

What's in a Voice? Exploring the Vocal Qualities of the Spanish Dubbed Voice in Emotionally-Loaded Scenes

Beatriz Naranjo

Universidad de Murcia, SPAIN

beatriz.naranjo@um.es

ABSTRACT

This paper aims at investigating whether intentional efforts of voice actors to adopt different levels of naturalness when dubbing scenes portraying different emotions could be detected by the target audience. Eight scenes representative of anger (4) and sadness (4) were used in this study and two dubbed versions into Spanish ('natural' and play-acted) were recorded for each one of them. Firstly, an acoustic analysis was carried out to detect potential distinctive vocal patterns of natural vs. play-acted dubbed voices. Later, we carried out a reception study to find out whether viewers perceived any differences between the two versions. Although the automatic voice analysis did not account for significant differences, some differentiating traits in play-acted versions (vocal fry, elongation of sounds, breathy voice) were observed. Natural versions were perceived as more natural than play-acted versions, with higher scores in the anger scenes. Finally, while participants generally preferred the natural over the play-acted style, a relevant percentage of participants expressed a preference for the play-acted versions of sad clips.

KEYWORDS: dubbese; dubbing; emotions; naturalness; vocal performance; voice quality

1. Voice and Emotion in Film

Together with the quality of the script itself, the actors' voices have always been considered an essential component of film dialogues. As well as playing a role in the identity of screen characters (Smith 2007), and conveying meaning (Klevan 2005), voice can also produce powerful emotional effects in the audience (Schingler 2006), as an essential part of the aesthetic experience.

The links between voice and emotion have been widely acknowledged by the research community. A wide array of previous theoretical and empirical studies back up this connection. When listening to voices, humans are able to attach meaning to sound, and therefore, accurately identify the speakers' emotional state, thanks to the ventral auditory stream (Johnstone et al. 2006). According to scholars specialized in the field of sound and emotion, sadness and anger

are the most accurately perceived emotions by listeners (Scherer 1995). Furthermore, voice does not only give us clues about speakers' mood, but it can also induce emotions in the listener. Higher levels of activation in specific brain regions have been detected when processing voices with different emotional cues, such as the middle temporal gyrus, involved in processing complex sounds including speech and emotional prosody, as well as the frontal gyrus, associated with emotional judgment and reward (*ibid.*: 247).

Previous literature shows how there are many vocal features involved in the expression and perception of emotions, such as pitch, segmental features and extralinguistic interjections (Scherer 1995: 238; Gobl & Chasaide 2003: 190, Johnstone et al. 2006: 242; Pietrowicz et al. 2017: 792-793). Firstly, pitch, which is defined as the subjective perception of the highness or lowness of a tone, is possibly one of the most acknowledged properties of the voice associated with emotions. The scholars mentioned above have pointed to several aspects related to the pitch, such as the fundamental frequency (f_0), which is the acoustic measurement of pitch, the pitch contour or rise and fall melody of pitch, as well as differences of intensity or power carried by sound waves. Secondly, segmental features related to the glottal and supraglottal activity of the vocal cords also seem to play a relevant role in emotional expression. Finally, studies refer to the affective meaning of both articulated and natural interjections or "affect bursts" (Scherer 1995:236) such as sighs and cries.

The so-called basic emotions in the field of psychology such as fear, anger, surprise sadness and joy, have been extensively associated with specific acoustic parameters, which are usually measured in a dichotomist scale (e.g. high vs. low pitch, up vs. down pitch contour, small vs. large amplitude variation). Some pairs of emotions present acoustic similarities, such as angry and happy vocal expressions, whereas others, such as sadness and surprise are considered antagonistic from the acoustic point of view (see Scherer 1995; Johnstone et al. 2006). Perhaps not surprisingly, some of the acoustic criteria used to define the properties of vocal expressions of emotions such as pitch level and intensity seem to coincide with the level of physiological activation (also called 'arousal') with which they have been associated in the field of psychology. For example, research has revealed how happiness is associated with higher levels of heart rate variability and skin conductance response than sadness (Khalfa et al. 2008: 24; Shi et al. 2017: 536-537). The two emotions that concern us in this study, namely anger and sadness, present contrary acoustic features; while anger is characterized by a strong intensity, high fundamental frequency, high rate of articulation (i.e., fast speaking speed) and upward

pitch contour, the opposite low and downward pattern prevails in sad vocal expressions (see Scherer 1995; Sobin & Alpert 1999; Johnstone et al. 2006; Niculescu et al. 2013).

However, a purely acoustic analysis may not always fully account for the spectrum of nuances that humans are able to identify in speakers' voices. Some of the most frequently mentioned qualities of the voice in film and play-acted scenarios include harsh or husky voice, vocal fry (also called creaky voice) and breathiness (audible air escape) or whispery voice, which have been identified in different languages such as Swedish (Gobl & Chasaide 2003), German (Jürgens et al. 2011), French (Bosseaux 2019) and different varieties of English (Shingler 2006; Pietrowicz 2017). Studies have shown how these vocal patterns can also play a role in terms of the expression and perception of emotions. In fact, as Gobl & Chasaide's study (2003) suggest, they could potentially be a rather universal indicator of "subtle variations in affective states" (2003, 208) since listeners are able to perceive them, even when they do not understand the language of the utterances. For example, whereas tense and harsh voice have been related to highly aroused states both with positive (happiness) and negative valence (anger, stress), breathy and creaky voices have been associated to low levels of activation also with both positive (relaxed, friendly, intimate) and negative valenced moods (sad, bored) (Gobl & Chasaide 2003). More subtle distinctive traits have also been identified as portrayals of specific emotions; for example, falsetto seems to be indicative of fear (*ibid.*), while deep throaty vocal expressions or 'husky' voice have been attributed to weary and cynical attitudes (Shingler 2006).

2. Voice and Naturalness in Dubbing

Another aspect of voice quality closely connected to the expression of emotions in the context of film studies is naturalness as opposed to playacting. Whereas traditionally, playacted expressions have been considered as more overacted, more aroused and more stereotypical than spontaneous speech, more up-to-date research has overthrown this misconception (Jürgens et al. 2011; Juslin et al. 2018).

