Parallèles 36(2) October 2024

Quality assessment tools for studio and AI-generated dubs and voice-overs

Giselle Spiteri Miggiani
University of Malta

Abstract

This paper proposes a quality assessment model designed for dubs and voice-overs, applicable to both studio recordings and AI-generated output. Drawing on a prior quality assessment proposal narrowed down to script adaptation (Spiteri Miggiani, 2022a), this paper introduces a broader model that includes an additional rubric to assess the overall quality of dubbed and voice-over output. The quality rating of the end product is determined by evaluating and assigning individual scores to a set of comprehensive quality indicators categorized into two main components: speech and sound. In contrast, the dubbing script is evaluated using the textual parameters rubric developed previously, which adopts a granular, error-based approach and combines a formula to calculate a percentage score. The newly revised quality assessment model thus enables a comprehensive or macro evaluation of a dubbed product from a viewer's perspective. Additionally, it provides another tool focused on textual parameters for a more detailed micro examination from the perspective of linguists and adapters. These tools have broad applications and account for recent AI advancements in dubbing and media localization. The model is intended for dubbing practitioners, trainers, evaluators, recruiters, dubbing managers, quality control specialists, and software developers interested in creating dubbing-related tools or enhancing localization management platforms with quality control features.

Keywords

Quality assessment, quality control, dubbing, voice-over, script, speech-and-sound, speech-and-sound post-editing, studio dubs, AI-dubs

1. Adjusting to recent developments in the field

This paper applies to various revoicing modalities: lip-sync dubbing, phrase-sync or lector dubbing, voice-over or UN-style voice-over, and voice-over narration. Lip-sync dubbing demands 'full' synchronization, meaning all synchronies must be respected, including 1) isochrony (Chaume, 2012) or timing (Spiteri Miggiani, 2021a, 2021b), which refers to the cue-in and cue-out of the utterances and duration; 2) phonetic sync or lip sync (Chaume 2012); 3) rhythmic sync or tempo (Spiteri Miggiani 2019, 2021a, 2021b), which pertains to internal speech tempo, pace, pauses, all of which determine mouth flap recurrences; 4) kinesics (Chaume, 2012), referring to the correspondence between the target language utterances and on-screen body language; and, 5) general semiotic cohesion, including synchronous and semantic correspondence between visuals and speech. Conversely, phrase-sync dubbing currently requires synchronization in terms of timing and rhythm, but not phonetic sync. Voice-over dubbing or UN-style voice-over approximates isochrony or timing with deliberate delayed or anticipated cues, in any case demanding a certain level of accuracy, while other synchronies are not essential, except for long pauses. Voice-over narration requires an approximation in terms of timing and duration to ensure semiotic cohesion. The differences between modalities also include the extent to which the original voice tracks are audible, whether completely muted or faintly heard in the background.

Dubbing is a complex mode of translation characterized by several constraints. Its effectiveness relies on the viewer's suspension of disbelief toward the mode itself, making credibility key. This credibility depends on a few factors, such as habituation to dubbing (Zabalbeascoa, 1993; Spiteri Miggiani, 2021a, 2021b; Sanchez Mompeán, 2023), but also on the quality of the overall dubbed and voice-over output, which is the core focus of this article. The multiple constraints managed by dubbing practitioners have led to script adaptation being considered 'constrained writing' (Mayoral et al. 1988; Titford 1982). The main challenge lies in molding the verbal text to match the visible mouth movements on screen, condensing or amplifying the utterances to match the timing and duration, creating natural pauses in the target language that coincide with those in the source text, and maneuvering syntax as well as target language equivalents to align with the body language. And this is just addressing the synchronization issues without delving into cultural, territorial, and other linguistic aspects.

The shift to cloud dubbing is intended to mitigate some challenges, facilitate the work of the professionals involved, and streamline processes and communication (Chaume & de los Reyes-Lozano, 2021). Cloud dubbing refers to an end-to-end workflow production and management system hosting the entire dubbing process on a cloud-based centralized platform: from script origination, script translation and adaptation to casting, auditioning, and recording. In some companies, the script adaptation process is still managed independently, sometimes using different software tools, while the cloud-based platform is mainly used for the recording process and can facilitate remote recording. When recording is done at home workstations rather than studios in other territories, different quality considerations must be made compared to traditional in-studio recording because of the non-studio recording environments and equipment and their implications. In this paper, the term 'studio dubs' or 'studio recordings' will encompass cloud-based recording, with the main distinction being drawn between human-generated and machine-generated dubs.

