EDICIONES COMPLUTENSE

Círculo de Lingüística Aplicada a la Comunicación
ISSN-e: 1576-4737

ARTÍCULOS

Towards a methodology for voice quality analysis and character coherence in dubbing

Alfonso Carlos Rodríguez Fernández-Peña
Universidad de Oviedo, Spain

https://dx.doi.org/10.5209/clac.90359 Received: 7 July, 2023 • Accepted: 26 April, 2025

ENG **Abstract:** This paper presents a methodology for voice quality analysis in dubbing and other audiovisual translation modes that include revoicing. The degree of correspondence between source and target voices, namely *character coherence*, is of utmost importance in audiovisual translation if we are to deliver a faithful translation and avoid unnecessary stereotypical cliches that could lead the audience to biased narrative. Although a few attempts at analysing voice quality have been made in audiovisual translation (Bosseaux 2015, 2018, Rodríguez Fernández-Peña 2020), we believe the methodologies proposed are incomplete since fundamental elements, such as the phonation types, voice group, and the fundamental frequency range are not included. Here, we present a methodology that complements the ones suggested by the scholarly tradition and which offers a comprehensive analysis of the actors' voice qualities resorting to the speech analysis software *Praat*, and which is explained in the analysis of Hollywood voices. Voice quality and identity constructions will be considered taking into account disciplines such as psychology, phonetics and phonology, audiology, and musicology. In addition, we will present the different types of phonation according to the most relevant researchers, which, together with other physical elements, are fundamental in order to analyse voice quality and their communicative value.
Keywords: Voice quality; dubbing; character coherence; voice match

How to cite: Rodríguez Fernández-Peña, A. C. (2025). Towards a methodology for voice quality analysis and character coherence in dubbing. *Círculo de Lingüística Aplicada a la Comunicación* 104 (2025): 123-137. https://dx.doi.org/10.5209/clac.90359

Introduction

The power of the voice is unquestionable. Voices carry some sort of genetic imprint that passes from generation to generation showing background personal information. We can raise our voice to impose authority, lower it to convey secrecy, even modulate it to imitate someone else. Everybody with the ability to speak plays with the voice. We can think of a teacher educating children at school, a lawyer making an objection in court, a politician giving a speech, or a doctor giving good news to a patient. Everyone can recall instances of these situations. In cinema, and acting in general, actors play with their voices to make their characters more believable, to make them sound sexy, evil, dumb, naïve, etc. Some actors are famous for having a special voice, for instance, Humphrey Bogart's lisp, James Earl Jones' deep resonant voice, Clint Eastwood's husky voice, Rhea Perlman's high and nasal voice, etc. Actors, with their movements, gestures, and voices, bring their characters to life in cinema. In dubbing, the gestures and movements cannot be and are not altered; however, the voices are always substituted by others in the target languages. These voices do not always show an equivalence to that of the original actor's and therefore, the faithfulness of the translated version could be jeopardized. Within dubbing, the role that voices play has traditionally been kept to a limited number of voice qualities, and favored the character or the expectation of the audience for that character, rather than the original actor's and the 'real' qualities with which the actor endowed it. This could be an issue since it could lead to stereotypical conceptions and lead the audience towards a biased 'reality' which has little to do with the original one. Therefore, we believe that character coherence should try to portray a faithful representation of the original actor and character's voice in the target text.

This paper offers a methodology to study and analyse voice quality, which can be used in audiovisual translation, Film Studies in general, and disciplines such as musicology and phonetics. We will start dealing with the concepts of voice and identity and how we establish mental representations of people depending on how their voices sound. Then we will delve into the different types of phonation and their relevance when identifying voice production. Next, we will explore the preceding studies of voice quality within the realm of Film Studies and the methodologies proposed for voice quality analysis. We will introduce our methodological proposal just next, followed by a voice quality analysis of four Hollywood actors and their Spanish dubbing counterparts. Finally, the results and conclusions sections explain the suitability of our methodology and how it could be used for future research.

1. Voice and identity

Voice refers to the way lung air creates vibration or friction of different types at the vocal folds in the larynx, generating sound that humans transform into speech and language. The human voice is a clear marker of identity in every language (Verhoeven 2002:179, Wells 2006:12, Kreiman and Sidtis 2011:03, Fernández Trinidad 2015:46). For Fasoli et al. (2016:04) "[...] it is how the voice sounds in the listeners' ears that impacts their judgements." The media frequently manipulate the powerful impressions provided by voice quality for different objectives, something crucial to building a product's image in advertising. Luxury automotive commercials use a low tone, breathy quality, and quick tempo to convey intimacy and allure. This voice suggests a mature, dynamic man with power, status, and 'coolness', aligning with the product's market and portraying its owners as strong, seductive, and wealthy. Moreover, documentary films boost their credibility by employing a male narrator whose voice bears the stereotype of a solid, mature, calm, highly intellectual, and dignified authoritative character. In the realm of Film Studies, a good example of how important voice perception is can be found in *The Iron Lady* (2011), which shows former Prime Minister Margaret Thatcher taking voice lessons from a speech therapist to speak more slowly, lower her average pitch and keep her pitch range under control in order not to sound out of control (Setter 2019:75). She had to sound less like a privileged conservative wife and more like the leader she could be, and for that she had to change her voice quality.

This indexical information offered by the human voice may be categorised into two groups according to Rose (2002:284): *extrinsic features*, which are under the control of the speaker, and *intrinsic features*, which are not. Thus, a regional or social accent, which in this context refers to an individual's pronunciation style, is an extrinsic indexical vocal trait. The accent is part of a person's idiolect, which identifies his or her geographical area and/or social status. According to Wells (1982) and Couper-Kuhlen (1986), an individual's pronunciation can also be indicative of the speaker's sex, age group, and educational level. Nevertheless, according to Rose (2002:46), current sociolinguistics holds that individuals do not have a "single invariant linguistic system" but rather employ several versions in sociologically-defined contexts. Therefore, the use of sociolinguistic variables, such as a more 'refined' pronunciation like RP English in a job interview, or a Cockney accent when watching football in a pub, is a matter of speaker's choice and, as such, it is extrinsic. The fundamental characteristics of a voice are the speaker's age, sex, physicality, and health, as well as his or her mental state. According to forensic phonetics, the voice can convey a significant amount of information on the biological state of the speaker. Thus, it is possible to determine if the speaker is healthy, premenstrual, deaf, suffering from a cold, inebriated, or even suicidal (Rose 2002:297).

Rose's (2002) examination of voice indexical information is based on Laver and Trudgill's (1979) typology of identity markers, which is a refinement of Abercrombie's (1967), and specifies three categories to characterise personal vocal traits (1979:03):

a) Group markers: those that indicate social traits such as geographical affiliation, social position, educational status, employment, and social role.
b) Individuating markers: those that mark physical characteristics, such as age, sex, physique and state of health.
c) Affective markers: those that mark psychological characteristics of personality and affective state.

In the light of this, one may see that group markers fall under the category of extrinsic indexical characteristics as defined by Rose (2002), whilst individuating and emotional markers belong under the category of intrinsic indexical features. This scholar provides a model for analysing the many components of a voice, which are comprised of four major parts: two inputs and two processes. The two inputs are labelled *communicative intent*, which maps into the *linguistic mechanism*, and *intrinsic indexical factors* (age, sex, physique, and health), which map onto the *voice mechanism*. The speech wave emanating from the vocal mechanism is the ultimate physical acoustic wave, carrying all the functions and sociolinguistic subtleties that make a voice distinctive.

In line with this, Poyatos (2002), considers that the voice characteristics that differentiate individuals and allow us to recognize them without knowing what they say are: timbre, resonance, intensity or volume, tempo, pitch (level, intervals, range), intonation range, syllabic duration, and rhythm. All these features can be found in Rose's (2002) model within the intrinsic indexical factors and the vocal mechanism (timbre, resonance, intensity, pitch) and the communicative intent (tempo, pitch range, intonation range, syllabic duration, and rhythm).

