

One important item which is worth noting here is that there were three different settings in the video talk show interviews, which were: (a) institutionalized interview such as political interview (b) semi-institutionalized interview (c) casual (ordinary) interview such as daily interviews between people and TV interviewer. In addition, it needs to be pointed out that the data used in the casual conversation were completely separated from an institutionalized or even semi- institutionalized situation. Therefore, the results of video talk show interviews about each of the above various settings of video interviews must not be generalized to all kinds of conversations. Both recorded video and audio files were brought into ELAN and then annotated.

ELAN is a special tool which has been utilized for the multi-level video and/or audio annotation that designed and developed by the Max-Planck-Institute (MPI) in Nijmegen. The ELAN tool was fundamentally planned for the transcription of different speech or conversation in American English. The MUMIN coding scheme was also used along with ELAN video annotation in order to have the accurate data. In this study, the non-verbal facial and head MUMIN coding scheme were annotated according to Table 1.

The kinds of gestures that appeared in the MUMIN coding scheme were hands gesture, fingers gesture and body posture which were significant and should be taken into account in the time of annotation of multimodal communication motions (Allwood et.al, 2008). In the present study, body gesture (e.g., hand and finger) and body posture were annotated based on Tables 2 and 3.

3.2.2 Questionnaire

To prevent participants from any confusing effect on their attempts to recognize the items written in English, the questionnaire was changed into the learners' native language, Persian. Additionally, the English version of the questionnaire was translated by the expert translator, and it was revised to avoid any parallelism.

The newly revised questionnaire on both verbal and non-verbal resources was distributed at the end of the semester among 37 participants who were asked to answer to the items of the questionnaire and determine their understanding and recognitions of the items. The Cronbach's alpha test was used to estimate the consistency of participants' responses to the questionnaire. The results showed a reliability coefficient of .81, which indicated that the responses to the questionnaire items were acceptable.

Table 1
Non-Verbal Facial and Head MUMIN Coding Scheme

Facial display aspects	*Expression form*	
Tier	*Value*	*Annotation in ELAN*
General face aspects	Smile	Smile
	Scowl	Scowl
	laughter	Laugh
Gaze	Up	Gaze-Up
	Down	Gaze-Down
	Besides	Gaze-Side
	Toward speaker	Gaze-Tow
Eyes	Both -closed	Eye-CB
	One-closed	Eye-CO
	Close repeatedly	Eye-RC
	Open-exaggerated	Eye-EO
Eyebrows	Raising	B-R
	frowning	B-F
Mouth and Lip	Closed mouth	Mouth-C
	Open mouth	Mouth-O
	Lip- rounded	Lip-RO
	Lip up	Lip-UP
	Lip down	Lip-Down
Head	Move forward	Head-MF
	Move backward	Head-MB
	Single nod down	Head-N
	Repeated nod down	Head-RN
	Sideway single tilt	Head-ST
	Sideway repeated tilt	Head-RT
	Move up	Head-MU
	Move down	Head-MD

An exploratory factor analysis was performed to recognize the construct validity of the elements presented in the learners' questionnaire responses. Moreover, principal component analysis method was utilized to test if there was empirical support for the researchers' hypothesis concerning the consistency of elements under several subscales of the questionnaire and to recognize some items that could be detached from the questionnaire.

Table 2
MUMIN Coding Scheme of Hand and Finger Gesture

Expression form		
Tier	*Value*	*Annotation in ELAN*
Hand/ Shoulder	Both hands up	BH-U
	Both hands down	BH-D
	Both hands sideways	BH-Side
	Single hands sideway	SH-Side
	Single hand up	SH-U
	Single hand down	SH-D
	Shoulders	Shoulders_UP
Finger	Fingers tapping	F-T
	Fingers pointing	F-P

Table 3
Body Posture MUMIN Coding Scheme

Body posture	*The form of gesture on body*	*Annotation in ELAN*
Torso	Torso turn right	TTR
	Torso turn left	TTL
	Torso bend forward	TBF
	Torso bend backward	TBB
	Torso lean right	TLR
	Torso lean left	TLL

During the semester, the researchers not only used various kinds of verbal elements such as intonation, rhythm, pitches, pauses within the utterances but they also utilized different types of non-verbal or visual signals such as hand motions, finger movements, facial displays, body gestures and body postures in the class speech in order to determine which of those signals was more effective for the better transmission of information to the students.