Admittedly, studies have demonstrated how listeners are able to clearly distinguish playacted and natural portrayals of emotional speech (Audibert & Rilliard 2010; Jürgens et al. 2011; Juslin et al. 2018) with posed clips being rated higher in emotional valence and intensity (Juslin et al. 2018). What research has not yet been able to clarify, though, is which specific acoustic

patterns differentiate natural from playacted talk, as studies in the field have often led to inconclusive results (Juslin et al. 2018). However, several recent works coincide in pointing toward fundamental frequency (f₀) as a distinctive key parameter between playacted and spontaneous talk. Analyses conducted by Audibert & Rilliard (2010) and Jürgens et al. (2011) reveal not only higher pitch or f₀ general levels, but also more variable f₀ or pitch contour in play-acted samples when compared to natural tokens.

In the field of Audiovisual Translation, some authors have also advocated for the study of voice quality in dubbing, due to its crucial role in the portrayal and credibility of screen characters (Palencia Villa 2002; Pérez González 2007; Bosseaux 2008; 2019; Whittaker 2012; Bruti & Zanotti 2017). Some studies (Naranjo 2016) have also tangentially touched upon the reproduction of distinctive phonological and prosodic traits used in the original version to preserve the identity of characters, such as the case of *falsestos* and rising final pitch contour when dubbing speakers of African American Vernacular English (AAVE) into Spanish. However, research carried out by Sánchez-Mompeán (2017; 2019; 2020) possibly constitutes the most relevant contribution to the field in this regard. To seek to gain a deeper insight into the vocal traits of dubbed dialogues, the author's work gives a pioneering and exhaustive account of the prosody and paralinguistic traits of dubbed speech in Spanish. Focusing on the analysis of dialogues from the sitcom *How I Met Your Mother*, several patterns which characterize the artificial paralinguistics of dubbed talk or *dubbitis* are identified. Such patterns include fluctuations in pitch, elongation of sounds at certain points while uttering the lines, faster or slower tempos than would be natural in spontaneous talk, as well as tense and precise (vs. slurred) articulation, which consists in sharp contrasts between strong and weak syllables combined with an extremely clear pronunciation. According to the author, two possible factors could explain these peculiarities of dubbed speech. The first one would be related to the attempt of voice actors to synchronize with the character's mouth, which may especially justify some of these features, such as tempo variations and elongation of sounds. The other factor has to do with the fact that dubbers may not have time to memorize their lines and they adopt the style of 'reading aloud' at the time of recording. Finally, Sánchez-Mompeán (ibid.) refers to the implications that these traits can have in terms of emotionality, since sometimes they can carry attitudinal and pragmatic implications not necessarily present in the original, such as pretentiousness, hesitation, less emotional involvement, etc.

3. The Study

Drawing from the distinctive vocal features found by previous literature in emotional play-acted vs. natural speech, I carried out a study which aimed at identifying paralinguistic patterns that distinguish between ‘natural’ and play-acted dubbing styles in Spanish from both an acoustic and a perceptual point of view. It is worth pointing out that the term ‘natural’ here does not refer to actual spontaneous speech since the voice actors commissioned to dub the scenes followed a script. They were instructed however to use a more natural style for one of the versions and a more artificial or play-acted style for the other version. I relied on the actors' experience and their own personal intuitive idea of natural vs. play-acted dubbing styles and no specific instructions were provided other than our definition of ‘naturalness’ in play-acted contexts, i.e., intended to sound like Spanish native speakers' speech engaged in spontaneous conversation.

3.1 Objectives

1. To find out whether intuitive natural and play-acted styles in dubbing differ in terms of acoustic traits.
2. To find out whether intuitive natural and play-acted styles in dubbing can be distinguished by the target audience in terms of naturalness.
3. To determine whether the emotion portrayed in the scene can influence perceived differences between natural and play-acted style.
4. To investigate potential viewers' preferences for one dubbing style (natural vs. play-acted) over the other.

3.2 Methods

3.2.1 Participants

A total of 59 students (22 male and 37 female) with a mean age of 19 participated in this experiment. All participants were undergraduates with Spanish as their mother tongue who did not have university training in translation or foreign languages. They were split into two groups (A and B) with 30 and 29 participants, respectively. Group B was more balanced in terms of participants' sex with 13 males and 16 females, whereas group A contained 9 males and 21 females. Most participants (70% of participants in group A and 72.4% in group B) in both

groups indicated having a habit of watching dubbed audiovisual products in Spanish at least weekly or monthly.

3.2.2 Materials

Two sets of four video clips (10-30 seconds) were chosen from the film *Like Crazy*, a 2011 American romantic drama directed by Drake Doremus and distributed by Paramount Vantage, starring Anton Yelchin and Felicity Jones. The clips portrayed conversations of high emotional intensity held between a young couple. Sadness and anger were the representative emotion of each set of clips. For each clip, two dubbed versions ('natural' and play-acted) in Spanish were recorded. Differences between versions lay in specific lexical choices as well as orality discourse markers to make the 'natural' version closer to real spontaneous conversation in Spanish, whereas script in the dubbed version contained some features of *dubbese*, the prefabricated language of dubbing (see scripts in Annex I). The four clips for sadness were titled 'SHOCKED', 'BUSY', 'BYE' and 'COME OVER' and the anger scenes were named 'LOOK AT ME', 'MESSAGES OUT', 'PARANOID' and 'SHOUTING'.

3.2.3 Procedures

I conducted two types of analysis in this study: an analysis of the voice recordings to detect potential acoustic differences between the natural and play-acted versions and a quantitative analysis of the viewers' assessment of naturalness and their preferences. Automatic voice analysis was carried out using *Praat* version 6.1.16, a software for speech analysis in phonetics and it was completed with a qualitative evaluation of other speech traits related to voice quality not identifiable by the software.

I carried out the reception study *in situ* in a university classroom, and students used their own laptops or tablets and earphones. Each group of students watched a different set of scenes. Participants in group A watched the natural and play-acted dubbed versions of the ANGER set of clips. Participants in group B did the same with the SADNESS scenes. To avoid carry-over effects¹, the natural and play-acted versions were presented in a different order each time. After

¹ When the same subjects are used for two or more experimental sessions with different stimuli, it is possible that the first condition may alter the subject's perception, behaviour or performance in subsequent experimental condition. This is known as 'carryover effect', i.e., an effect that is 'carried over' from one experimental condition to the other.

watching each pair of clips, I requested the participants to complete a brief questionnaire in which they were asked to assess the level of naturalness of both versions in terms of script and intonation following a 4-point Likert scale². In the last two questions participants were to choose the version that had best suited their preference and the reasons for such preference. Both the video clips and the questionnaires were inserted in a *GoogleForm* which allowed me to collect participants' answers online.