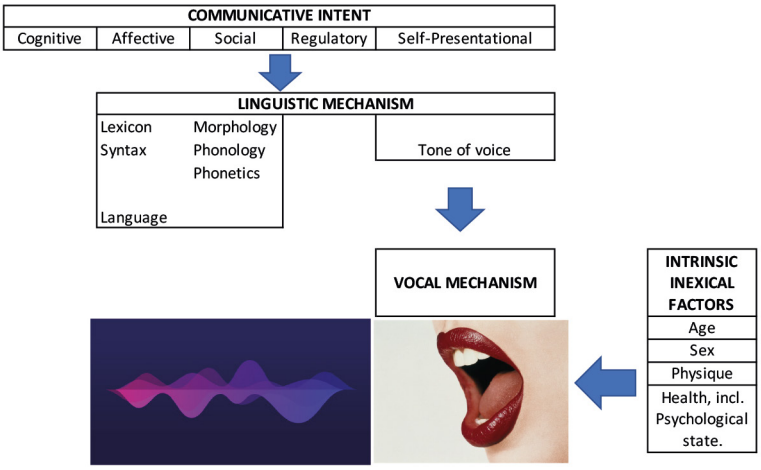
Rose's (2002) model is certainly quite interesting, relevant and useful to identify the communicative value of the messages in an oral source text prior to its translation. While some elements within the linguistic mechanism are provided by the script (lexicon, morphology, syntax), phonology, phonetics, the communicative

intent, and the intrinsic indexical factors highlighted by Poyatos (2002) belong to the voice/dubbing actor in charge of the delivery of the script. All these relate to what Fodor coined as *character synchrony* (1969: 386), for whom:

> There must be certain correspondences between the source and target sound sequences in point of phonetic attributes such as individual timbre, pitch, intensity and speech tempo, peculiarities, that is which are revealed to the spectator by the exterior, temperament and deportment of the actor impersonating the character. If the correspondence is of the right degree we have synchrony in character, if it falls short of a certain level we experience dischrony in character. (1969)

We completely agree with Fodor's view on the degree of correspondence between the original actor and the one providing the translation. However, since there are no time or synchronic requirements, *character coherence*, as suggested by Chaume (2012: 70), is a more accurate term in our view. According to this, and for the sake of faithfulness to the source text, dubbing actors should have a voice that matches (as closely as possible) that of the original actor so that there is no dissonance between the two and to prevent stereotyping procedures and cultural boundaries (Fasoli et al. 2016:04).

Image 1. Rose's (2002) model for voice analysis



2. Voice quality description

Describing voices can be an arduous and challenging endeavor because of interpretation and subjectivity (Bosseaux 2015:12; 2018:225-226), although a wide range of acoustic measures and frameworks have been developed in different disciplines. We could compare it to describing smell and taste; some people could find that some food can taste sweeter or saltier than others, and the same could be applied to voice perception. Depending on our hearing ability we could hear voices higher or lower in pitch, louder or softer, and more or less nasal, for example. Moreover, it can also depend on the listener's ear training or musical education. Therefore, someone with *absolute pitch,* the faculty of recognizing and defining the pitch of a tone without use of a reference tone (Bachem 1955:1180), will not analyse or judge an opera singer as someone with *relative pitch,* the one most people have and which happens when after a reference tone is given, a second tone can be recognized as identical, lower or higher (Bachem 1955:1180), or even as someone with hearing loss. In addition, due to the numerous fields and disciplines that study voice and voice quality (acoustics, biology, phonetics, music, speech science, etc.), both terms arise in a variety of settings. For this reason, their precise usage relies on context and viewpoint, making it impossible to define either term precisely. There are several definitions of voice quality, depending on the focus and interest of each field. For instance, in otorhinolaryngology, voice quality is "a multidimensional construct and cannot be measured monodimensional such as pitch (measured in Hertz Hz) or loudness (measured in decibels dB)" (Barsties and De Bodt 2015:183). This definition contrasts with that of computer science and engineering, where there are systems that evaluate voice quality signals using four parameters: fundamental frequency (F0 measured in Hz), jitter, shimmer (measured in dB), and harmonic-to-noise ratio (Ueng et al. 2012:618). And if we consider the field of phonetics, Abercrombie (1967:91) described voice quality as "those characteristics which are present more or less all the time that a person is talking: it is a quasi-permanent quality running through all the sound that issues from his mouth." Finally, within the realm of audiology, Kreiman and Sidtis (2011) agree with Zemlin (1998) and consider that:

> [...] from the point of view of voice production, quality in its narrow sense can be related to changes in the tension and mass of the vocal folds, to the symmetry of their vibration, to the forcefulness with which they are held together (medial compression), and to the amount of subglottal pressure. Some kinds of changes in the manner of vocal fold vibration do result in reliably perceptible changes in the sound of a voice. (2011:62)

The approach towards voice quality that this work will follow is based on the one used in phonetics and audiology, since we will evaluate voice production and perception. Kreiman et al. (2003:116) find that "the most

common approach to the problem of specifying voice quality is simply to create a long list of terms to describe listener's impressions." For these scholars, ancient descriptive systems incorporated adjectives pertaining to the personality and emotional condition of the speaker (confused, despondent) as well as phrases pertaining to articulation and rhetorical ability (articulate, distinct). Modern dictionaries nowadays include terminology such as *breathy* and *nasal* that are used often in the study of vocal pathology. However, commonalities among traditions greatly exceed differences, and several terms have been in use for centuries as shown in Kreiman and Sidtis (2011:12-13).

Within the realm of Translation Studies there have been some attempts at categorizing voices and analyzing voice quality, such as Ávila (1997), Bosseaux (2015, 2018), Rodríguez Fernández-Peña (2020; 2023), and Imani and Khoshsaligheh (2022). From these, the works of Bosseaux, based on Van Leeuwen's (1999), have been truly revealing. However, we believe that this proposal is incomplete and can lead to some confusion since it mixes phonation types with elements like pitch, vibrato and nasality, which have more to do with singing than with speaking. This paper will present a methodology for voice quality analysis based on Laver (1980, 1994), Kreiman and Sidtis (2011), and Rodríguez Fernández-Peña (2020), which allows the presence of the characteristics within Van Leeuwen's (1999) and Bosseaux's methodology. This new model can provide a more in-depth and comprehensive voice analysis. However, before that, we have to bear in mind a concept that is fundamental for voice analysis: *phonation types*.

2.1. Phonation Types

The changes in the tension and the mass of the vocal folds results in what is commonly known as phonation types. When we speak about phonation, we refer to the interaction between the vocal folds and the egressive pulmonic airstream, which results in a sound source we call voice (Ashby 2011:18). If there is no vibration of the vocal folds, there is no voice. This way we can differentiate voiced and voiceless consonant phonemes like /z/ and /s/, for instance. Depending on the scholar and the discipline, there can be different types of phonation, which can have different combinations. The three main phonation types for a phonetician such as Ashby (2011) are modal (or normal) voice, creaky voice, and breathy voice. Although Ashby (2011) mentioned only three types of phonation, there are other types that can also take place in speaking.

Thus, Laver (1980) found there were eight types of phonation (modal voice, falsetto, whisper, creak, whispery creak, breathy, harsh and ventricular voice), which can be combined with each other in several ways to get a larger number of composite phonatory qualities. Most of those combinations are shown in Esling and Moisik's (2022) taxonomy and can be accessed and heard in the iOS app *iPA Phonetics app* developed by John Esling. For Laver (1980), there are two criteria to classify the phonation types: first, whether they can occur alone as a *simple* type; and second, whether these can occur in combination with other phonation types as a *compound*. The first category is formed by *modal voice* and *falsetto*. These two types can occur alone, and they can combine with other groups as compound types, but they cannot combine with each other. Modal voice is considered a neutral phonation and occurs in ordinary voicing. Sometimes it is referred to as *normal* voice, but that would imply that other phonation types are abnormal and, therefore, most scholars prefer to use the term *modal* rather than *normal*. Modal voice can be subdivided into *head* voice and *chest* voice, the only distinction being resonance properties, "(...) "feeling" the voice predominantly in the chest or in the head." (Gick et al. 2013:113). *Falsetto* is characterised by the high fundamental frequency (F0), which usually is an octave higher than modal voice and is exploited both in singing and normal speaking. Laver (1980:120) finds that falsetto is not exploited for linguistic purposes but, instead, it has a paralinguistic function governed by conventional usage in several cultures. A good example of falsetto can be found in Dustin Hoffman's performance in *Tootsie* (1982).

The second category for Laver (1980) is formed by *whisper* and *creak*, which can happen alone, as simple types, or together as a compound, thus getting *whispery creak*. A whispery voice can be defined as a tense voice with audible friction in the vocal folds. Laver (1980) calls it the "library voice" and acknowledges that its use for paralinguistic purposes is widespread, one example being to signal secrecy or confidentiality (1980: 112). *Creak* is characterized by an extremely low pitch that allows the individual vocal fold vibrations to be heard. What occurs is that the vocal folds vibrate so slowly that each vibration is audible as a pulse of noise. It is also called *vocal fry* or *glottal fry* (Laver 1980:122), and "it sounds like running a stick along a fence." (Biemans 2000:197). According to Setter (2019:83), vocal fry is currently in vogue amongst younger English native speakers since it is used by wealthy, glamourous and upwardly mobile celebrities like the Kardashians in the US and the cast of *Made in Chelsea* in the UK. These three phonation types can be combined with either member of the first group, thus giving *whispery voice* and *whispery falsetto*, *creaky voice* and *creaky falsetto,* and *whispery creaky voice* and *whispery creaky falsetto*.