Regarding script adaptation, when hosted on a cloud-based platform, script adapters can rely on tools such as the rythmo band, along with its rhythmic cues and lip-sync markers, to facilitate the text synchronization process. This allows them to focus on the linguistic aspects, potentially improving the quality of the script. Merging the translator and adapter roles, as required by cloud-based platforms, can also impact the quality of the translation. Training approaches and skill requirements differ, and these platforms can also facilitate quality control (QC) processes by integrating user-friendly tools to flag and review errors.

The migration to cloud-based platforms is not the only ongoing shift occurring in the dubbing industry that can affect the quality of the outcome. Recent advancements in AI technologies address the main synchronization challenges by reversing the traditional approach. Instead of relying on word-to-lip adaptation techniques, these technologies facilitate lip-to-word adaptation through specific tools and algorithms (Spiteri Miggiani, 2022b). In other words, the lip movements in the visuals are synthesized to match the translated audio. Examples of such tools include TrueSync by Flawless (https://www.flawlessai.com/truesync), the large-scale multilingual audiovisual dubbing tools developed by Google's DeepMind researchers (Yang et al. 2020, https://deepmind.google/), and Lenseup (https://www.lenseup.com/en/), among others. In contrast, Lipdub (https://www.lipdub.ai/) allows the recording process to be handled by voice talents and studios, then artificially modifies the lip movements to achieve phonetic sync. This means script adapters are relieved from the effort required to achieve one type of synchronization, while still catering to others. The sound mixing and editing of the newly adapted visual content also fall into the hands of sound technicians.

Apart from word-to-lip maneuvering, software developers have focused efforts on voice cloning, transcription, translation, and, in some cases, adaptation. An example is Dubly (https://dubly.ai/), which, at the time of writing, clones the original voice, transcribes and translates the speech utterances, and then, rather than artificially modifying the lip movements, adapts the text based on a sync algorithm that analyses the length and timing of the original audio. It then provides clients (and their adapters) with the ability to edit the script adaptation and regenerate the merged video and audio file accordingly. It is important to note that all the above-mentioned examples are a continuous work in progress seeking further development and enhancement.

Considering the shifts brought about by this AI revolution and the incorporation of such tools to facilitate the dubbing process (to various extents), it seems necessary, if not urgent, to reconsider the notion of quality and identify new challenges and issues that may arise, along with their impact on the overall quality of dubbing or voice-over. Currently, the decision-making, observational, critical, and analytical processes in defining quality, and controlling and determining the parameters, heavily depend on scholars and practitioners. To this end, the revisited quality assessment model proposed in this paper considers parameters specific to AI-generated output while including parameters common to both studio and AI processes. Ultimately, they share a common goal: ensuring satisfaction of the product, and this is determined by practitioners and viewers.

1.1. Aims

Drawing on a previous quality assessment proposal narrowed down to script adaptation (Spiteri Miggiani, 2022a), this paper provides a broader model that seeks to evaluate the quality of the overall dubbed or voice-over output. The ultimate objective is to provide an additional tool for practitioners, stakeholders, and trainers to measure the overall quality, and pinpoint, identify, and label any issues or glitches, and intervene as necessary. The previously developed script-focused rubric based on textual parameters is integrated as a separate analytic tool within the same model. The following sections delve into quality assessment in media localization and specifically dubbing, outline the existing textual parameters rubric and measuring system, and then present the revisited quality assessment model with its newly developed rubric. The section before the conclusions focuses on its application and integration into professional workflows, tools, and training contexts.