After utilizing those signals during teaching time in the class, the two final sessions of the semester were devoted to showing the whole video talk shows to the students in order to recognize the influence of these signals on the learning materials by the students. They were asked to detect which one of the verbal or non-verbal resources was effective in conveying data. Finally, in the final session, the questionnaires were distributed among the students, and they were asked to determine the extent to which the non-verbal resources were impressive for their learning in comparison to the verbal signals.

3.3 Data Collection Procedure

After downloading each of the video conversational data, the initial step in the data pre-processing processes was to change each recording file to the segregate MP4 file. In order to make each video prepared for the transcription, both MP4 recorded files were transferred into the software program 2.2.0 version of VLC media player and also Pot Player software program which contributed to VLC program in order to control the system and the various parts of video via the system-wide hotkeys. Controlling different parts of video files through system-wide hotkeys and VLC media player enabled the users to control the media playback.

The main reason for using Pot player was related to its video system-wide hotkeys for eliciting the non-verbal cues by means of changing the speed of playback of video talk shows then annotate each special part of visual or non-verbal resources which took place during the speech between the speaker one as an interviewer and the speakers two and three as interviewees.

In the present research, Pot Player and VLC media player were simultaneously utilized not only for providing exact evaluation of the visual signals by means of using slow-motion button but also for highlighting the non-verbal elements through mutual speech between or among individuals, so that the procedure of transcription was carried out simpler. The second step was to make the transcription ready along with the downloaded interactional data.

ELAN utilizes a special model of data which is time dependent (i.e., Vid1, Vid2) and the output file format which is produced by ELAN in XML format is so easy to change into other tools. Modifications can be produced in this regard simultaneously, new values, subjoining or eliminating new elements are considered in a straight line of the crossing point in ELAN. About oncoming work, there will be an opportunity in ELAN in order to broaden the annotation scheme by subjoining some linguistic information such as the concurrent processing with other contents in the framework of time or unit limitations.

4. Results and Discussion

The purpose of this study was to determine which one of the non-verbal or verbal signals was more significant in the transmission of information. As it previously noted, two types of data were collected. The first was related to the video talk show interviews and the second one was related to the questionnaire. Therefore, the results section is presented in two parts. The first section was allotted to the elicitation of data from the annotation of two video talk show interviews, which were carried out by the ELAN video annotation.

The second one was allocated to the data extraction performed by the statistical software program SPSS v.20.

4.1 Investigation on Annotation

For the purpose of understanding the non-verbal actions, there should be a particular attention to the occasion that the interviewer or even the interviewee finds him/herself in and the role of individual(s) who participate in the interaction and the context. As non-verbal signals play the role of special organizers in the interaction, they transmit information. For instance, when using hand movement or facial aspects. In both George Clooney and Jim Carrey interviews, the setting was casual (ordinary) and the interviewer or even the interviewees only had certain time to speak.

As tables 4 and 5 show, in the first interview, there were 495 verified non-verbal resources and in the second interview, there were 503 confirmed visual cues.

In each video talk show interview, the number of distribution of non-verbal signals between interviewer and interviewee are shown in Tables 4 and 5.

Table 4
Distribution of Non-verbal Signals of the First Video Talk Show

	David Letterman and George Clooney`s Interview (First Video)							
	Interviewee				*Interviewer*			
Visual Signals	*Gesture*	*Facial aspects*	*Body*	*Total*	*Gesture*	*Facial aspects*	*Body*	*Total*
No. of signals	178	125	52	355	65	42	33	140
Each Percent	50.1%	35.2%	14.7 %		46.4%	30%	23.6%	
Total Percent	50.1%	35.2%	14.7 %	100%	46.4%	30%	23.6%	100%

As shown in Tables 4 and 5, the total number of visual cues distribution of interviewee (n_G = 355, n_J = 350) was more than that for the interviewer (n_1 = 140, n_2 = 153) in both videos; therefore, the time of the interview which was assigned to each interviewee was more than the time allotted to the interviewer, which actually is crystal clear basically with a glimpse to both talk shows.