3.3 Analysis

I first carried out automatic analysis with *Praat* delivered voice reports for each clip recording, providing several acoustic measures. I selected values from the complete reports according to their usefulness for this study, which are as follows:

- Number of voice pulses. Glottal pulses are defined as short bursts or 'puffs' of air.
- Number of voice breaks. These refer to the number of intervals of silence between pulses.
- Pitch mean (M) and standard deviation (SD). Defined as the frequency of a sound wave commonly expressed in Hz, pitch refers to the highness or lowness with which sounds are perceived and it determines tone and intonation in voice analysis. Since previous studies have pointed to pitch variability as a feature of play-acted and dubbed speech (Audibert et al. 2010; Jürgens et al. 2011; Sánchez-Mompeán 2017; 2019; 2020), we have calculated the coefficient of variation (CV), which shows the extent of variability of pitch measures in each clip in relation to the mean.

Besides this automatic analysis, other relevant vocal features not included in these voice reports were also detected by my listening to the recordings and examining the spectrograms and pitch contour graphs provided by *Praat* for each recording. These vocal features include instances of elongation of sounds, presence of vocal fry (or 'creaky voice') and the number of pulses containing breathy voice, as they have also been highlighted as traits of acted and dubbed speech in previous literature (Jürgens et al. 2011; Pietrowicz et al. 2017; Bosseaux 2019; Sánchez-Mompeán 2017).

² A Likert scale is a rating scale (typically ranging from 1 to 5 or 1 to 7) used to collect participants' attitudes and opinions in survey research. Normally, they are used to ask participants how much they agree or disagree with several statements.

With the survey data I had collected during the reception study, I conducted descriptive and inferential statistical analyses using the SPSS package (Statistical Package for the Social Sciences) to find out about perceived naturalness and preferences for the two dubbed versions. For this purpose, I ran a two-way ANOVA (Analysis of Variance) test to measure the effects of the emotion in the scene (anger vs. sadness) and in the dubbed versions (natural vs. play-acted) on viewers' perception of naturalness. Finally, I then calculated the percentages to compare the participants' preference for the play-acted or 'natural' versions in both scenes, as well as their reasons for such preferences, which I classified in four categories:

- identification with characters,
- involvement with the scene
- professionalism of voice-actors and
- credibility of screen actors.

4. Results

4.1 Voice Analysis

I carried out separate analyses for the clips corresponding to each scene. Table 1 (below) shows results for the automatic voice report (pitch variability, total number of pulses and voice breaks) as well as other significant vocal parameters (elongation of sounds, vocal fry and breathy voice) observed in the recordings from the sad scene. Potentially significant differences were only found for pitch variability in the clip 'BYE'. It was the only case where I found that the coefficients of variation (CV) differ by almost 0.2 points between the play-acted and the natural version [Play-acted=0.27; natural=0.11]. Number of pulses and voice breaks also presented differences due to the different number of syllables contained in the scripts of the two versions. This is not surprising as the scripts were not completely identical. The higher values in play-acted vs. natural versions may be due to higher articulatory precision with which play-acted lines were uttered. In 'COME OVER', though, the replacement of '*quieres*' (do you want) by '*te apetecería*' (would you feel like) together with the addition of the colloquial discourse marker '*en plan*' (like) resulted in a higher number of syllables, which explains the higher values in number of pulses and voice breaks. Play-acted versions presented more instances of elongation of sounds and number of pulses with breathy voice than natural versions in all clips.

Finally, vocal fry was present in the play-acted version and absent in the natural version in 3 out of 4 clips.

Table 1: Results of the voice analysis for clips extracted from the sad scene.

Clip name	Pitch (Hz) M(SD)/CV	Total no. of pulses	Voice breaks	Elongation of sounds	Vocal fry	No. pulses containing breathy voice
SHOCKED						
<i>Play-acted</i>	193(46)/0.24	586	11	4	Yes	2
<i>Natural</i>	205(46)/0.23	537	12	1	No	1
BUSY						
<i>Play-acted</i>	180(56)/0.31	1054	24	3	Yes	3
<i>Natural</i>	151(53)/0.35	701	23	1	No	0
BYE						
<i>Play-acted</i>	163(43)/0.27	232	10	2	Yes	5
<i>Natural</i>	177(19)/0.11	173	6	0	No	1
COME OVER						
<i>Play-acted</i>	274(75)/0.27	975	20	2	No	2
<i>Natural</i>	244(72)/0.30	1182	26	0	No	0

For the anger scene, as displayed in Table 2, I found similar results in the recordings. In this case, large differences in pitch variability were not found in any of the clips. Again, the different number of pulses and voice breaks between both versions can be explained by the different scripts. More occurrences of elongation of sounds and breathy voice in the play-acted versions were found in 2 out of 4 clips. As in the sad scene, vocal fry was present in the play-acted version and absent in the natural version in 3 out of 4 clips. Together with a higher number of pulses, the analysis also reveals a higher number of voice-breaks for the natural version of the clip 'SHOUTING', contrarily to what happens with the rest of the clips. Again, actors' higher articulatory precision when uttering the syllables in the play-acted versions may explain that there are more voice breaks or separations between pulses. In the case of 'SHOUTING', however, the deliberate high articulation rate (speaking speed) which with the first line was uttered in order to portray the angry tone may be the cause that the natural version resulted in a higher number of pulses and breaks.