The third category, for Laver (1980), includes modificatory settings that can only occur in compound types of phonation and never by themselves as simple types: *harshness* and *breathiness*. The combination of phonation and the frication of air at the glottis defines a breathy voice, in which the muscular effort is low, resulting in keeping the glottis somewhat open, resulting in a higher rate of airflow than in modal voice. It can be easily recognized in Marilyn Monroe's singing of "Happy Birthday Mr. President" for J.F.K. in 1962. Harsh voice is usually called *rough* or *throaty* voice and involves an additional compression of the vocal folds. A harsh voice produces for Laver (1980), citing Milisen (1957), "a rasping sound associated with excessive approximation of the vocal folds". If we combine harshness with modal voice and falsetto, we get *harsh voice* and *harsh falsetto*. An example of harsh voice can be found in Max Cavalera, former lead singer of the thrash metal band Sepultura. A raucous voice indeed. However, breathiness can only combine, for Laver (1980:113), with modal voice to give

breathy voice and is incompatible with falsetto. In addition, this scholar states that harshness and breathiness cannot combine with the phonation types from the second category unless a member of the first category, modal voice or falsetto, is also present. Nonetheless, due to the incompatibility between breathiness and falsetto, the only compound types of this category are *harsh whispery voice*, *harsh whispery falsetto*, *harsh creaky voice*, *harsh creaky falsetto*, *harsh whispery creaky voice*, and *harsh whispery creaky falsetto*.

Finally, *ventricular voice*, which Laver (1980:114) considers a subcategory of the harshness setting, is a type of phonation which occurs when the *false vocal folds*, often called the *ventricular folds*, placed above the "true" vocal folds vibrate intentionally along with the true vocal folds, intentionally or unintentionally, thus getting an "extremely harsh voice" (1980:118). However, for Laver (1980) and Gick et al. (2013), the function of the ventricular folds is still not well understood and cannot be considered a type of phonation per se, but a sub-category of harsh voice. We have examples of ventricular voice in traditional inuit singing and beatboxing (Stowell and Plumbley 2008).

Esling and Moisik's (2022) taxonomy of phonation types is the same as Laver's (1980); however, it includes *pitch register* for *harsh voice* (harsh voice high pitch, harsh voice mid pitch, harsh voice low pitch), which, for us, is an extra, as we could also label *modal voice* as high, mid and low in pitch, and *breath* as a phonation type, which according to Laver (1980:112) never occurs as simple type. For this reason, as Laver (1980), we believe there are 20 different phonation types and one sub-category (ventricular voice), shown in table 2 below.

Table 2. Types of phonation and their combinatory possibilities.

Categories / **Phonation Type**		**2nd Category**			**3rd Category**	
		Whisper	**Creak**	**Whispery Creak**	**Harsh**	**Breathy**
1st Category	Modal Voice	✓	✓	✓	✓	✓
	Falsetto	✓	✓	✓	✓	x
3rd Category		Harsh	Harsh	Harsh		
Sub-category	Ventricular Voice					

The 20 possible phonation types shown in table 2 are: modal voice, falsetto, whisper, whispery voice, whispery falsetto, creak, creaky voice, creaky falsetto, whispery creak, whispery creaky voice, whispery creaky falsetto, harsh voice, harsh falsetto, harsh whispery voice, harsh whispery falsetto, harsh creaky voice, harsh creaky falsetto, harsh whispery creaky voice, harsh whispery creaky falsetto, and breathy voice. In addition, we have ventricular voice as a sub-category for harsh voice.

2.2. Methodologies for voice quality analysis within Translation Studies

The study of voice quality in Translation and Film studies has been scarce (Bosseaux 2008, 2012, 2015), and although it is still an under-researched discipline, recent studies have emerged, including Fasoli et al. (2016), or Rodríguez Fernández-Peña's (2020). Despite these new attempts at analyzing voice quality in Translation and Film studies, for us, Bosseaux's (2015, 2018) methodology and proposal is currently the most complete and has been used by other scholars, such as Imani and Khoshsaligheh (2022). Bosseaux's model includes 7 criteria for voice analysis in dubbing, taken from Van Leeuwen (1999), which are: tension, roughness, breathiness, loudness, pitch register, vibrato, and nasality. These criteria can be defined by a series of adjectives, as shown in table 3.

Table 3. Criteria for voice analysis according to Van Leeuwen (1999) and Bosseaux (2015, 2018)

Quality	**Contrasting adjectives**
Tension	Tense/low
Roughness	Rough/smooth
Breathiness	Breathy/Smooth
Loudness	Loud/soft
Pitch register	High/low
Vibrato	Vibrato/plain
Nasality	Nasal/clear

Tension refers to "what happens when you tense the muscles of your throat [...] The voice becomes higher, sharper, brighter, and above all, more tense [...]" (Van Leeuwen 1999:130). When you relax your voice and

expand your throat, the opposite occurs. The voice softens and becomes more mellow. The two suggested adjectives to describe this criterion are "tense" and "low".

Roughness is a characteristic found in voices "[...] in which we can hear other things besides the tone of the voice itself — friction sounds, hoarseness, harshness, rasp." (Van Leeuwen 1999:131). The opposite of a rough voice is a clean, smooth and 'well oiled' sound. The adjectives suggested for this criterion are "rough" and "smooth".

Breathiness occurs when breath is mixed with the tone of the voice itself (Van Leeuwen 1999:133) and for this parameter voices can be described as "breathy" or "smooth".

Loudness has to do with the volume of the voice measured in decibels. People can speak loud or soft depending on the context and their personality.

Pitch register refers to "the scale from the very low 'in the chest' voice to the very high voice (falsetto for men) [...]" (Van Leeuwen 1999:134). Speakers can change the pitch throughout a sentence or conversation to show excitement (high pitch) or boredom (low pitch).

Vibrato, for Van Leeuwen (1999:134), "means what it is. The vibrating sound literally and figuratively *trembles*." Bosseaux (2015) considers that the sound quality of vibrato can be used to describe what Barthes (1977) called the *grain of the voice*, "the materiality of the body speaking its mother tongue" (1977:182), "the body in the voice as it sings, the hand as it writes, the limb as it performs" (1977:188). Vibrato is present when expressing emotions such as love, sadness, even anger.

Nasality occurs when we utter sounds with the soft palate lowered and the mouth open so the air escapes both through the mouth and the nose. A good example of nasal voice is that of Janice in the sitcom *Friends* (1994-2004), whose high pitch and nasal quality characterize and define this character. Voices can then be nasal or clear according to Van Leeuwen's (1999) taxonomy.

Now that we have described the methodology used by Van Leeuwen (1999) and Bosseaux to analyse voice quality, let us comment on the possible problems that we have identified in this model and suggest a new paradigm for voice analysis, which, in our view, can be more complete and provide more information about voice quality when we aim to analyse voices in dubbing and audiovisual translation in general.

The first issue we find in Van Leeuwen's (1999) proposal is the item related to tension, "what happens when you tense the muscles of your throat" (1999:130). Characterizing voices as tense or lax could be confusing, since according to Laver (1980), as highlighted by Gobl and Chasaide (2003), the physiological correlates of voice quality are described in terms of three muscular tension parameters: *adductive tension* (the action of the interarytenoid muscles adducting the arytenoids), *medial compression* (the adductive force on the vocal processes adducting the glottis ligament), and *longitudinal tension* (the tension of the vocal folds themselves). With this in mind, modal voice is characterized by having moderate laryngeal tension, and tense voice is characterized by more tension in the whole vocal tract compared to a neutral voice. In addition, breathy voice is known for little laryngeal tension. As we can see, *tension* is a parameter that can be misleading since there will always be some sort of tension in any phonation type. For this reason, we believe that if we want to analyse voice quality from a phonetic approach, this item should be removed.