Table 5
Distribution of Non-verbal Signals of the Second Video Talk Show

David Letterman and Jim Carrey`s Interview (Second Video)								
	Interviewee				*Interviewer*			
Visual Signals	*Gesture*	*Facial aspects*	*Body*	*Total*	*Gesture*	*Facial aspects*	*Body*	*Total*
No. of signals	160	129	61	350	70	55	28	153
Each Percent	45.7%	36.9%	17.4%		45.8%	35.9%	18.3%	
Total Percent	45.7%	36.9%	17.4%		45.8%	35.9%	18.3%	

The comparison between the distribution of the non-verbal signals of the first and the second video interview revealed that the number of gestures used by George Clooney (n_1= 178) was more than that by Jim Carrey (n_2= 160); however, the number of facial features, body gestures, and even body postures used by Jim Carrey (n= 129 and 61) was more than that by George Clooney (n= 125 and 52). This contrast could be due to their different characteristics.

It needs to be pointed out that the highest difference took place in each turn during the interviews. The other issue which was found during the talk show interviews was allotted to non-verbal resources for both the interviewer and interviewee that could be more sophisticated or even simpler by utilizing their facial features such as gaze, eyebrows, nods, or hand movements (e.g., both hands or just single hand movement), both shoulders or single shoulder movement, fingers movements etc. and torso (i.e., the motions of upper part of the body such as torso turn right, torso turn left, torso bend forward, torso bend backward, torso lean right, and torso lean left).

In Tables 6 and 7 and also Figures 2 and 3, the comparison between the visual elements used by the interviewee and the interviewer in both video talk shows were presented at the certain period of time.

As shown in Table 6 and Figure 2, at certain period of time in the first video talk show, the non-verbal signals used by the interviewer and interviewee were extracted by the ELAN video annotation. For instance, the interviewer utilized ‘single hand sideways’; whereas the interviewee used ‘both hands repeated up and down’ and also ‘both hands sideways’ at the previously specified time of interview.

Table 6
Comparing the Visual Elements Used by the Interviewee and Interviewer for the First Video Talk Show

George Clooney and David Letterman`s Interview (First Video)			
The Period of Time	*Non-Verbal Signals*	*Interviewer*	*Interviewee*
07:30.485 to 07:31.685	Gaze (Up, Down, Besides, Toward speaker)	Gaze down	(Gaze side) and (Gaze down)
	Head (Sideway single tilt, Sideway repeated tilt and etc.)	Head nod	Single Tilt
	Hand (Both hands up, Both hands down and etc.)	Single hand sideways	(Both hands repeated up and down) and (Both hands sideways)
	Body posture (Torso bend forward , Torso bend backward)	Torso Lean Left	Torso Bend Backwards

As revealed in Figure 3 and Table 7, at the specified period of time in the second video talk show, the visual resources utilized by the interviewer and the interviewee were elicited by the ELAN video annotation. For example, the interviewer utilized ‘Gaze down’ and ‘Gaze Toward speaker’, while the interviewee used only ‘Gaze down’ at the previously pointed out time of interview.

4.1.1 The Most Practical Non-verbal Cues Elicited From both Video Talk Shows

Given the extracted files from the ELAN video annotation obtained from both video talk show interviews, it was revealed that the rate of the usage of some non-verbal signals was more than that for other visual resources; in other words, some of which were more practical and useful than the other known NV cues. For instance, in the facial features of the non-verbal elements, the rate of the usage of gaze, eyebrow movements, head motions and smile was more than the other facial aspects. In addition, in body gestures and movements, the rate of the usage of hand motions was more than finger movements and in the body postures that was only torso with various subsections, the rate of the usage of *torso bend backward*, *torso lean left*, *torso lean right* was more than the other torsos, respectively.