2. Quality control and quality assessment in media localization

QC in dubbing and media localization is increasingly becoming a crucial part of professional workflows. This is also evident from the fact that some stakeholders publish their QC processes and vendor expectations online (Netflix, 2024). A typical in-sutdio dubbing or voice-over workflow involves several steps, as illustrated in Figure 1. These include:

1) Preparation of working materials. This involves retrieving or preparing the original post-production scripts or dialogue lists or transcription, pivot language translation (if applicable), and video;
2) Translation and adaptation of the speech or dialogue. This may involve two separate professionals, one for translation and another for adaptation;
3) Linguistic QC, when feasible. This step ensures that the translated dialogue is accurate and culturally appropriate, respects the original creative intent, and adequately caters to the synchronies;
4) Loop or take segmentation (if applicable). This involves dividing the dialogue into smaller segments for recording purposes;
5) Voice casting: This involves selecting the appropriate voice actors (often via an auditioning process);
6) Recording: This is the process of recording the speech utterances;
7) Mixing: This involves combining the recorded dialogue with the original audio and music and effects track;
8) Technical and linguistic QC: This ensures that the dubbed content meets the required technical and linguistic standards.

This is followed by further adjustments and editing before the deliverable is finalized and sent to the client. While this workflow is an ideal scenario, in reality, a linguistic QC is not always conducted before recording due to various reasons, including the challenging task of evaluating the synchronization aspects of an adapted text without the actual recording. This will more likely lead to further audio editing, reviewing, and retouching of the dubbed content, which may require re-recording specific segments in the case of traditional in-studio dubbing. Workflows vary depending on the companies. For instance, a typical workflow is one in which the linguistic and technical QC are performed by two separate professionals, with differentiating skills, especially if the technical QC requires direct intervention on the audio tracks. In such cases, the same practitioner can perform the technical QC across multiple languages. Another workflow may require the same practitioner to perform both linguistic and technical QC in a specific language.

The availability of quality assessment tools can play a significant role in facilitating the integration of QC in various phases of the workflow. These tools can be implemented as features in cloud-based platforms or as manual tasks in more traditional processes, as will be discussed briefly in Section 4.

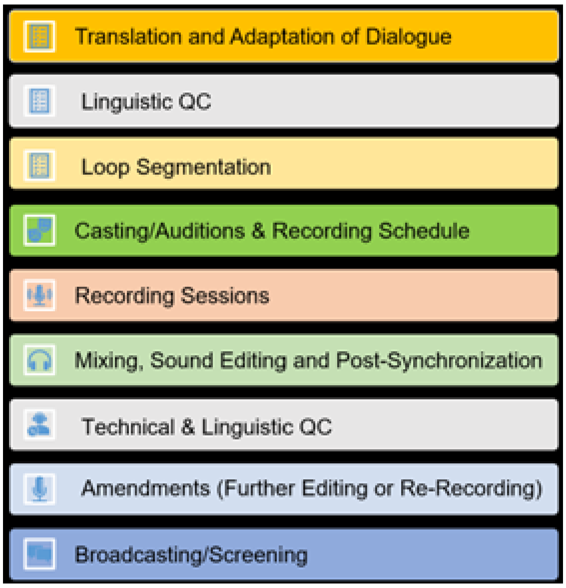

Figure 1. Dubbing and Voice-Over Workflow. Source: Author

Academic research has introduced various quality-related tools for trainers in the profession. While some tools are originally designed for professional use, they can also be applied within university or corporate training environments. They serve as valuable resources for trainers to evaluate and provide feedback, as well as for trainees to engage in self-assessment and improvement. Bolaños García-Escribano (2025) addresses assessment in audiovisual translation specifically for educational settings by proposing a model that embraces all modalities. The landscape of subtitling quality assessment is diverse, with various models catering to different needs and perspectives. The FAR model (Pedersen, 2017) adopts a product-oriented approach, focusing on the viewer's experience. It employs an error-based assessment method, evaluating functional equivalence, acceptability, and readability. By assigning penalty points to errors, the FAR model generates a final score reflecting overall quality. This approach provides clear guidelines for evaluation, ensuring consistency and objectivity. In contrast, the NER model (Romero-Fresco & Martínez Pérez, 2015) caters specifically to intralingual live subtitling. It focuses on error detection, analyzing the number of words, editing errors, and recognition errors. Errors are classified as minor, standard, or serious, providing a straightforward assessment suitable for the fast-paced nature of live situations. The NTR model (Romero-Fresco & Pöchhacker, 2017), designed for interlingual live subtitling, also employs error detection. It considers the number of words, translation errors, and recognition errors, classifying errors as minor, major, or critical. This model incorporates translation quality into the assessment, ensuring accuracy and faithfulness to the original content. The CIA model (Künzli, 2017, 2021) takes a unique approach by incorporating the perspective of professional subtitlers. It focuses on interlingual subtitling and relies on subjective assessment. The model evaluates correspondence, intelligibility, and authenticity, assigning maximum scores to each dimension and deducting penalty points for errors. This approach acknowledges the expertise of subtitlers and emphasizes the importance of a flowing viewing experience.