The second issue in Van Leeuwen's model has to do with the items of *roughness* and *breathiness*. One could easily define a voice as rough and breathy with this system, when it should rather be 'whispery' as indicated by Laver (1980):

> [...] (I)t would not be possible, for example, to accept a label which combined 'breathiness' and 'harshness', such as *harsh breathy voice*, for the voice often described as 'husky' or 'hoarse', because of the mutually exclusive prerequisites of breathiness as here defined and harshness. Such quality would instead be labelled *harsh whispery voice*. (1980:133)

Since breathiness and harshness are incompatible, we believe that these two items can be misleading if they are shown in a table in which we have to select "rough or smooth" and "breathy or smooth" because one could tick "harsh" and "breathy" when wanting to describe a husky voice, and that, for most phoneticians, would be an error. Moreover, breathiness and harshness are the only phonation types that are shown in Van Leeuwen's and Bosseaux's model, even when these are incompatible. Therefore, if a thorough voice analysis is to be conducted, we suggest that these two criteria be removed and substituted by one single item labelled "phonation type" in which all the different types of phonation can be included.

The third element that can be problematic is the one related to *pitch register*, which has to be judged as "high" or "low". According to Bosseaux (2018:224) *pitch* corresponds to the musical note of a voice and, for Van Leeuwen (1999:134), *register* is the scale that goes from the very low 'in the chest' voice to the very high voice (falsetto for men). We believe that these scholars use *pitch register* here as musicologists do, since in Music Studies, as Cobeta and Mora (2013:522) declare, there are three registers in classical music: *chest register* for lower notes, *medium* for medium notes, and *head* for higher notes (this applies for each voice type). In musicology, singing voices are classified according to pitch register and musical notes as shown in table 4.

As can be seen, each voice type ranges from a lowest note to a highest one, and within that range classical singers can move freely in their performances. Casanova (2013) offers a more detailed voice type classification for singing, including early, classical, and modern music.

Titze (2000:206) offers a classification of speaking voices like that of singing voices, based on the speakers' fundamental frequency (F0), as shown in table 5. This table and speaking voice classification will be used in our voice quality analysis model.

Table 4. Voice Types and pitch register in Musicology (Cobeta and Mora 2013:522)

Voice Type	Lowest note	Frequency	Highest Note	Frequency
Soprano	G3	196 Hz	E6	1175 Hz
Mezzosoprano	M3	165 Hz	A5	880 Hz
Contralto	E3	147 Hz	M5	659 Hz
Tenor	D3	131 Hz	D5	523 Hz
Baritone	G2	98 Hz	G4	392 Hz
Bass	M2	82 Hz	M4	330 Hz

Table 5. Average speaking F0 for voice types according to Titze (2000:206)

Bass	G2 (98.0Hz)
Baritone	B2 (123.5 Hz)
Tenor	E3 (164.8 Hz)
Contralto	F3 (174.6 Hz)
Mezzo-Soprano	G3 (196.0 Hz)
Soprano	B3 (246 Hz)

Apart from *register*, Van Leeuwen (1999) and Bosseaux (2015) also consider *pitch level*, which relates to vocal effort. Therefore, the higher the pitch level, the greater the effort needed and vice versa. This has to do with volume, as stated by Bosseaux (2015:99), for whom "in terms of physical, someone raising their voice usually increases both loudness and pitch". However, *loudness* is one of the seven criteria within Van Leeuwen's (1999) model and has two possible labels: loud, and soft. For this reason, we do not fully understand why these scholars talk differently of *pitch level* on the one hand, and *loudness* on the other, if both things are (as they say) the same.

Finally, Van Leeuwen (1999) and Bosseaux (2015) also consider *pitch range* in a similar fashion as phoneticians such as Cruttenden (2014), Roach (2009), Estebas Vilaplana (2014), and Gómez González and Sánchez Roura (2016) do. Estebas Vilaplana (2014:179) defines pitch range as "the amount of pitch displacement while uttering a sentence" and highlights its importance in translation since pitch differences can prompt misinterpretations. For this scholar, English speakers tend to use a broader pitch range than Spanish speakers (2014:260), and this should be considered during the dubbing process so that the delivery does not sound too excited or monotone. Bosseaux (2008:224) is also aware of this situation when she studied and compared the different voices used to dub Julianne Moore in French.

Cruttenden (2014) and Roach (2009) refer to such a variation in the width of the pitch range as *key*, a term that for Van Leeuwen (1999:207) is "another term for *pitch register*". However, pitch register is defined by Van Leeuwen as "the sound quality we have in mind when we refer to voices, instruments or other sounds as high or low in character" (1999:210). Here, again, we believe that Van Leeuwen is using a musicological view of pitch and key, since from a phonetic point of view, according to Collins and Mees (2013:303) pitch is "the property of a sound (related to frequency) which enables a listener to perceive it as high or low." Although both definitions sound similar, Van Leeuwen's does not mention frequency, which is the way we can measure pitch in acoustic phonetics, and, for us, is paramount when describing voices. As we could see, Titze (2000) and Cobeta and Mora (2013) include the fundamental frequency (F0) in their classifications. F0 is also used by Rodríguez Fernández-Peña (2020) to describe voice quality and analyse voice match in his research of voice-over translation, and we believe it is paramount in voice quality analysis.

For all these reasons, we believe that Van Leeuwen's (1999) and Bosseaux's (2018) criterion of *pitch register* should just be labelled *pitch* and include the F0 of the speaker under analysis and the speaking voice type suggested by Titze (2000). Then, once we know the speaker's voice type, we can look for instances of broad and narrow pitch range to identify how emotions are expressed.

The next item in Van Leeuwen's (1999) and Bosseaux's (2015, 2018) taxonomy that is confusing is *vibrato*, defined as "a sound quality in which the sound has some kind of 'grain' (...) Its semiotic potential derives from the fact that the voice wavers at moments of emotion, whether positive (passion) or negative (fear)" (1999:214). Bosseaux agrees with this definition, adding that "the traces of the physical can be heard in the voice in the form of a vibrato" (2015:106). We agree with these scholars' vision of vibrato as a specific characteristic of certain voices in moments of emotion and tension. Nonetheless, since these moments are not present all the time in films, and are usually sporadic, we reckon this criterion should not be considered as a key element for voice description in dubbing. Kreiman and Sidtis (2011:393) find that vibrato also distinguishes singing from speaking, and it develops "quasy-automatically during singing training, and occurs in popular singing as well as in classically trained voices". In line with this, Cobeta and Mora (2013:521) claim that vibrato is, indeed, a

very common resource in singing voices, used as an ornament in dramatic opera scenes. For these scholars, an opera singing voice without vibrato sounds flat and not attractive since vibrato confers richness and vitality to the voice, differentiating it from psalmodic singing. Again, we can see how Van Leeuwen's and Bosseaux's model is influenced by musicological terms that could mislead a voice quality analysis for dubbing purposes, since vibrato is usually associated to singing rather than speaking.

The final criterion in Van Leeuwen's model is *nasality*, which for this author usually carries negative associations (1999:208). In normal speaking, nasality occurs due to excessive use of the nose as a resonator, usually due to anatomophysiological alterations of the soft palate (Coll 2013:139). Some nasal traits can be found in projected voices, used, for instance, by radio casters and voice actors, who project the voice towards the palate, thus generating a "more voluminous" voice in terms of timbre (Farías 2013:509). This type of nasality is done on purpose, like the one used by professional singers in order to amplify the vocal harmonics (Worsly 2012: 259). Therefore, we can have two types of nasality; involuntary nasality when the physical characteristics of a speaker produce a nasal voice, and a voluntary one when the speaker utters sounds with a nasal touch to provide intensity or amplify the harmonic spectrum. In addition, as Van Leeuwen states, nasality can also be heard in moaning, wailing and screaming (1999:136), and even during a mocking imitation. We believe that nasality is a relevant characteristic to include in voice analysis and it should be kept in our methodology.

As we have seen, Van Leeuwen's (1999) and Bosseaux's (2015) voice quality analysis model can be tricky for various reasons, and we consider that a new proposal for voice quality analysis in dubbing is needed.

3. A new proposal for voice quality analysis in audiovisual translation

In this section we will continue our search for a comprehensive voice quality analysis methodology that can be applied to audiovisual translation modes in which revoicing is present, most notably for dubbing and voice-over. For this purpose, we have found that Kreiman and Sidtis' (2011:68) description of Mel Blanc's vocal characteristics can be a good starting point since it includes the criteria present in Poyatos (2002), Van Leeuwen's (1999) and Bosseaux's (2015, 2018) methodology, together with other items such as F0 range, F0 mean and mode, pronunciation, speaking rate, voice quality (phonation type) and miscellanea. Table 6 shows the voice analysis for some of the Looney Tunes characters voiced by Blanc as per Kreiman and Sidtis (2011:68).