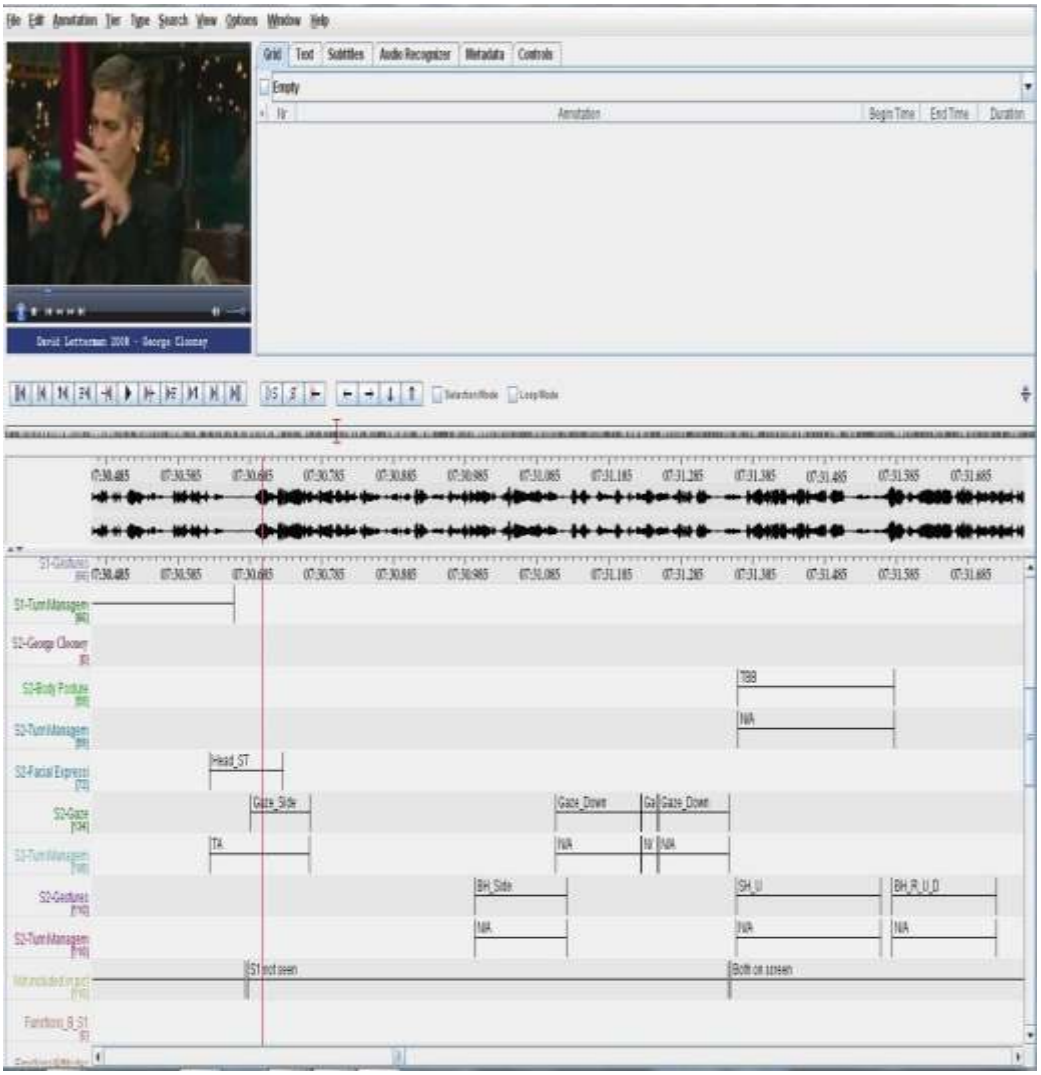

Figure 2. Comparing the Visual Elements Used by the Interviewee and the Interviewer for the Second Video Talk Show Elicited by the ELAN Video Annotation

4.2. Questionnaire: *Assessment of Non-verbal and Verbal Signals*

In this section, the two subcategories of the data obtained from the students` questionnaires are discussed. First, the statistical differences between the non-verbal and verbal signals were investigated. To this end, an independent sample t-test was performed. Then, the differences among the non-verbal items of the questionnaire were presented. In this case, a chi-square test was conducted to show which of the visual elements was more effective in the transmission of information to the students compared to the other NV

resources. The entire abbreviations utilized in the present research are presented in Appendix C. The results of the students` questionnaires were comprehensively discussed in the section below.