Conversely, there are fewer quality assessment tools specific to dubbing. While some academic courses have their evaluation rubrics, such as those applied at Universitat Jaume I and Universitat de València, in Spain, the need for a standardized assessment rubric prompted the development of the Textual Parameters quality assessment model (TP model) (Spiteri Miggiani, 2022a) focusing on script translation and adaptation. This model combines an assessment rubric with an error-based formula to identify and categorize errors, provide tailored feedback, and calculate evaluation scores. It aligns with universally accepted quality standards, focusing on aspects such as lip synchronization, natural dialogue, coherence with visuals, faithfulness to the source text, pleasant phonaesthetics, and script functionality. Some scholars have delved into specific quality parameters in dubbing within training contexts, emphasizing skills like synchronization (Chaume, 2007, 2008) and natural dialogue delivery (Baños, 2021) as pivotal for viewer immersion. Different perspectives exist regarding the prioritization of quality standards, ranging from prioritizing a realistic oral register over lip synchrony (Martínez Sierra, 2008, p. 58, drawing on Whitman-Linsen, 1992, p. 55; Chaume, 2012, pp. 85–86, drawing on Caillé 1960, p. 107), to focusing on phonetic equivalence over semantic or pragmatic equivalence in the case of close-up shots (Chaume 2012, p. 74). The TP model assigns equal importance to all quality parameters to ensure objectivity but also provides the option to differentiate between minor and major errors in the evaluation process.

2.1. The Textual Parameters Model

The TP model was designed exclusively for script adaptation and has been integrated as a separate analytic tool within the enhanced and expanded dubbing and voice-over quality model proposed in this paper. It condenses six core error categories based on established textual quality standards. These categories, which focus on both the process and end product, offer a comprehensive framework for assessing adaptation quality:

1. Synchronization: This category encompasses technical issues related to timing, tempo, and lip-sync accuracy. Precise timing and matching lip movements are crucial for maintaining viewer immersion and avoiding jarring discrepancies.
2. Language: This category evaluates the naturalness of the adapted dialogue, while also focusing on grammar, vocabulary, style, and register. It also focuses on smooth flow, cohesion between dialogue events, and clear comprehension. This category also includes source language interference issues, particularly those arising from the ‘dubbese’ register. It is distinct from the Translation category as it can be assessed solely through the target language, without the need to reference the original text.
3. Visuals and Sound: This category examines the cohesion between the target language words and the visuals on screen, including body language, and the retained original soundtrack. Inconsistencies can disrupt the viewer’s understanding and engagement.
4. Translation: This category focuses on the accuracy and fidelity of the translation, identifying mistranslations, unnecessary omissions or additions, awkward phrasing, and undue non-inclusive or overly sensitive language use.
5. Phonaesthetics: This category assesses the euphony of the dubbed dialogue, avoiding cacophonic utterances, excessive repetition, or unwanted rhyme that could detract from the listening experience.
6. Script Functionality: This category delves into process-oriented issues encountered during post-production script processing. These include practical issues that can disrupt the dubbing workflow, such as formatting inconsistencies, missing dialogue, or wrong character attribution. Orthography is also included in this category because errors in spelling or writing are typically not detectable by dubbing viewers. However, they can disrupt actors during the recording process, thus representing a functional issue rather than a linguistic one in this modality.