The analysis model proposed by Kreiman and Sidtis (2011) is very interesting and complete since it includes relevant information, such as the F0 range, mode and mean concerning pitch, which is very useful to determine the voice type of the character following Titze's taxonomy (2000); and the voice quality, which has to do with the type of phonation used and allows for all the types (not only harshness and breathiness as in Van Leeuwen's model).

Our model will be based on Kreiman and Sidtis (2011), but it will not include the speaking rate item since we will compare the voices of dubbed and synchronized texts, and therefore, the two voices will have a similar or exact pace. The speaking rate could be included in the analysis of voiced-over texts, in which the original and target voices can be perceived by the audience. In addition, for the analysis of English and Spanish voices in dubbing, counting syllables per second to measure speaking rate could be tricky. English and Spanish are measured differently in terms of rhythm. English is considered a stressed-timed language (Cruttenden 2014:271, Roach 2009:107, Estebas Vilaplana 2014:208, Gómez González and Sánchez Roura 2016:263, Collins and Mees 2013:136) and is measured in feet, while Spanish is a syllable-timed language (Estebas Vilaplana 2014:207, Gómez González and Sánchez Roura 2016:263, Ashby 2011:166) measured in syllables. Consequently, the speaking rate item will be ruled out in our model. However, we can comment on and describe whether the speaking delivery is fast or slow if we find any dissonance between the voices within the pronunciation section. The same applies to the criterion devoted to loudness, which will not be included in our model for voice analysis in dubbing. The target text will offer a voice that has to fulfill not only isochronic purposes but also kinetic ones. Therefore, the voice actor's delivery should be in sync with the gestures and attitude of the person on screen. For Schwarz (2011:400), "an actor's delivery gives rhythm to syllables and emphasizes the important parts of speech". This implies that the loudness used by the two voices (original and translated) will be the same. Nonetheless, if there are instances in which loudness differs, we could note that within the miscellanea section.

There are two items from Poyatos (2002) which will be included in our model: timbre and resonance. *Timbre* is described by Cleveland (1977:1622) as "that particular attribute of a given voice, which distinguishes that voice from another, when the vowel and the pitch are the same." In line with this, Bustamante (2012:83) considers that timbre (or *colour*) is the personal sounding characteristic of each voice both in singing and speaking. For her, it is this *colour* of voice that tells us whether a voice is perceived as velvety, dark, clear, metallic, etc. However, the way of measuring timbre is tricky. Two voices can have the same F0 and, therefore, belong to the same voice type (tenor, soprano, etc.) but sound differently, have a different colour. In the field of psychology, voice-induced synaesthesia leads to multisensory perceptions while hearing someone's voice (Moos et al. 2011), and studies such as Moos et al. (2011, 2013) indicate that light colours are usually attributed to higher F0 and dark colours to lower F0. Fales (2002:58) finds timbre a "slippery concept and a slippery percept, perceptually malleable and difficult to define in precise arranged units". The reason behind the slipperiness of timbre is that it is mainly determined by the harmonic content of a sound. Therefore, as Fales (2002:72) states, "timbre is dominated by several primary harmonics that characterize the tone's timbre". The harmonics of a voice can be graphically established by means of speech analysis software that includes harmonic spectrum capabilities. Rodríguez Fernández-Peña (2020:379) used this system to describe the voices of voiced-over texts following the traditional methodology of voice identification in forensic phonetics, which

Table 6. Voice description pattern of some Looney Tunes characters according to Kreiman and Sidtis (2011: 68)

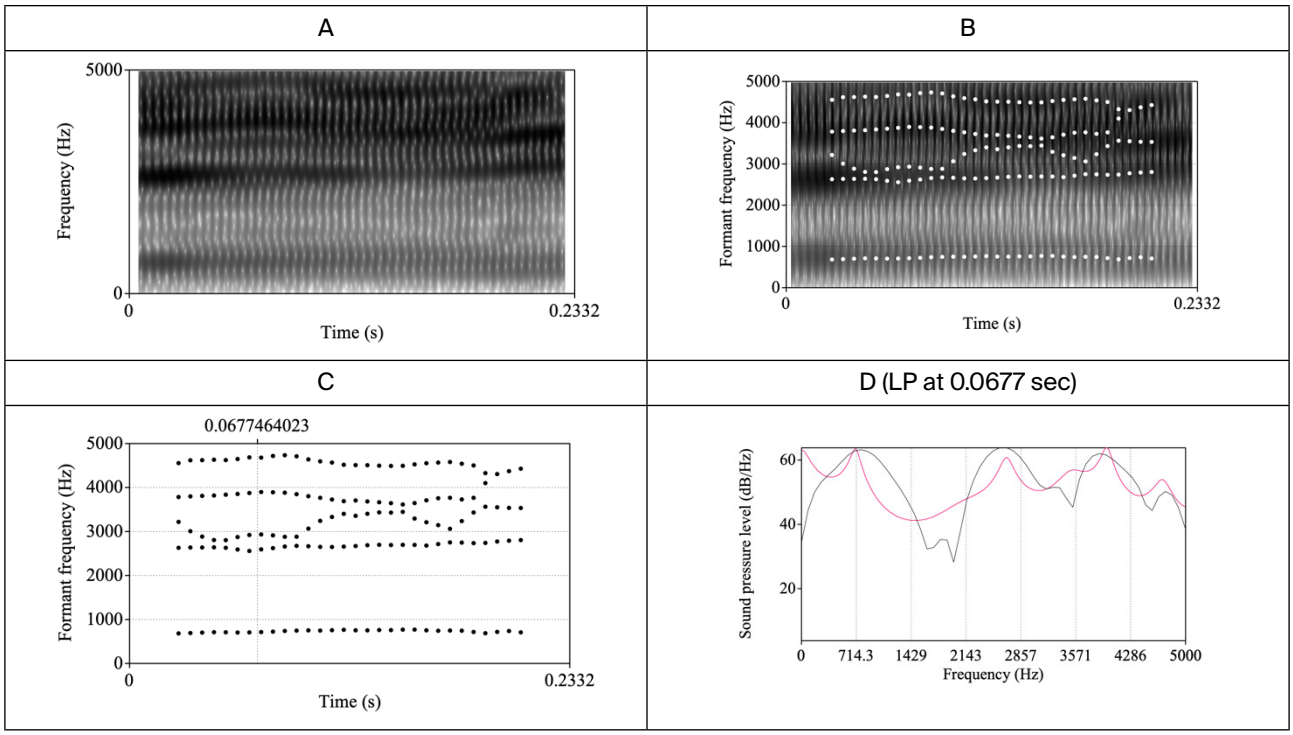
	Mean F0/ Mode F0	**F0 range**	**Loudness**	**Speaking rate (syllables/ sec)**	**Voice quality**	**Pronunciation**	**Miscellanea**
Bugs Bunny	232 Hz / 160 Hz	76-486 Hz	Moderate	3.4	Varies widely. Often uses falsetto, vocal fry; occasional breathiness during imitations.	Variable; often with New York accent.	Imitates many voices. Often uses exaggerated pitch contours.
Daffy Duck	223 Hz / 220 Hz	75-390 Hz	Loud	4.0	Harsh, with vocal fry and turbulent noise	Lisp. Prolongs syllables, /s/ sounds.	Tends to be agitated
Porky Pig	238 Hz / 220 Hz	70-400 Hz	Moderate	5.2	Modal to slightly pressed; no creak or breathiness	Stutters	F0 contours never exaggerated
Taz	Undefined	Wide	Loud	2.8	Rough/harsh, with frequent period doubling and vocal fry.	Snorts, roars, grunts, and other animal noises, with occasional pidgin English.	-
Tweety Bird	323 Hz / 220 Hz and 400 Hz (bimodal)	75-575 Hz	Moderate	4.7	Very high falsetto	“Baby talk”	Excursions into ultra-high falsetto on about half of the utterances.
Yosemite Sam	Low end of male range	Wide	Usually shouting	3.8	Very rough; often period doubled.	Southern accent, extends stressed syllables.	Sometimes uses falsetto
Pepe le Pew	190 Hz / 130 Hz and 235 Hz (bimodal)	50-371 Hz	Soft	3.3	Alternately breathy and vocal fry	French accent.	-

consists in examining aurally and spectrographically similar samples (same word or vowel). This method is currently used by the FBI (VIAAS/IAI protocol) (Rose 2002:111) and we consider it can be useful since it can help us show evidence of the different colour of voices even when the F0 is the same. Needless to say, the FBI voice discrimination model and others used in forensic phonetics use more parameters apart from harmonic distribution to identify speakers. However, in our study, since we already know that the voices belong to two different people, what we want to observe is their difference in timbre, and for that purpose we only focus on harmonic and formant frequency distribution. Therefore, we will indicate the frequency for the first, second, and third major frequencies, which are called *formants* —the resonant frequencies created in the vocal tract as the sound source (the egressive airstream) passes through—. We will extract the values of these formants using the speech analysis software Praat, which allows us to extract the formant contour of utterances by means of a spectrogram and linear prediction (LP) analysis. The relevance of this acoustic structure is regarded by Rose (2002:207), for whom it “not only encodes linguistically relevant pitch and vowel quality, but also carries the imprint of the vocal tract that produced it.” Moreover, although a speaker may never utter exactly the same sound twice, as Latinus and Belin (2011:R143) consider, listeners extract these invariant features in the vocal signal to build representations of a speaker’s identity. For this reason, we believe the colour of the voice, represented in formant frequency distribution, is significant in voice quality analysis of dubbing voices.