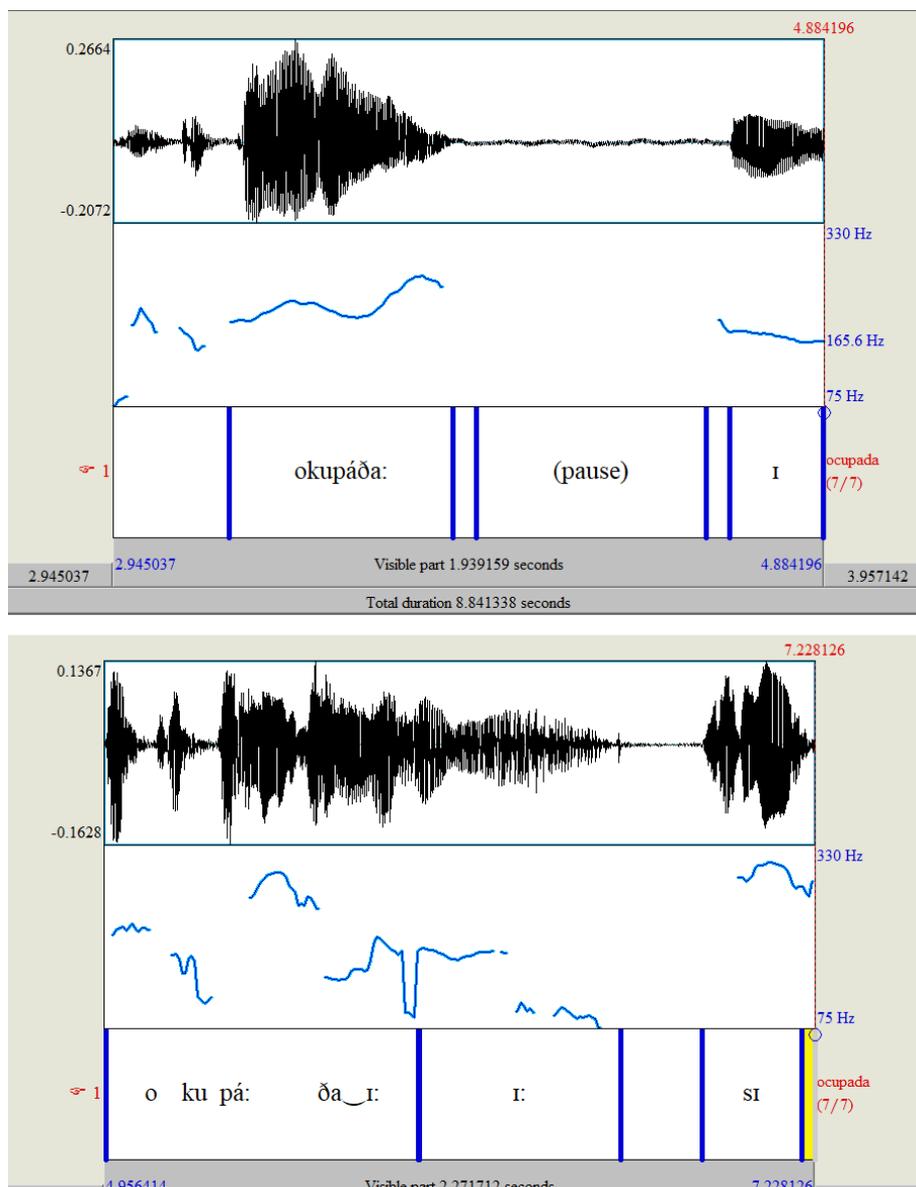
Table 2: Results of the voice analysis for clips extracted from the anger scene.

Clip name	Pitch (Hz) M(SD)/CV	Total no. of pulses	Voice breaks	Elongation of sounds	Vocal fry	No. pulses containing breathy voice
LOOK AT ME						
<i>Play-acted</i>	149(60)/0.40	544	10	2	Yes	0
<i>Natural</i>	120(77)/0.43	566	8	0	No	0
MESSAGES						
<i>Play-acted</i>	186(37)/0.20	437	14	2	Yes	2
<i>Natural</i>	183(46)/0.25	305	10	0	No	0
PARANOID						
<i>Play-acted</i>	188(63)/0.34	780	20	0	No	0
<i>Natural</i>	180(57)/0.32	974	13	0	No	0
SHOUTING						
<i>Play-acted</i>	147(40)/0.27	996	17	0	Yes	3
<i>Natural</i>	176(50)/0.28	1230	19	1	No	0

A closer look at the elongation of sounds reveals that not only are they more numerous in the play-acted than the natural version. They are sometimes accompanied by a gliding vowel or vocalic *liaison* to another vowel sound, which is substituted by a simple pause or unvoiced segment in the natural version, as Figures 1 and 2 show.

In the clip 'BUSY', the character in the original version elongates the sound *\a* in the word 'and'. To synchronize with the character's lips, this elongation is also present in both dubbed versions. The voice actress decided, perhaps, to do this at different points in each dubbed version of the script in an attempt to make the two versions distinguishable. In Figure 1, the natural version (at the top) shows the elongation of the */a/* sound at the end of the word "*ocupada*" (busy) followed by a pause. The sound */i/* from the Spanish conjunction "y" (and) is uttered after the pause. In contrast, the play-acted version (bottom) presents elongation of the *\a* sound at the end of the word "*ocupada*" in diphthong with the conjunction */i/* sound. The sound */i/* continues until the following word in the sentence (*sí*) is uttered. Also, the graph reveals visible differences in terms of pitch contour with more ups and downs in the play-acted than in the natural version.