Textual parameters	**Generic Error Tag**	**Error category**	**Specific Error Tag**	**Error specifics**
Adequate lip synchronization	[S]	Synchronization	[…]	Too short
			[--]	Too long
			[R]	Rhythmic issues (mouth flaps mismatch)
			[L]	Labial consonants mismatch
			[V]	Vowels or semivowels mismatch
Natural-sounding language	[L]	Language	[GR]	Incorrect grammar
			[SC]	Source calque
			[REG]	Unsuitable register
			[COMP]	Lack of clarity & comprehension
			[NAT]	Lack of naturalness
			[FLOW]	Lack of flow & cohesive dialogue exchanges
Semiotic cohesion	[VS]	Visuals & Sound	[VIS]	Lack of cohesion between words & visuals (such as body language)
			[SND]	Lack of cohesion between words & sound belonging to the original audio track (music & effects, lyrics, noise)
Fidelity to source text	[T]	Translation	[MIS]	Mistranslation
			[OM]	Unnecessary omission
			[ADD]	Unnecessary addition
			[LOSS]	Unnecessary loss (semantic)
			[AWK]	Awkward rendering
			[IMP]	Improper translation (such as undue non-inclusive, offensive or derogatory terms that are not functional to the plot or characterisation)
Phonaesthetics	[PH]	Phonaesthetics	[CAC]	Cacophonic utterances
			[REP]	Annoying repetition
			[RHY]	Unintended rhyme
Script functionality	[F]	Functionality	[CON]	Lack of consistency (non-compliance with glossary sheets; inconsistent use of names/ nicknames, forms of address & terminology within the same script or across serial production scripts)
			[REAC]	Missing or wrong reaction
			[NOT]	Missing or wrong notation
			[/]	Missing pause marker
			[FOR]	Layout or format issues
			[DS]	Unsuitable dialogue segmentation
			[OR]	Orthography mistakes
			[CH]	Wrong character allocation
			[D-?]	Missing or redundant dialogue
			[B-?]	Missing or inadequate background walla
			[PUN]	Misleading punctuation
			[TC]	Missing or wrong time code
			[G/P]	Non-compliance with guidelines & policies
			[PRON]	Tricky articulation or pronunciation
			[MISC]	Miscellaneous

Table 1. Script Rubric

These main error categories are broken down further into 37 error specifics, as shown in Table 1. Every error category and error specific has a tag for ease of use during an evaluation or review process. These tags can be used to flag a specific issue by inserting them in a specific point within the text, therefore indicating the exact issue and location. The rubric can then be used as a legenda to interpret the tags. The evaluator or QC specialist can review a script by adopting the 6 generic error categories and generic tags, without delving into the error specifics within each category. In other words, this rubric offers two possibilities: a simplified and more detailed variant, depending on the level of granularity required.

During the evaluation process, the individual errors (tagged as generic or specific) are quantified and the total number of errors is then incorporated into a formula to calculate a percentage score: S% = 100 – (E/W)*100, where S is the total percentage score indicating quality levels; E is the total number of errors; W is the total number of words in the source text sample. This basic formula gives equal weight to each error. Variations on the formula are possible and these can consider different levels of severity, by flagging errors as major or minor (or major or critical, or whichever marked distinction is preferred), and also varying levels of difficulty of the texts in question. In this case, the formula can integrate these elements as follows: S% = 100 – [(Emaj*3 +Emin)*O/W]*100, where S is the total percentage score indicating quality levels; Emaj is the number of major errors; Emin is the number of minor errors; W is the total number of words; O is the error 'offset', a parameter which varies according to the degree of difficulty of the text; O will be taken as a number between 0.5 and 1, based on 3 degrees of source text difficulty: Low: O = 1, Medium: O = 0.75, High: O = 0.5. These formulas and examples of their application are explained and exemplified in further detail in a previous article (Spiteri Miggiani 2022a). Tables 2 and 3 illustrate how errors can be quantified and tagged to calculate a percentage score based on the total number of errors while flagging the issues to address.