Table 7
Comparing the Visual Elements Used by the Interviewee and the Interviewer for the Second Video Talk Show

Jim Carrey and David Letterman Interview (Second Video)			
The Period of Time	*Non-Verbal signals*	*Interviewer*	*Interviewee*
	Gaze (Up, Down, Besides, Toward speaker)	Gaze down	(Gaze down) and (Gaze Toward speaker)
08:26.957	Hand (Both hands up, Both hands down and etc.)	Single hand sideways	Both hands sideways
to	Eyebrows (Raising and Scowling)	Raising and Scowling	Raising
08:28.057	Body posture (Torso bend forward , Torso bend backward)	Torso Lean Left	Torso Bend Forward

4.2.1. Investigating the Differences between Non-verbal and Verbal Signals of the Questionnaire

Before investigating whether verbal or non-verbal signals could lead to better transmission of information in the mutual conversation, a group statistic was carried out to calculate the visual representation score. Table 8 shows the results of this test.

As appeared in Table 8, the mean score of non-verbal representation resources was greater than that of verbal elements. It can be claimed that there was an obvious difference between the visual and the verbal cues in conveying information from the students` viewpoint [$M_{(NV)}$ = 49.56, $M_{(V)}$ = 38.64]. However, the standard deviation of NV elements was lower than that of verbal signals [$SD_{(NV)}$ = 3.73, $SD_{(V)}$ = 4.49]. It implies that the NV cues instruction did not have the same influence on the students; therefore, it can be claimed that the visual elements were more effective than the verbal cues in the transmission of information from the students' viewpoint.

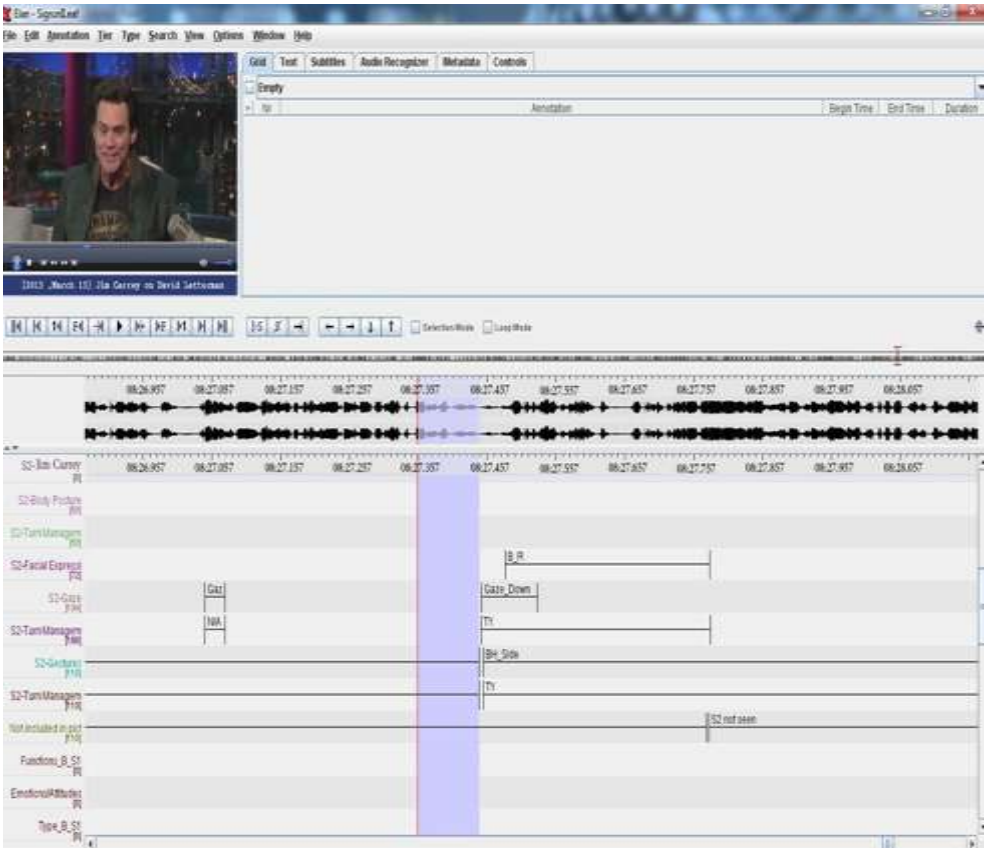

Figure 3. Comparing the Visual Elements Used by the Interviewee and the Interviewer for the Second Video Talk Show Elicited by the ELAN Video Annotation