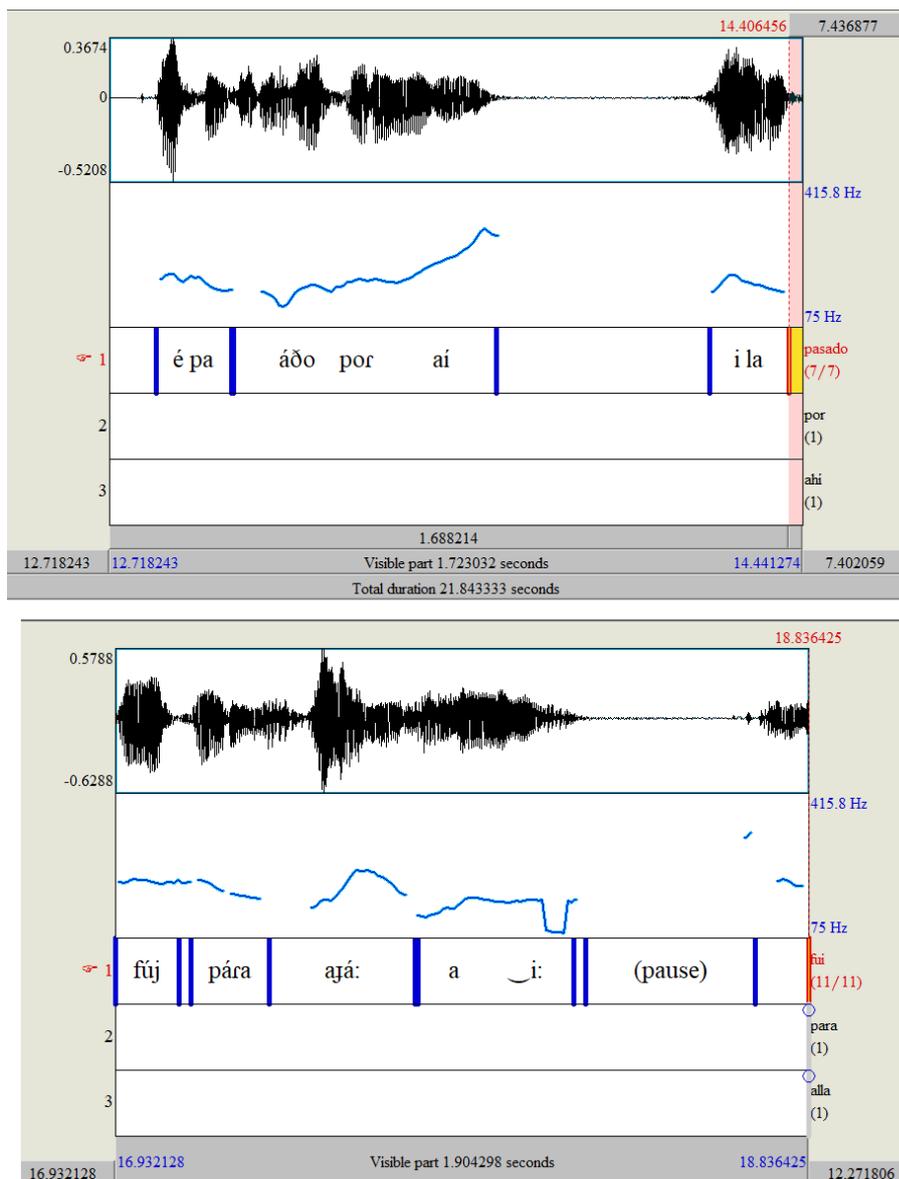
Figure 1: Spectrogram and pitch contour from segment from the sad scene in clip 'BUSY'. The top spectrogram shows an example of elongation followed by a pause in the natural version, whereas on the bottom spectrogram, corresponding to the play-acted version, we can observe the elongation accompanied by a gliding vowel instead.



In the clip 'MESSAGES', a similar pattern is observed again with conjunction "y". In Figure 2, the top graph, corresponding to the natural version, reveals a pause after the word "ahí" (there) is uttered with a rising intonation and then the sentence continues with "y la conversación estaba abierta" ("and your messages were out" in the original). For the play-acted version, however, the graph shows elongation of the final /a/ sound from the word "allá" (there) which is dragged finally transitioning to sound /i/ from conjunction "y". The pause in this case is made after and not before uttering the sound \ ɪ \ and then the sentence is resumed with "los mensajes

estaban en la pantalla" (source-text "your messages were out"). As in the case of Figure 1, we can also notice a more syncopated pitch contour line in the play-acted than in the natural version.

Figure 2: Spectrogram and pitch contour from segment from the anger scene in clip. The top spectrogram shows an example of a short sound with a pause in the natural version, whereas on the bottom spectrogram, corresponding to the play-acted version, we can observe an elongation followed by a gliding vowel.



Finally, as Figure 3 shows, it is also worth pointing out that, whereas in the example of 'BUSY' above, elongation was justified by the attempt to synchronize with the characters' mouth, in

'MESSAGES' the character appears with her back turned to the camera and sound elongation was, therefore, not strictly required:

Figure 3: Screenshots of clips 'BUSY' (left) and 'MESSAGES' (right) at the time point where elongation of sounds with gliding vowels occurs showing the different positions of the character on the screen. Source: American film *Like Crazy* (2011). Frame 00.30.32 (left) and 1.04.38 (right)³.



4.2 Viewers' Perception of Naturalness and Preferences

To find out the effects of play-acted vs. natural style depending on the emotion of the scene (anger vs. sad) -and regardless of the specific clip-, I merged data from all the play-acted versions and natural versions of clips for each scene to conduct the analysis in order to isolate the variable 'naturalness' vs. 'play-acting', regardless of the scene. Moreover, a separate analysis was carried out with 'script' and 'intonation' as dependent variables.

Table 3 shows the results of the statistical analysis I carried out to compare the perception of both versions in terms of script. The data reveals that the difference between the two versions (natural vs play-acted) was statistically significant ($p < 0.001$), with scores for the natural versions being significantly higher (2.95 vs. 2.45 in the anger scene and 2.47 vs. 2.24 in the sad scene), regardless of the group (or scene). Moreover, the effect of the scene was also significant since scores for naturalness were higher in the anger scene than in the sad scene ($p < 0.001$). It is possible that actors found it easier to make anger sound more natural than it was

³ Film directed by Drake Doremus, written by Drake Doremus and Ben York Jones; produced by Jonathan Schwartz and Andrea Sperling, distributed by Paramount Vantage; starring Anton Yelchin, Felicity Jones and Jennifer Lawrence. Production companies: Indian Paintbrush and Super Crispy Entertainment.

to make sadness. However, the effect of the interaction between scene and version was not significant ($p=0.083$), which indicates that there are no differences in terms of perceived naturalness between the natural and play-acted version depending on the scene.

Table 3: Descriptive and inferential statistics for each scene (anger and sadness), comparing perceived differences in the scripts of natural and play-acted versions

	Dubbed version		Tests of between-subject effects					
	Natural	Play-acted	Version		Scene		Scene*Version	
	Mean (SD)	Mean (SD)	F	p (η^2)	F	p (η^2)	F	p (η^2)
Scene								
Anger (N=120)	2.95 (0.87)	2.45 (0.88)						
Sad (N=116)	2.47 (0.86)	2.24 (0.84)	20.83	.000 (0.43)	19.08	.000 (0.39)	3.02	0.083 (0.06)
Total (N=236)	2.71 (0.90)	2.35 (0.86)						

η^2 : Partial eta squared (effect size)

Table 4 shows the comparison of the participants' perception of both versions in terms of intonation. The results obtained again reveal statistically significant differences between the natural and play-acted version, regardless of the scene ($p < 0.001$), with scores for the natural versions being significantly higher (2.83 vs. 2.21 in the anger scene and 2.53 vs. 2.18 in the sad scene). In this case, no effect of the scene ($p=0.056$) or the interaction between scene and version ($p=0.111$) were reported.