Adapted dialogue	Percentage score
FARSHID (OFF-ON) Here it is. / The wonder of wonders. // I remember the first time I heard you play. [S] (gasps) [F] I'd never heard anything like it. Really, it was [S]… When I think about it, I want to cry. [S]/ I… [F] Do you see this tear running? [S] / Just there. Do you see?	Total no. of words: 57 (W) Total number of errors: 6 (E) S% = 100 – (E/W)*100 **89.5%** = 100 – (E/W)*100

Table 2. Error tagging and quantification applying the 6 generic categories

Adapted dialogue	Percentage score
FARSHID (OFF-ON) Here it is. / The wonder of wonders. // I remember the first time I heard you play. [...] (gasps) [/] I'd never heard anything like it. Really, it was [...]… When I think about it, I want to cry. [L]/ I… [/] Do you see this tear running? [...] / Just there. Do you see?	Total no. of words: 57 (W) Total number of errors: 6 (E) **S%** = 100 – (E/W)*100 **89.5%** = 100 – (6/57)*100

Table 3. Error tagging and quantification applying the 37 error specifics

This model, implemented in professional and training settings (Spiteri Miggiani, 2023; Spiteri Miggiani, 2024), yielded valuable insights into error patterns among established and trainee translators. Analysis of the results revealed that synchronization and script functionality emerged as recurring error categories across both groups. This suggests that these areas pose particular challenges for translators, regardless of their level of experience. Further research is ongoing to provide a more detailed analysis of these error patterns and identify potential interventions to improve accuracy and efficiency in script adaptation for dubbing.

That said, the advancements and rapid changes in the field outlined earlier necessitate the expansion of the model to provide a more comprehensive, versatile, and rapid tool. This broader model evaluates not only individual script adaptation elements but also the overall effectiveness and coherence of the final dubbed product and is based on a wider range of quality parameters.

3. Script, Speech and Sound Quality Assessment Model

The revisited quality assessment model can be referred to as the 'Script, Speech and Sound (SSS) Model' reflecting its all-encompassing approach to evaluating dubbing quality through its three main components, which are used to categorize the key performance indicators. The model can also be referred to as a QC model since it applies to the final product. That said, the term 'assessment' will be adopted in this proposal, considering the thorough evaluation process through comprehensive analysis, the possibility of providing feedback, the script evaluation process that can occur during the production cycle, the rating and score calculations, and the different contexts in which the model can be applied, especially corporate training and university settings.

The model encompasses two analytic rubrics:

1) the newly developed Speech-and-Sound Rubric that offers a comprehensive evaluation of the entire dubbed product, including a high-level assessment of the script's impact on overall quality;
2) the already established Script Rubric (or Textual Parameters rubric) illustrated in Table 1 that provides a detailed and focused evaluation of the script, specifically targeting error analysis. Integrating the Textual Parameters rubric into the broader dub and voice-over quality assessment model prompts the adoption of the simpler term Script Rubric to clearly differentiate it from the Speech-and-Sound Rubric.

As a result, the model now includes an additional rubric that covers speech and sound components, allowing for the evaluation of the dubbed version of the target language, after it has been recorded/produced by humans or generated by machine. This addition facilitates a comprehensive, 'macro' evaluation of the dubbed product from a viewer's perspective. This evaluation is intended to be performed by QC specialists, supervisors, or, in some cases, dubbing managers, or other designated individuals within the workflow. A possible process-oriented solution prior to recorded output is also discussed later, along with the potential integration of this rubric into dubbing workflows.

The Speech-and-Sound Rubric employs an acceptability score system of 0 to 10, evaluating various quality indicators through rapid assessment due to its 'perceived quality' approach. This approach provides a quick and efficient overview of the overall quality without requiring a comparative analysis with the source version. In contrast, the Script Rubric delves deeper with a granular, error-based approach, requiring a thorough review of the entire script against the original product. This more in-depth 'micro' evaluation allows for precise identification and quantification of errors, facilitating targeted improvements. While the Script Rubric's error quantification offers a degree of objectivity, the Speech-and-Sound Rubric may benefit from multiple raters to reduce subjectivity, which is one of the main limitations of the assessment process. Additionally, while random sample-checking can be sufficient for the Speech-and-

Sound Rubric, a comprehensive script review is essential for the Script Rubric's accuracy. The model can be applied to a project using either one or both rubrics, with the latter being preferred when the script-related parameters in the broader rubric are insufficient and a more detailed script assessment is needed.

In the Speech-and-Sound Rubric, the speech component encompasses those parameters related to the voice, vocal delivery, and synchrony with the visuals, including those dependent on the script and technical factors, but ultimately conveyed through the actor's performance and speech output in general. The sound component focuses on the cohesive whole at a technical audio level, emphasizing the quality of the recording itself and the roles of mixing and editing to shape the final product's overall quality. These could make or break the product, potentially undoing all the effort and standards achieved through the other parameters (Spiteri Miggiani 2021). Table 4 summarises the main differences between the Script Rubric and the Speech-and-Sound Rubric.