Table 7 shows four images that graphically describe the phoneme / ɪ / extracted from the word *Bing*, uttered by Maggie Wheeler in one episode of *Friends*. Image 7A shows a spectrogram of that utterance, which consists in “a two-dimensional plot of frequency versus time, with amplitude shown by the darkness of the trace” (Rose 2002:223). As can be seen, the horizontal axis shows the time (0.2332 seconds), and the uttered phonemes are distributed along the time axis as they occur. The horizontal axis shows the frequency (in Hz) and the amplitude of those particular frequencies which is represented by the intensity or darkness of each point in the image. We can identify at least three horizontal bands of dark frequencies, which belong

to the first three formants, the second one being the largest. These are displayed on image 7B by means of dotted lines across the spectrogram. Here we can see that for the second formant, there are two sets of dotted lines. Image 7C shows the formant contour on an empty spectrogram. There are 4 formants, and we have set the cursor at 0.067 sec. to measure them since it seems the point at which they are most stable. F1 (formant 1) occurs at around 700 Hz, F2 just over 2500 Hz, and F3 almost at 4000 Hz. We have conducted a linear prediction (LP) analysis, which is a well-established signal processing technique to digitally determine the formant center frequency, at 0.0677 seconds of the utterance, and the result is shown in image 7D. We can observe the harmonic structure (in black), with a line that shows a contour of four rounded peaks, and the linear prediction spectrum (in magenta) which resolves the major peaks that belong to F1, F2, F3 and F4. We can appreciate that F1 happens at around 710 Hz, as estimated by eye, F2 at around 2700 Hz, F3 at around 3900Hz, and F4 at around 4700 Hz. This proves that our formant analysis by eye from the spectrogram was right, and that we can use this system to get the formant frequencies from the samples selected for study.

Table 7. Spectrographic analysis



The second item from Poyatos (2002) which will be included in our model has to do with *resonance*, which can be pharyngeal, oral, or nasal, depending on where the vocal folds vibrations find the largest resonator based on the size and form of the pharyngeal, oral and nasal cavities. Voices that are very oral, are known also as resonant, strong and rich, easily associated to rather large body size if the speaker is not seen; they suggest positive male's characteristics like masculinity, energy, good health, resourcefulness, etc., and some similar ones are also attributed to women. The opposite could be labelled as *thin* voices. If pharyngeal resonance is understood as 'throatiness', it reflects good characteristics in males (e.g., maturity, old age) and negative characteristics in women (e.g., mannish, unemotional). Nonetheless, persistent nasal resonance does not appear to be indicative of any beneficial physical or psychological traits in anyone.

Our model will include the following sections:

- F0 mean: This will give the average mean of the speakers' fundamental frequency to measure pitch. With this information, we can categorize the voices following Titze's (2000) model. We will use Praat, a speech analysis software, in order to get this information.
- F0 range: this information will provide information about the lowest and highest peaks concerning pitch.
- Voice type: Once we get the F0, we can classify the voices following Titze's (2000) taxonomy: bass (98 Hz), baritone (123.5 Hz), tenor (164.8 Hz), contralto (174.6 Hz), mezzo-soprano (196.0 Hz), soprano (246.9 Hz).
- Timbre / Formants: here we will show the frequency of the first three formants (F1, F2, F3). These values will be obtained analyzing same words or same vowel phonemes in both texts using Praat.
- Voice quality: this section will describe the phonation types used by the speakers following Laver's classification.
- Resonance: whether the voice is oral, nasal, or pharyngeal.
- Pronunciation: here we will describe the pronunciation issues or characteristics of the speakers. Do they stutter or speak with a lisp? Do they speak with a foreign or regional accent? Is the voice nasal? Are there instances of vibrato?

– Miscellanea: this section will include all the aspects that do not fit within the previous categories. Here we can resort to Poyatos' (2002:25-26) paralinguistic voice qualifiers or voice types and use some of the impressionistic labels to complement our description. Besides, we can add other information worth commenting on. Do the voices imitate other voices? Are the imitations similar? Are there instances of *dubbitis* in the Spanish text? According to Sánchez Mompeán (2020:148), dubbitis can appear in two ways: first, as a monotonous, flat and aloof melody triggered by the frequent repetition of low pitch contours; and second, as a more overacted and over-involved melody, motivated by a characteristic rhythmicality and tension, with irregular variations in loudness and tempo.

We believe that these categories and the information they provide will offer a detailed voice quality analysis with which we will be able to compare character voices in dubbing and other revoicing translation modes. Table 8 provides the grid with the criteria suggested and some of the possibilities for each of them, and it will be used in the analysis provided in section 4.

Table 8. Voice quality assessment rubric for revoicing translation modes in audiovisual translation

	F0 Mean	F0 Range	Voice Type	Timbre Formants	Voice Quality	Resonance	Pronunciation	Miscellanea
Actor			Bass Baritone Tenor Contralto Mezzo-soprano Soprano		Modal Falsetto Whisper Whispery voice Whispery Falsetto Creak Creaky voice Creaky falsetto Whispery creak Whispery creaky voice Whispery creaky falsetto Harsh voice Harsh falsetto Harsh whispery voice Harsh whispery falsetto Harsh creaky voice Harsh creaky falsetto Harsh whispery creaky voice Harsh whispery creaky falsetto Breathy voice	Oral Nasal Pharyngeal Laryngeal Chest	Stutter Lisp Foreign accent Vibrato Rhotacism	Screeching Hollow Bleating Moaning Overarticulated Strident

4. Voice quality analysis

In this section we will analyse the voice quality of four actors and characters from four well-known films and series following our suggested model: Humphrey Bogart's in *Casablanca* (1942), Clint Eastwood's in *Million-dollar baby* (2004), Maggie Wheeler's (Janice) in *Friends* (1994-2004) (season 1, episode 14), and Scarlett Johansson's in *Match Point* (2005). The reason behind the election of these four voices is that they belong to movie celebrities whose distinct and characteristic voices are well known to the general English-speaking public. The Spanish dubbing actors are José Guardiola (Bogart), Constantino Romero (Eastwood), María Moscardó (Wheeler) and Marta Bárbara (Johansson). From the films under study, we have selected clips with no background music so that the audio is clean, and the analysis is more accurate. In addition, in order to extract the formants, we have isolated similar phonemes from some words left untranslated in the target text which are pronounced similarly in both languages (mainly character names in the films). Thus, for *Casablanca* the formants were extracted from /ɪ/ in "Victor"; for *Million Dollar Baby* from /ʌ/ in "Sully"; for *Friends* from /ɪ/ in "Bing"; and for *Match Point* from /e/ in "Eleanor".

Analysis 1 shows the voice qualities of Humphrey Bogart and José Guardiola. As can be seen, Guardiola's tone of voice (F0) and voice range is lower than Bogart's, making him a bass, while Bogart could be categorized as a baritone following Titze's model. The phonation type used by both actors is modal and with nasal resonance, although in the case of Guardiola it is also somewhat throaty as we can hear vocal fold vibration in the lower frequencies. In terms of pronunciation, Bogart speaks with a lisp due to the lip scar he

had. In addition, Bogart barely moves his mouth while speaking, keeping a stiff upper lip almost all the time. Guardiola's pronunciation is relaxed and well-articulated. In terms of timbre and colour we could say that Guardiola's voice is dark, while Bogart's is quite clear, probably due to the nasal resonance.

Analysis 2 depicts the voices of Scarlett Johansson and her Spanish dubbing actor Marta Bárbara. We can see that both voices differ in terms of F0 and range, making Johansson a mezzo-soprano (203 Hz) and Bárbara a contralto (176 Hz). We can hear Johansson's characteristic harsh whispery phonation, which is compensated by Bárbara with a modal to breathy phonation, trying to imitate the original delivery. However, Johansson's resonance is pharyngeal, and we can hear some degree of hoarseness and some voice cracks, while Bárbara's voice is clean, oral and there is no sign of laryngeal fatigue. This makes both voices have a different colour, Johansson's being darker in our view. It is worth considering that during this scene Johansson was smoking, something which Bárbara did not do while recording her lines in the studio.