Table 4: Descriptive and inferential statistics for each scene (anger and sadness), comparing perceived differences in intonation used in natural and play-acted versions

	Dubbed version		Tests of between-subject effects					
	Natural	Play-acted	Version		Scene		Scene*Version	
	Mean (SD)	Mean (SD)	F	p (η^2)	F	p (η^2)	F	p (η^2)
Scene								

Anger (N=120)	2.83 (0.94)	2.21 (0.95)						
Sad (N=116)	2.53 (0.94)	2.18 (0.86)	33.16	0.000 (0.066)	3.68	0.056 (0.008)	2.55	0.111 (0.005)
<i>Total</i> (N=236)	2.69 (0.95)	2.19 (0.96)						

*eta*²: Partial eta squared (effect size)

After I had assessed the level of naturalness of the two versions for each clip, I asked the respondents to choose one of the two versions (presented as 'scene 1' and 'scene 2'), according to their own preference. In Table 5, I show the percentage of participants who opted for each version. Overall total percentages for each scene reveal a stronger preference of the natural over the play-acted version in both scenes, with 61.67% of respondents for the anger scene and 57.76% for the sad scene. However, if I examine the clips separately, we can observe some differences between them. Whereas respondents in group A chose the natural versions in 3 out of 4 clips in the anger scene ('LOOK AT ME', 'MESSAGES' and 'SHOUTING'), respondents in group B only indicated a preference for the natural version in 2 out of 4 clips ('SHOCKED' and 'BUSY'). Moreover, we can also observe a striking difference in the clips 'PARANOID' and 'COME OVER', where participants clearly preferred the play-acted version. In the first case, this may have been due to a potentially perceived mismatch between the higher intensity with which the line of the male character was uttered and the apparent mild intensity of the emotion reflected in the character's face, giving the impression that the voice was quite angry when the character's facial expression actually showed only slight annoyance. In the case of 'COME OVER', the character gets very emotional and the actress may have found it difficult to reproduce the emotional cues (crying, sobbing) in an identical fashion in both scenes. Despite her efforts, they may have come out as more 'natural' in the play-acted scene, according to the participants' perception.

Table 5: Results of participants' preferences for natural vs. play-acted versions in both scenes

Clip name	Preference for play-acted version		Preference for natural version	
	Participants (N)	%	Participants (N)	%
LOOK AT ME	8/30	26.67	22/30	73.33
MESSAGES	4/30	13.33	26/30	86.66
PARANOID	28/30	93.33	2/30	6.67
SHOUTING	6/30	20	24/30	80

ANGER SCENE (TOTAL)	46/120	38.33	74/120	61.67
SHOCKED	5/29	17.24	24/29	82.76
BUSY	9/29	31.03	20/29	68.70
BYE	16/29	55.17	13/29	44.83
COME OVER	19/29	65.52	10/29	34.48
SAD SCENE (TOTAL)	49/116	42.24	67/116	57.76

Finally, as displayed in Table 6, the respondents were subsequently asked about the reasons for their preference of one version over the other. Credibility of the screen actors was the most frequently mentioned aspect, followed by involvement, with more than 40% and more than 20% of mentions in both scenes, respectively.

Table 6: Percentage of times that participants selected each reason for preference of one version over the other

Clip name	Reasons for preference (% mentions)	
	Anger scene	Sad scene
Identification with characters	12.5	19.4
Immersion in the scene	28.0	23.1
Professionalism of voice actors	12.5	13.4
Credibility of screen actors	47.0	44.1

5. Discussion and Conclusions

This study has attempted to examine the acoustic properties and viewers' perception of different levels of voice naturalness ('natural' vs. play-acted) in emotionally-loaded scenes dubbed into Spanish. I carried out two types of analysis including clips extracted from two previously dubbed scenes (anger scene and sad scene). For the first one I mainly concentrated on the study of the voice by complementing the automatic acoustic analysis of our recordings with direct observation in order to detect other vocal traits present in play-acted speech.

No clear automatically-detected acoustic patterns were observed which could consistently define how play-acted and natural dubbing styles differ. Coefficient of variance did not reveal a higher level of variability in terms of pitch in play-acted versions compared to deliberately 'natural' versions, a result that would have been consistent with previous literature about acted and dubbed speech (Audibert et al. 2010; Jürgens et al. 2011; Sánchez-Mompeán 2017; 2019; 2020). It is possible that a more detailed analysis of individual words or syllables is needed to

detect such fluctuations of pitch. Sánchez-Mompeán (2017), for example, especially observed that this contrast tends to occur between strong and weak syllables (328). Digitally-generated graphs and spectrograms in *Praat* did reveal more variable pitch contour lines in the play-acted versions for the segments analyzed, which coincides with results found in previous literature (Audibert & Rilliard 2010: 4; Jürgens et al. 2011: 5).

Via this study, I was able, however, to observe other vocal features in the recordings that may characterize more accurately the differences between the two dubbing styles. Vocal fry was more frequently present in the play-acted version and absent in the natural version in both scenes. Moreover, the play-acted versions presented more instances of elongation of sounds and number of pulses with breathy voice than the natural versions, especially in the sad scene. More specifically, some instances of elongation of sounds were also found in the form of inter-word diphthongs in the play-acted versions, which were replaced by a pause in the natural version.

The second analysis, in which I focused on the reception of the product by the target audience, consisted in exposing a homogeneous group of viewers to the natural and play-acted versions of both sets of scenes and studying their self-reported perception of naturalness and preferences in both cases. The statistical analysis in this case revealed that scores for perceived naturalness were higher for the natural version than for the play-acted version in terms of both script and intonation, regardless of the scene. However, scripts were perceived as more natural in the anger scene than the sad scene.

As for the viewers' preference, the answers provided by our participants revealed that the natural version was mostly preferred over the play-acted version by approximately 60% of participants in both scenes. Most respondents indicated that the main reason why they preferred one version over the other was that actors had more credibility. Their preference for naturalness was less strong than expected, which can be due to a more extensive presence of play-acted vs. natural voices in the dubbed versions of audiovisual products that they usually watch. Continuous exposure to overacted dubbing styles may enhance suspension of disbelief and prevent viewers from questioning whether the style adopted is more or less similar to spontaneous talk and even lead them to associate overacted styles with professional dubbing.