ASPECT	SPEECH-AND SOUND RUBRIC	SCRIPT RUBRIC
FOCUS AREA	Entire dub or voice-over product, including speech, sound, and high-level script assessment	Textual parameters, language, translation/adaptation accuracy
EVALUATION CRITERIA	15 quality indicators covering performance, voice, sound, synchronies, dialogue	6 categories and 37 specific errors focusing on the script; can apply two levels of severity and three degrees of difficulty
SCORING SYSTEM	0–10 scale for each criterion; overall average score calculated.	Error quantification; formula calculating percentage score based on errors vs. word count
ASSESSMENT APPROACH	Analytic, comprehensive evaluation of various aspects separately; perceived quality; viewer-centred	Analytic, detailed analysis of translation/adaptation and error-specific impact; quantification of errors; linguist-adapter centred
OUTCOME	Overall quality score for dubbed product.	Percentage score for script quality and accuracy
UTILITY	QC tool and metrics; performance comparison; identification of issues; recommended action	Possibility to give feedback; specific error flagging; translator ranking & recruitment; self-revision tool
WORKFLOW	Applied after recording or machine-generated output; prior to finalization and delivery	Can be implemented before or after recording or machine-generated output; ideally applied before
EFFECTIVENESS	Provides accurate evaluation; less cost-effective	Could offer less accuracy; more challenging; more cost-effective if applied before output
PROFILE	Technical and linguistic QC specialist	Linguistic QC specialist with technical adaptation skills

Table 4. Script Rubric versus Speech-and-Sound Rubric

3.1. Speech-and-Sound Rubric

Table 5 illustrates the newly developed Speech-and-Sound Rubric, which consists of 15 quality indicators that will be discussed in this section. Each parameter is assigned equal weight for simplification purposes, though this can be tailored to the client's needs, priorities, and project requirements. Each quality indicator is assigned an acceptability score from 0 to 10 and an overall average score can also be calculated at the end. If a specific quality indicator does not apply to a given project, a full score of 10 can be assigned, provided that the final average is calculated over 15 indicators. The suggested rating scale is 0–6: Action required, 7–8: Enhancements recommended, 9–10: No action required. However, this is customizable. It is at the discretion of the model user to define their desirable acceptability rating criteria.

The specific scores attributed to individual indicators serve as valuable tools to pinpoint areas that require attention or specific actions. While individual indicator scores are crucial to identify and address issues, the overall rating (the sum of all the scores divided by 15) can take precedence. Even if some indicators have low scores while others have higher scores, no immediate action may be necessary if the overall rating is deemed satisfactory. Assessing the overall rating proves beneficial, especially for comparative analyses. Comparing overall ratings from different team members or suppliers, products, or components of the same serial production enables the identification of potential challenges, areas for improvement, or the need for further professional development. Likewise, comparing overall ratings of the same product calculated by different evaluators can reduce the degree of subjectivity.

SPEECH	**Quality indicator**	**Descriptors**	**Acceptability Score (0–10)**
1.	Voice Symbiosis	- Suitable voice casting according to age, gender identity, physique du rôle, characterization, narrative-related features - Suitable voice qualities in terms of range (e.g., mid-to-low, high, mid-range female), pitch, timbre, adequate volume rises	
2.	Vocal Output	- Recording quality (including, mic and home station equipment considerations) - Voice quality in terms of naturalness and credibility (the way the voice sounds) - Warmth and naturalness of human voice - Continuity and consistency of character voice and style - Sufficient voice varieties - Avoidance or minimization of synthetic qualities (e.g., derived from voice synthesis, or pitch adjustments, e.g., in the case of adults dubbing children) - Suitable degree of synthetic qualities if deliberate	