Analysis 3 portrays the voice differences of Clint Eastwood and Constantino Romero. The main difference that can be perceived when we listen to both voices is the type of phonation. Eastwood has a harsh whispery voice, which makes it sound husky and dysphonic. On the other hand, Romero's voice is modal, round and well projected. In addition, both differ in terms of F0 and range, Romero's voice being lower and deeper. In terms of colour, Romero's voice is dark, and Eastwood's is pale/gray due to his characteristic dysphonia.

Analysis 4 shows the analysis of two actors using a voice that is not their everyday voice, but one created ex profeso for the character in the series (Janice). Maggie Wheeler is the original actor and she created Janice's voice that way to endow the character with a sense of humor and to make it sound annoying at times. This made Janice one of the most recognizable and imitated characters of the series. As shown in the analysis both voices are similar in terms of phonation type and resonance (nasal). In addition, they are fairly high-pitched, Moscardó's being even higher (soprano) than the original (mezzo-soprano), probably trying to make the character sound even more annoying when speaking. Both voices have a similar colour since the Spanish voice provides an imitation of the original. Moscardó's voice is high and nasal, just as Wheeler's, and mimics the intonation and cadence of the original actor, which results in a similar voice perception.

Table 8. Voice quality analysis of Humphrey Bogart (José Guardiola), Scarlett Johansson (Marta Bárbara), Clint Eastwood (Constantino Romero), and Maggie Wheeler (María Moscardó)

	Actor	F0 Mean	F0 Range	Voice Type	Timbre-Formants	Voice Quality	Resonance	Pronunciation	Miscellanea
1	**Humphrey Bogart**	137.328 Hz	86-171 Hz	Baritone	F1 500 Hz F2 1700 Hz F3 3500 Hz	Modal voice	Nasal	Speaks with a lisp.	Speaks with a stiff upper lip. Does not gesticulate much.
	José Guardiola	108.729 Hz	75-145 Hz	Bass	F1 400 Hz F2 2000 Hz F3 3300 Hz	Modal	A bit nasal and throaty	relaxed	We can clearly hear the low frequencies from the throat. Dark colour of voice.
2	**Scarlett Johansson**	203.983 Hz	85-278 Hz	Mezzo-Soprano	F1 900 Hz F2 3010 Hz F3 4400 Hz	Harsh whispery voice	Pharyngeal	relaxed	Smokes during scene. Voice sounds husky, hoarse at times. It cracks at certain points.
	Marta Bárbara	176.553 Hz	117-367 Hz	Contralto	F1 600 Hz F2 2200 Hz F3 3000 Hz	Modal to breathy voice	Oral	relaxed	No room noise — sound booth
3	**Clint Eastwood**	163.403 Hz	103-494 Hz	Tenor	F1 700 Hz F2 2100 Hz F3 3050 Hz	Harsh whispery voice	Pharyngeal	Elision of /h/ phoneme in "him" Tense articulation	Husky voice / Dysphonic Some room noise in recording setting
	Constantino Romero	133.322 Hz	74-498 Hz	Baritone	F1 560 Hz F2 1300 Hz F3 2400 Hz	Modal voice	Oral	Tense articulation	No room noise — sound booth Clean, resonant projected voice.
4	**Maggie Wheeler**	225.901 Hz	219-236 Hz	Mezzo-Soprano	F1 710 Hz F2 2700 Hz F3 3900 Hz	Modal voice	Nasal	She has a peculiar way of speaking. Her intonation is characteristic. Elongates some vowels.	The pitch and laugh are so high that can be annoying — that is characteristic of the character. She gesticulates a lot while speaking.
	María Moscardó	298.688 Hz	223-328 Hz	Soprano	F1 400 Hz F2 3300 Hz F3 4050 Hz	Modal voice	Nasal	Imitates original intonation.	Uses a nasal high pitch similar to that of the original actor. Projected voice.

5. Conclusions

In this study we have presented the reader with a methodology to analyse voice quality in a comprehensive way. We have delved into previous studies within the realm of Film Studies that are aimed at analyzing voice quality and found that these could be complemented with items present in other disciplines such as phonetics, forensics, psychology and musicology. We have covered relevant features of voice quality, as stated by the numerous works cited, such as phonation types, fundamental frequency and range, formant distribution, etc., which have not been included in any voice analysis within the realm of Film Studies and Translation before. In addition, we have used speech analysis software to prove our findings and showed graphical evidence of them with spectrographic and LP analysis. The result is, for us, an easy-to-use tool that could be complemented with other elements, and which we believe can be helpful in future research concerning voice analysis and voice matching for audiovisual texts such as films, musicals, series, or documentaries.

The relevance that voices play in cinema is undeniable. We associate voices with actors, movie or series characters (even cartoons), and each voice remains in our minds echoing every time we think of that actor or character. Choosing the right voice for a dubbed version of a movie seems to be a hard task and there may be several factors in play. Sometimes, casting directors may decide to use a similar voice that imitates the original English one, as we have seen in the examples of the two female actors. María Moscardó perfectly imitates Janice's original tone of voice and manner of speaking providing a faithful representation of the original character. Moreover, Marta Bárbara, with her low voice, uses a breathy phonation trying to achieve the husky imprint of Scarlett Johansson's voice. On the other hand, as in the case of the two male protagonists, the voices are chosen according to what is expected of the character, even if that is a cliché. We have seen that the Spanish voices of Humphrey Bogart and Clint Eastwood are both low and resonant, usually associated to dandy-macho characters (Ávila 1997:122). However, the English original characters do not sound that way even if their characters behave as male classical heroes. This brings the question of what voice is more suitable for the translated text: should target voices match the original ones? Or should these voices match what is expected of the characters they portray? If we wish to achieve character coherence, we have to deal with these two options: respect the original actor's voice type and find a match in the target language or find a voice that favors the expectation of the target audience (even if that voice differs significantly from the original actor's). It would be interesting to research the behaviour of casting directors from a diachronic perspective to see how voice quality selection and character coherence has evolved over the years. With this work we present a methodological framework for analysis hoping that future researchers will use it in forthcoming studies.

6. References

Abercrombie, David (1967). *Elements of General Phonetics*, Edinburgh, Edinburgh University.

Ashby, Patricia (2011). *Understanding Phonetics*. Understanding Language Series. Hodder Education.

Ávila, Alejandro (1997). El doblaje. Vol. 46. Ediciones Catedra.

Bachem, Albert (1955). "Absolute pitch." The Journal of the Acoustical Society of America 27, no. 6 (1955): 1180-1185. (https://doi.org/10.1121/1.1908155).

Barsties, Ben, & De Bodt, Marc (2015). "Assessment of voice quality: current state-of-the-art." *Auris Nasus Larynx 42*, no. 3: 183-188. (DOI:10.1016/j.anl.2014.11.001).

Barthes, Roland (1977). *Image — Music — Text*. Glasgow: Fontana Press.

Biemans, Monique (2000). *Gender variation in voice quality*. Utrecht: Lot. Utrecht. The Netherlands.

Bosseaux, Charlotte (2008). "Buffy the Vampire Slayer Characterization in the Musical Episode of the Tv Series." *The Translator 14*, 343-72. (https://doi.org/10.1080/13556509.2008.10799262).

Bosseaux, Charlotte (2012). "Some Like It Dubbed: Translating Marilyn Monroe*".* In Helen Julia Minors (Ed.), *Music Text and Translation* (117-137). Bloomsbury Publishing Plc.

Bosseaux, Charlotte (2015). *Dubbing, film and performance: Uncanny encounters*. Peter Lang Publishing Group. (DOI:10.3726/978-3-0353-0737-5).

Bosseaux, Charlotte (2018). "Voice in French dubbing: the case of Julianne Moore." *Perspectives 27*:2, 218-234. (https://doi.org/10.1080/0907676X.2018.1452275).

Bustamante, Carmen (2012). "El cantante lírico". In Inés Sánchez Bustos (Ed.), *La voz: la técnica y la expresión* (76-97). Paidotribo.

Casanova, Cori (2013). "Características de la voz cantada y estilos vocales". In Ignacio Cobeta, Faustino Núñez, & Secundino Fernández (Eds.), *Patología de la Voz* (525-533). Barcelona: Marge Médica Books.