It is also worth pointing out that a closer look at individual clips reveals that, even if respondents were inclined towards the natural version for the anger scene in most clips, preference for the natural versions in the sad scene was only revealed in half of the clips. A possible explanation for one clip in which the play-acted version was preferred ('COME OVER') is that the naturalness/play-acted style was disguised by the fact that the character is sobbing and crying. The high emotional intensity of the scene in this clip may have interfered in the ability to distinguish between natural and play-acted style both at the production and reception stages. Even if the natural script contained spontaneous traits of orality in Spanish such as *oye* (listen), *es que* (it's just that..) or *en plan* (like), which were purposefully absent in the play-acted version, other vocal features that seem to have characterized the play-acted dubbing style in this study, such as vocal fry or breathy voice, may not have been identifiable on this particular occasion. As a matter of fact, in the second part of my vocal analysis I was not able to find any instances of vocal fry in 'COME OVER' and, although some occurrences of breathy voice were observed in the recording, these may have coincided with the moments in which the character —and the dubbing actress too— was sobbing. Another possibility is that viewers actually prefer to see a more play-acted style in scenes of maximum emotional intensity, as a more dramatic style may seem more suitable in these cases. This idea was actually pointed out in the questionnaire by some of the participants who claimed having preferred the play-acted scene. For example, when asked about the reasons for preferring the play-acted vs. the natural version, participant 19 stated: "because, when there is emotionality in speech, everything is better understood and conveyed in a different way, so it influences whether you like it or not".

Finally, it is necessary to acknowledge the limitations of this study. These results can only be considered as preliminary findings. The size of our sample of recordings is too small to make firm claims about the features of naturalness in dubbing; moreover, the fact that the recordings only contained the voices of two dubbing actors does not allow us to account for the existing variety of vocal features present in dubbed speech. The possibility that the vocal traits observed here are merely a personal preference of the voice-actors involved in this study cannot be ruled out. Another limitation was the changes made in scripts for the natural and play-acted versions, which may have influenced dubbers' predisposition to adopt different suprasegmental attitudes. The same effect may have occurred with our respondents, who may have been conditioned by the differences of the script in terms of orality discourse markers when they were to determine whether intonation was or was not different. In fact, this would confirm Spiteri Miggiani's

(2019) suspicion that, when asking spectators about the naturalness of intonation in dubbed dialogues "the answer would probably consist of puzzled reactions pointing to a lack of discernment" (37), contrarily to what may happen if participants were to compare OV Spanish voices with dubbed dialogues. The author argues that, as long as it does not fall into a parodic effect, dubbed intonation is tolerated by the audience, who are not aware of its distinctive character when it appears in the context of film-watching. Identical scripts should be provided to voice actors in subsequent studies to truly look for differences in the perception of intonation and other vocal features.

At any rate, the study of the interplay between naturalness and emotions seems to be worth pursuing. Follow-up studies including larger and more varied corpora of dubbed recordings alongside a more fine-grained acoustic analysis are needed to verify whether the traits found here can be considered as genuinely distinguishable features of dubbed speech.

Emotions undoubtedly constitute an essential part of fictional films. In fact, one of the main reasons why viewers decide to engage in film watching is to have some exposure to the emotional experience they offer. This study reinforces the evidence that actors' voices play a decisive role in such experience and, more specifically, that different styles of dubbing speech (rather play-acted vs. rather natural) can also make the difference in terms of viewers' preferences, depending on the specific emotions portrayed. Awareness of these types of preferences should possibly be taken into account in professional dubbing settings, especially by the agents involved in this type of decision-making in the studio, since they may also even determine the success of the film once it has been released for the public.

Acknowledgements

This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and FEDER/UE funds (grant number FFI2017-84187-P). This funding source played no role in the collection, analysis and interpretation of data, or the decision to submit the article for publication.

This study has been carried out under the framework of the program "*Ayudas Moving Minds-CMN para la realización de estancias de investigación y formación técnica en centros extranjeros para el PDI y PAS de la Universidad de Murcia para el año 2020*".

References

- Audibert, Nicolas, Véronique Aubergé e and Albert Rilliard (2010) 'Prosodic correlates of acted vs. spontaneous discrimination of expressive speech: A pilot study', Paper presented at *Proceedings of Speech Prosody*, Chicago, 2010. Available online at [<https://pdfs.semanticscholar.org/359c/b9ce6b04a05b9e77a5806300f62797dcc397.pdf>] (accessed 30 September 2020).
- Baños Piñero, Rocío, and Frederic Chaume (2009) 'Prefabricated orality. A challenge in Audiovisual Translation', *inTralinea*. Available online at [<http://www.intralinea.org/specials/article/1714>] (accessed 30 September 2020).
- Bosseaux, Charlotte (2008) 'Translating Britishness in the French versions of Buffy the Vampire Slayer', *Quaderns de Filologia*, 13:85-103.
- Bosseaux, Charlotte (2019) 'Voice in French dubbing: The case of Julianne Moore', *Perspectives: Studies in Translatology*, 27(2): 218-234.
- Bruti, Silvia and Serenella Zanotti (2017) 'Exploring the Sensory Dimension of Translated Films: An Analysis of The King's Speech', *Palimpsestes. Revue de traduction*, 30: 154-177
- Chaume, Frederic (2001) 'La pretendida oralidad de los textos audiovisuales y sus implicaciones en traducción', in Rosa Agost and Frederic Chaume (eds). *La traducción en los medios audiovisuales*, Castellón: Universitat Jaume I, 77-88.
- Gobl, Christer and Ailbhe N. Chasaide (2003) 'The role of voice quality in communicating emotion, mood and attitude', *Speech Communication*, 40:189-212.
- Johnstone, Tom, Carien M. van Reekum, Terrace Oakes and Richard J. Davidson. (2006) 'The voice of emotion: an fMRI study of neural responses to angry and happy vocal expressions', *Social cognitive and affective neuroscience*, 1(3): 242-249 .
- Jürgens, Rebecca, Kurt Hammerschmidt and Julia Fischer (2011) 'Authentic and play-acted vocal emotion expressions reveal acoustic differences', *Frontiers in Psychology*, 2(180): 1-11.
- Juslin, Patrik N., Petri Laukka and Tanja Bänziger (2018) 'The Mirror to Our Soul? Comparisons of Spontaneous and Posed Vocal Expression of Emotion', *Journal of Nonverbal Behavior*, 42: 1-40.
- Khalfa, Stéphanie, Mathieu Roy, Pierre Rainville, Simone Dalla Bella and Isabelle Peretz (2008) 'Role of tempo entrainment in psychophysiological differentiation of happy and sad music?', *International Journal of Psychophysiology*, 68 (1), 17-26
- Klevan, Andrew (2005) *Film Performance*. London: Wallflower press
- Naranjo, Beatriz (2015) 'Translating blackness in Spanish dubbing', *RESLA (Revista Española de Lingüística Aplicada)*, 2(28): 416-441.
- Niculescu, Andrea, Betsy van Dijk, Anton Nijholt, Haizhou Li and Swee L. See (2013) 'Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy', *International Journal of Social Robotics*, 5, 171-191.
- Palencia Villa, Rosa María (2002) *La influencia del doblaje audiovisual en la percepción de los personajes*. Doctoral thesis, unpublished. Barcelona: Universidad Autónoma de Barcelona.