Table 8
Descriptive Statistics of Visual and Verbal Resources

	Category		*M*	*SD*	*Std. Error Mean*
Visual Representation Score	Visual	37	49.56	3.73	.61
	Verbal	37	38.64	4.49	.73

In order to find out which one of verbal or non-verbal resource was more beneficial for the transmission of data, an independent samples t-test was performed. Table 9 presents the results of the comparison between the verbal and the visual cues.

As Table 9 shows, the result of *t* test revealed a significant difference between the scores of non-verbal and verbal elements with regard to the use of NV resources in the speech, t (72) = 11.35, p= .000. The magnitude of the

differences in the means (mean difference = 10.91, 95% CI: 9.00 to 12.83) was very large. The statistic of eta squared (.64) indicated a large effect size. The previous statistic showed that NV instruction was completely effective in conveying the information to the students.

4.2.2. Investigating the Differences between the Non-verbal Questions of the Questionnaire

In this section, the purpose was to investigate which of the NV cues had more impact on the students in the transmission of information comparing to other visual elements. Thus, to this end a chi-squared test was carried out.

In order to compare each NV cues` questions together, researcher should distinct the questions` number of each visual signal from those of verbal resources; then compared each one of them together. After comparing each 2 by 2 questions, the results showed that there were significant differences between 2 questions.

Table 9
Independent Sample Test Comparing Visual and Verbal Signals

	Levene's Test for Equality of Variances	t-test for Equality of Means						
							95% Confidence Interval of the Difference	
	F	t	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances assumed	.74	11.35	72	.000	10.91	.96	9.00	12.83
Equal variances not assumed		11.35	69.67	.000	10.91	.96	9.00	12.83

4.2.2.1. Comparison between the Hand Movement and Head Motions

The result of the comparison between these questions revealed that there was a significant difference between the hand movement and the head motions (p=.007). Table 10 presents the result of the test.

As Table 10 indicates, there was a significant difference between the hand motions and head movement in conveying information from the students` perspectives (p = .007). The frequency of the hand movement (66.7%) in the transmission of information was greater than that of head motions (63.6%).

Table 10
Comparing the Hand Movement and Head Motions by Chi-Square Tests

	Value	*Df*	*p*
Pearson Chi-Square	22.818[a]	9	.007
kelihood Ratio	20.370	9	.016
Linear-by-Linear Association	1.323	1	.250
N of Valid Cases	37		

a. 15 cells (93.8%) have expected count less than 5. The minimum expected count is .32.

4.2.2.2. Comparison between the Smiling and Laughing

In this section, students` views on the impact of the smiling and laughing on the transmission of information were compared.

As Table 11 shows, there was a significant difference between smiling and laughing in conveying information from the students` views (p= .000). The frequency of laughing (81.2%) in conveying data was greater than that of smiling (72.2%).

In order to have a better interaction in conversation, this study recommends that TEFL instructors familiarize their students with the importance of using non-verbal elements through their speech because visual signals can convey more information to the listeners than verbal resources.

The results of ELAN video annotation obtained from both the first and the second video talk shows showed that the frequency of some non-verbalelements such as hand movements, eyebrow motions, gaze (gaze down and gaze toward speaker) and torso (torso lean left or right and torso bend forward or backward) were more than the other NV cues. The result of this investigation was similar to the study on non-verbal representation signals of multimodal analysis conducted by Pajo (2013) and also Koutsombogera and Papageorgiou (2009) in the field of hearing impairments` individuals and deaf people. Additionally, the results of the study of Koutsombogera and Papageorgiou (2009) showed several functions of non-verbal communication; in other words, interlocutors conveyed numerous messages by means of NV representation elements. In fact, the result of present study was along with the data of above study which revealed that non-verbal cues were more effective in the transmission of information than the verbal cues.