Chaume, Frederic (2012). *Audiovisual Translation: Dubbing.* St Jerome Pub. (https://doi.org/10.4324/9781003161660).

Cleveland, Thomas F. (1977). "Acoustic properties of voice timbre types and their influence on voice classification." *The Journal of the Acoustical Society of America 61*, no. 6: 1622-1629. (https://doi.org/10.1121/1.381438).

Cobeta, Marco, & Mora, Elena (2013). "Fisiología de la voz cantada". In Ignacio Cobeta, Faustino Núñez, & Secundino Fernández (Eds.), *Patología de la Voz* (516-524). Barcelona: Marge Médica Books.

Coll, Roxana (2013). "Valoración logopédica del paciente disfónico". In Ignacio Cobeta, Faustino Núñez, & Secundino Fernández (Eds.), *Patología de la voz* (525-533). Barcelona: Marge Médica Books.

Collins, Beverley, & Mees, Inger Margrethe (2013). *Practical Phonetics and Phonology: A Resource Book for Students*. Routledge.

Couper-Kuhlen, Elizabeth (1986). *An Introduction to English Prosody*. Edward Arnold.

Cruttenden, Alan (2014). *Gimson's Pronunciation of English*. 8th ed. London: Routledge.

Esling, John H., & Moisik, Scott R. (2022) "Voice Quality." In Rachael-Anne Knight, & Jane Setter (Eds.), *The Cambridge Handbook of Phonetics* (237-58). Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press.

Fales, Cornelia (2002). "The paradox of timbre." *Ethnomusicology 46*, no. 1: 56-95. (https://doi.org/10.2307/852808).

Farías, Patricia (2013). "Voz hablada", In Ignacio Cobeta, Faustino Núñez, & Secundino Fernández (Eds.), *Patología de la Voz* (500-515). Barcelona: Marge Médica Books.

Fasoli, Fabio, Mazzurega, Mara, & Sulpizio, Simone (2016). "When Characters Impact on Dubbing: The Role of Sexual Stereotypes on Voice Actor/Actress' Preferences." *Media Psychology 20*, 450-476. (https://doi.org/10.1080/15213269.2016.1202840).

Fernández Trinidad, Marianela (2015). "La percepción de la cualidad de voz y los estereotipos vocales." *Revista Española De Lingüística 45* Fasc. 1 2015 45-72. (http://revista.sel.edu.es/index.php/revista/article/view/91).

Fodor, István (1969). "Linguistic and psychological problems of film synchronization (part II)". *Acta Linguistica Academiae Scientiarum Hungaricae*, Vol. 19, No. 1/2 (1969), pp. 69-106.

Gick, Bryan, Wilson, Ian, & Derrick, Donald (2013). *Articulatory Phonetics*. Malden MA: John Wiley & Sons. (https://www.wiley.com/en-us/Articulatory+Phonetics-p-9781118438084).

Gobl, Christer, Ní, Chasaide, & Ailbhe (2003). "The role of voice quality in communicating emotion, mood and attitude". *Speech Communication*, *40,* Issues 1-2, Pages 189-212. (https://doi.org/10.1016/S0167-6393(02)00082-1).

Gómez González, María de los Ángeles, & Sánchez Roura, Teresa (2016). *English Pronunciation for Speakers of Spanish: From Theory to Practice*. Berlin, Boston: De Gruyter Mouton. (https://doi.org/10.1515/9781501510977).

Imani, Sima, & Khoshsaligheh, Masood (2022). "Voice Quality and Characterization in Dubbing: The Case of Sherlock in Persian Dubbing." *Language and Translation Studies (LTS) 55*, no. 2: 1-30.

Kreiman, Jody, Vanlancker-Sidtis, Diana, & Gerratt, Bruce R. (2003). "Defining and measuring voice quality." In *ISCA Tutorial and Research Workshop on Voice Quality: Functions, Analysis and Synthesis*.

Kreiman, Jody, & Sidtis, Diana (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. John Wiley & Sons. (DOI:10.1002/9781444395068).

Latinus, Marianne, & Belin, Pascal (2011). "Human voice perception." *Current Biology 21*, no. 4: R143-R145. (10.1016/j.cub.2010.12.033).

Laver, John (1994). *Principles of Phonetics*. Cambridge University Press. (https://doi.org/10.1017/CBO9781139166621).

Laver, John, & Trudgill, Peter (1979). "Phonetic and linguistic markers in speech." In Klaus Rainer Scherer, & Howard Giles (Eds.), *Social markers in speech* (1-32). Cambridge: Cambridge University Press.

Laver, John (1980). "The phonetic description of voice quality." *Cambridge Studies in Linguistics 31*: 1-186.

Milisen, Robert (1957). "Methods of evaluation and diagnosis of speech disorders." In Travis Lee Edward (Ed.), *Handbook of speech pathology* (234-258). Appleton Century Crofts.

Moos, Anja, Simmons, David, & Smith, Rachel (2011). "Voice Quality Perceptions by Synaesthetes, Phoneticians and Controls." In *ICPhS*, pp. 1426-1429. (http://www.internationalphoneticassociation.org/icphsproceedings/ICPhS2011/OnlineProceedings/RegularSession/Moos/Moos.pdf).

Moos, Anja, Simmons, David, Simner, Julia, & Smith, Rachel (2013). "Color and texture associations in voice-induced synesthesia." *Frontiers in Psychology 4*: 568. (https://doi.org/10.3389/fpsyg.2013.00568).

Poyatos, Fernando (2002). *Nonverbal communication across disciplines. Paralanguage, kinesics, silence, personal and environmental interaction*. J. Benjamins Publishing Company. (https://doi.org/10.1075/z.ncad2).

Roach, Peter (2009). *English Phonetics and Phonology. A Practical Course*. 4th Edition. Cambridge University Press.

Rodríguez Fernández-Peña, Alfonso Carlos (2020). Translating intonation and prosody in English-Spanish voice-over: strategies, resources and professional practices. [Unpublished PhD thesis]. University of Oviedo. (https://digibuo.uniovi.es/dspace/handle/10651/57698).

Rodríguez Fernández-Peña, Alfonso Carlos (2023). "An analysis of the Spanish dubbing of the video game Iron Man 2". Digital Translation: International Journal of Translation and Localization, 10 (1), 58-87.

Rose, Phil (2002). *Forensic speaker identification*. Taylor & Francis (https://doi.org/10.1201/9780203166369).

Sánchez-Mompeán, Sofía (2020). *The Prosody of Dubbed Speech*. Springer International Publishing. (https://link.springer.com/book/10.1007/978-3-030-35521-0)

Schwarz, Barbara (2011). "Translation for dubbing and voice-Over." In Kirsten Malmjaer, & Kevin Windle (Eds.), *The Oxford handbook of translation studies* (394-409). Oxford University Press.

Setter, Jane (2019). *Your Voice Speaks Volumes: It's Not What You Say, But How You Say It*. Oxford University Press.

Stowell, Dan, & Plumbey, Mark D. (2008). "Characteristics of the beatboxing vocal style." *Dept. of Electronic Engineering, Queen Mary, University of London, Technical Report, Centre for Digital Music C4DMTR-08-01*.

Titze, Ingo R. (2000). *Principles of voice production*. The University of Iowa and the Denver Center for the Performing Arts. National Center for Voice and Speech Iowa City IA.

Ueng, Shy-Kuang, Luo, Cheng-Ming, Tsai, T. Y., & Chang, H. (2012). "Voice quality assessment and visualization." In *2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, pp. 618-623. IEEE. (10.1109/CISIS.2012.62).

Van Leeuwen, Theo (1999). *Speech, music, sound*. Bloomsbury Publishing. (https://link.springer.com/book/10.1007/978-1-349-27700-1).

Verhoeven, Jo (2002). "The communicative setting and markers in speech." In Jo Verhoeven (Ed.) *Phonetic work in progress,* (pp. 177-190). Antwerp: University of Antwerp.

Estebas Vilaplana, Eva (2014). *Teach Yourself English pronunciation: An Interactive Course for Spanish Speakers*. Netbiblo.
Wells, John Christopher (2006). *English Intonation. An Introduction*. Cambridge University Press.
Wells, John Christopher (1982). *Accents of English: Volume 1*. Cambridge University Press.
Worsly, Selina (2012). "Voz, vibración y armonía. Descubriendo y explorando la voz terapéutica". In Inés Sánchez Bustos (Ed.) *La voz: la técnica y la expresión* (246-269). Paidotribo.
Zemlin, Willard R. (1998). *Speech and hearing science: Anatomy and physiology*. Needham Heights, Massachusetts. Allyn & Bacon.