- Pavesi, Maria (1996) 'L'allocuzione nel doppiaggio dall'inglese all'italiano', in Christine Heiss and Rosa María Bollettieri Bosinelli (eds) *Traduzione multimediale per il cinema, la televisione e la scena*, Bologna: Clueb, 117-130.
- Perez-Gonzalez, Luis (2007) 'Appraising dubbed conversation: Systemic functional insights into the construal of naturalness in translated film dialogue', *Translator*, 13(1): 1-38.
- Pietrowicz, Mary, Hasegawa-Johnson, Mark, Karahalios, Karrie (2017): 'Acoustic correlates for perceived effort levels in expressive speech', *The Journal of the Acoustical Society of America*, 142(2): 792-811
- Sánchez-Mompeán, Sofia (2017) *The rendition of English intonation in Spanish dubbing. La naturalidad de la entonación en doblaje y su traducción del inglés al castellano*. Doctoral thesis, unpublished. Murcia: Universidad de Murcia.
- Sánchez-Mompeán, Sofia (2019) 'Prefabricated orality at tone level: Bringing dubbing intonation into the spotlight', *Perspectives. Studies in Translation Theory and Practice*. 28 (2): 284-299.
- Sánchez-Mompeán, Sofia (2020) *The prosody of dubbed speech. Beyond the character's words*, Cham (Switzerland): Palgrave, MacMillan.
- Smith, Susan (2007) 'Voices in Film', In: John Gibbs and Douglas Peye (eds) *Close up 02*, Londond: Wllflower Press, 159-238
- Scherer, Klaus. R. (1995) 'Expression of emotion in voice and music', *Journal of Voice*, 9(3): 235-248.
- Schlinger, Martin (2006) 'Fasten your seatbelts and prick up your ears: the dramatic human voice in film', *Scope*, 5:1-12.
- Scherer Klaus R. (1994) 'Affect bursts', in Stephanie H.M van Goozen, Nanne E. Van de Poll and Joseph A. Sergeant (eds) *Emotions: essays in emotion therapy*, Hillsdale, New Jersey: Erlbaum, 161-196.
- Scherer, Klaus. R. (1995) 'Expression of emotion in voice and music', *Journal of Voice*, 9(3): 235-248.
- Shi, Hongyu., Licai Yang, Lulu Zhao, Zhonghua Su, Xueqin Mao, Li Zang and Chengyu Liu (2017) 'Differences of Heart Rate Variability Between Happiness and Sadness Emotion States: A Pilot Study', *Journal of Medical and Biological Engineering*, 37,527-539 .
- Sobin, Christina and Alpert, Murray (1999) 'Emotion in Speech: The Acoustic Attributes of Fear, Anger, Sadness, and Joy', *Journal of Psycholinguist Research*, 28 (4):347-365.
- Spiteri Miggiani, Giselle (2019) *Dialogue Writing for Dubbing An Insider's Perspective*, Cham, Switzerland: Palgrave Macmillan.
- Whittaker, Tom (2012) 'Locating "la voz": The space and sound of Spanish dubbing', *Journal of Spanish Cultural Studies*, 13(3): 292-305.

Annex I: Scripts for the Sadness and Anger Clips in Spanish

SADNESS			
Clip name	Original script (EN)	Play-acted version (SP)	Natural version (SP)
SHOCKED	I'm OK, yeah. I just, um... I just... It was a bit of a shock.	(OFF) Estoy bien, sí. (ON) Solo ...solo un poco sorprendida.	(OFF) Bien, bien, es que ... (ON) Es que me ha sorprendido un poco.
BUSY	I'm just working really hard and... I'm busy and... Yeah, stuff's going... going well.	(OFF) He estado trabajando mucho y he estado...(ON) amm..ocupada y sí, las cosas van...van bien.	(OFF) Pues trabajando mucho, (ON) y em..ocupada y...bueno, las cosas van...bien.
BYE	OK, bye. Bye, Jacob.	(ON) Vale, adiós, adiós, Jacob.	(ON)Venga, vale, Jacob. Hasta luego.
COME OVER	[sobs] Um... Yeah, do you want to come over? [chuckles] Come over now? And, uh... I'll just be here?	(ON)(Sollozando)¿Quieres venirte? (risa entre lágrimas) ¿Venirte ahora mismo y yo te espero aquí?	(ON)(Sollozando) Oye, ¿te apetecería venirte? (risa entre lágrimas) ¿En plan venirte ahora y yo te espero aquí?
ANGER			
Clip name	Original script (EN)	Play-acted version (SP)	Natural version (SP)
LOOK AT ME	Why can't you look at me and tell me? Look at me and tell me!	(ON) ¿Por qué no puedes mirarme y decírmelo? ¡Mírame y dímelo!	(ON) ¿Por qué no me lo dices a la cara? ¡Que me mires y me lo digas!
MESSAGES OUT	I didn't go through your phone. I walked over there... and your messages are out.	(OFF) No te he cogido el móvil./ Me fui para allá y...los mensajes estaban en la pantalla.	(OFF) No te he cogido el móvil./ He pasado por ahí...y la conversación estaba abierta.
PARANOID	Please don't go through my phone. It's just going to make you paranoid.	(OFF) Por favor, no vuelvas a mirar mi teléfono. (ON) Solo vas a volverte paranoica.	(OFF) Haz el favor de no mirarme más el teléfono, (ON) que luego te entra la paranoia.
SHOUTING	- Why did you just shout at me? - I'm sorry that I shouted.	(ON)-¿Por qué me estás gritando? (OFF)- Siento haberte gritado.	(ON)-¿Por qué me tienes que gritar? (OFF)-Perdona, no quería gritarte.