Moreover, NV resources of communication made clear the conversation of the interlocutors since they were powerful resources for self-expression. The result of this research also showed that visual signals were more practical and effective in the transmission of data. This finding was in contrast with that of Chen (2011) in that the verbal signals such as pitch and silence played significant role in the language instruction showing that the non-verbal

elements were more effective in the transmission of information than the verbal cues.

Table 11
Comparing the Smiling and Laughing signals

	Value	*df*	*p*
Pearson Chi-Square	35.567[a]	9	.000
Likelihood Ratio	38.929	9	.000
Linear-by-Linear Association	16.580	1	.000
N of Valid Cases	37		

a. 13 cells (81.2%) have expected count less than 5. The minimum expected count is .05.

5. Conclusion and Implications

The elicited data of the present study showed that in the video talk show conversation, by the omission of vocal resources, the purpose of the speaker(s) was not only interrupted but also the listener continued to understand the speaker`s speech through indicated vocal words within the sentence(s) just before the interruption of vocal part of speech, lip reading and the visual representations. Although, during such a kind of speech, the listener encounters with some problem(s), but it does not mean that any utterances of the speaker(s) are unknown.

In the best of our knowledge this is the first time that used ELAN video annotations in order to understand, which of NV cues could transmit more information to the listener. The result of the this study from ELAN video annotations and the questionnaire revealed that visual signals were more impressive than the verbal signals and also the verbal elements played the role of complementary factor for the non-verbal resources. Thus, visual cues were considered as independent elements and verbal signals as the dependent factors. The outputs of ELAN for both video interviews showed that the frequency of some NV elements such as hand movements, eyebrow motions, and torsos were more than the other visual cues.

The findings of independent *t*-test revealed a significant difference between the scores of non-verbal and verbal elements with regard to the use of NV resources in speech. Therefore, it could be stated that the non-verbal elements were the main part of each utterance. Additionally, the result of chi-square test showed a significant difference in learning materials via using certain non-verbal signals such as hand movements, general face aspects, and head motions. The results of the present study can be utilized to expand our recognition about multimodal communication and to make an accurate direction for face-to-face interactional action in conversation.

It is important to note that the present study was not only beneficial for instructors to better transmission of the lessons` contents to the students but

also useful for students to receive data better and use non-verbal elements in order to have a better interaction in conversation. Additionally, this research could be of use for individuals with hearing impairments and deaf people. Moreover, this study could be of use for teaching the video and cassette translation course in the field of B.A. English translation at the university.

This study was performed on NV descriptive signals for the interactional action in conversation. It would be beneficial to conduct more research investigating the effect of NV functional resources for conversation. Furthermore, the focus of this study was on the non-verbal elements. Other studies can investigate the role of verbal cues for the interactional action in conversation. In addition, this study was only carried out on male students. It could be valuable to investigate the effect of visual resources on female students` conversation. More research can be conducted to improve the design of the study with various types of interviews such as other forms of video talk shows with institutionalized and semi-institutionalized interactions in order to study the modifications in the multimodal branches between different forms of interview settings.

References

Allwood, J., Cerrato, L., Jokinen, K., Navarretta, C., & Paggio, P. (2008). The MUMIN coding scheme for the annotation of feedback, turn managements and sequencing phenomena. Multimodal corpora for modeling human multimodal behavior. *Journal on Language Resources and Evaluation, 41*(4), 273-287.

Aran, O., & Perez, D. G. (2011). Analysis of social interaction in group conversations: Modeling social verticality. In A. Salah, & T.Gevers (Eds.), *Computer analysis of human behavior* (pp. 293-322). Springer.

Ba, S., & Odobez, J. M. (2011). Multi-person visual focus of attention from head pose and meeting contextual cues. *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 33*(1), 101-116.

Boersma, P., & Weenink, D. (2007). *Praat: Doing phonetics by computer 5.0.02.* University of Amsterdam: Institute of Phonetic Sciences. Retrieved from http://www.praat.org.

Carroll, W. R., & Bandura, A. (1982). The role of visual monitoring in observational learning of action patterns: making the unobservable observable. *Journal of Motor Behavior, 14*(2), 153-167.

Chen, H. K. Y. (2011). *Sound patterns in Mandarin recycling repair* (Doctoral dissertation). Department of Linguistics. Faculty of the Graduate School, University of Colorado.

Esfandiari, G. B., & Ágnes, A. (2013). An overview of multimodal corpora, annotation tools and schemes. *Debreceni Egyetemi Kiadó, 9,* 86-98.

Foster, M. E., & Oberlander, J. (2007). Corpus-based generation of head and eyebrow motion for an embodied conversational agent. *Proceedings of the International Language Resources and Evaluation Conference (LREC), 41*(3/4), 305-323.

Haddington, P., & Kääntä, L. (2011). *Language, body and interaction: A multimodal perspective into social action*. Helsinki: Finnish Literature Society (SKS).

Jewitt, C. (2006). *Technology, literacy and learning: A multimodal approach.* London: Routledge.

Jewitt, C. (Ed.). (2009). *Handbook of multimodal analysis*. London: Routledge.

Jokinen, K. (2009). Gaze and gesture activity in communication. In: C. Stephanidis, (Ed.), *Universal access in human-computer interaction* (pp. 495-506). Helsinki: Gaudeamus Helsinki University Press.

Jokinen, K., & Vanhasalo, M. (2009). Stand-up gestures – annotation for communication management. *Proceedings of the Multimodal Workshop at Nodalida Conference*, Denmark.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Knight, D. (2009). *A multimodal corpus approach to the analysis of back channeling behavior* (Doctoral dissertation). The University of Nottingham. Retrieved from www.core.ac.uk/download/pdf/98821.pdf.

Koutsombogera, M., & Papageorgiou, H. (2009). Multimodality issues in conversation analysis of Greek TV interviews. *Multimodal signals: Cognitive and algorithmic issues LNAI, 5398, 40-46.* Retrieved from www.springer.com/content/pdf.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, New Jersey: Lawrence Erlbaum.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University Of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. Chicago and London: University of Chicago Press.

Kipp, M. (2004). *Gesture generation by imitation: From human behavior to computer character animation* (Doctoral dissertation, Saarland University). Retrieved from www.Dissertation.com.

Musgrave, D. J. (2012). A multimodal analysis of the communicative utterances of a language competent bonobo (*Pan paniscus*) (Master´s thesis). English Studies Department, Faculty of the Graduate School, Iowa State University.

O'Halloran, K. L., Smith, B. A., Tan, S., & Podlasov, A. (2010). Challenges in designing digital interfaces for the study of multimodal phenomena. *Information Design Journal, 18*(1), 2-12.

O'Halloran, K. L., & Smith, B. A. (2012). *Multimodal text analysis.* Singapore: Singapore University Press.

Pajo, K. (2013). *Joint multimodal management of hearing impairment in conversations at home* (Unpublished doctoral dissertation). Behavioral Sciences Department, Faculty of Behavioral Sciences, University of Helsinki.

Pallant, J. (Ed.). (2010). *SPSS survival manual.* New York, NY: Open University Press.

Pastra, K., & Wilks, Y. (2004). Image-language multimodal corpora: Needs, lacunae and an AI synergy for annotation. *In Proceedings of the Language Resources and Evaluation Conference* (pp. 767-770). Athens: Institute for Language and Speech Processing.

Saferstein, B. (2004). Digital technology- methodological adoption: Text and video as a resource for analytical reflectivity. *Journal of Applied Linguistics*, 12, 197-223.

Skelt, L. (2006). *See what I mean: Hearing loss, gaze and repair in conversation*. The Australian National University: Canberra.

Streeck, J., Goodwin, C., & LeBaron, C. (2011). Embodied interaction in the material world: an introduction. In J. Streeck, C. Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and the body in the material world* (pp. 1-26). Cambridge: Cambridge University Press.

Vilhjálmsson, H. H. (2009). Representing communicative function and behavior in multimodal communication. In A. Esposito, A. Hussain, M. Marinaro, & R. Martone (Eds.), *Multimodal signals: Cognitive and algorithmic issues lecture notes in artificial intelligence* (pp. 47-59). Berlin Heidelberg: Springer-Verlag.