3.	Performance & delivery	- Convincing role interpretation (conveying characterization & emotions & intensity through voice dynamics) - Natural emphasis (elongated or stressed syllables) - Natural tone and intonation (balance between extreme dubbese and over-domestication) - Natural speech melody, target-appropriate rising and falling tones, avoidance of monotone intonation - Clear diction (articulation) - Voice projection - Correct pronunciation of proper nouns, specialized jargon, foreign language utterances - Suitable language variation (accents or flavour) - Sufficient/adequate non-verbal sounds and reactions
4.	Body Language	- Speech-to-face correspondence: Semantic correspondence and synchrony between utterances and facial expressions - Speech-to-body correspondence: Semantic correspondence and synchrony between utterances and body gestures
5.	Timing	- Adequate synchronization of target language audio track to visuals in terms of cueing in and out of utterances, and duration - Genre-appropriate sync (e.g., voice-over TL deliberate lag, phrase sync, or lector dubbing cues & pauses) - Pauses & pace if applicable
6.	Lip Sync	- Matching lip articulatory movements (labial consonants and lip rounded vowels, semi consonants) - Internal speech tempo/rhythm (matching mouth flaps)

7.	Narrative cohesion	- Narrative integrity/holistic storytelling experience - Cohesive, seamless flow between dialogue exchanges and character interactions; smooth and well-connected dialogue events or speech utterances or narration - Logical sequence of utterances and lack of ambiguity and disjointed exchanges - Consistency and continuity across scenes or episodes of the same serial production; plot progression - Consistent characterization - Understanding of context, e.g., determining what to render in the target language or source language - Suitable attribution and distribution of multilingual words or utterances (different characters or same-character speech)
8.	Translation	- Fidelity to original creative intent - Faithful rendering/translation - Cultural appropriateness (e.g., honorifics or culture-bound items or idiomatic expressions in target version) - Accurate use of specialized jargon - Appropriate rendering of sensitive content and inclusive terms, expressions and pronouns as per the narrative or characterization and creative intent - Evidence of accurate transcription from original source when applicable
9.	Language	- Natural-sounding dialogue (e.g., avoidance of source calques), spontaneous spoken discourse and interjections, well-balanced *dubbese* features - Suitable linguistic style and registers - Appropriate use of slang, colloquialisms, or dialects - Correct use of language, grammar, and syntax or lack thereof where applicable (e.g., if character-appropriate) - Consideration of phonaesthetics, overall pleasant-sounding speech, that is avoidance of cacophonic sounds (consonant clusters, hissing sounds, annoying repetition, unintended rhyme)

10.	Wider visual & aural context	- General semantic correspondence between dubbed speech utterances and overall visuals and sound (e.g., reactions of other characters to dialogue; canned laughter) - General correspondence between speech output (dubbed or retained original) and visible on-screen mouth movements (technical), that is, avoidance of missing dialogue - Correspondence between speech utterances/ dubbed output and on-screen graphics, forced narratives, or subtitles in the target version
SOUND		
11.	Volume balancing	- Well-blended volume levels across newly recorded tracks - Well-blended volume levels between newly recorded voice tracks and original voice tracks - Avoidance of unwanted original dialogue bites - Suitable volume levels for overlapping speech - Suitable volume levels voice-over/lector & original voice tracks
12.	Camera shots	- Suitable adjustment of volume levels and voice positioning based on camera shots in terms of distance (long shots versus medium and close-up shots) and camera angle (over the shoulder or profile view, or other)
13.	Background murmur & M/E	- Sufficient depth to general audio conveyed through ambient sounds - Sufficient crowd murmur density - Adequate balance between background and foreground dialogue and noise - Well-blended music and effects track

14.	Room tones & effects	- Voice in spatial context, that is, the narrative setting - Credibility in specific ambiance achieved through general acoustic effects - Room type/venue adaptability - Outdoor/indoor space adaptability - Avoidance of recording studio tone (suitable levels of echo, reverberation, or lack thereof where applicable) - Recording environment considerations in the case of home station recording - Adequate addition of filters and effects for voices in another spatial context (differing from that of the camera) TV, radio, computer, phone, or behind physical barriers. Addition of filters and effects. - Lack of unnecessary noise or interference picked up involuntarily in the recording studio (e.g., script rustling, hitting microphone, unnecessary pop)	
15.	Voice transition	- Smooth transition in the case of different alternating voices for same-character multilingual utterances	
	Overall SCORE (0–10)	Addition of all scores ÷ 15	

Table 5. Script-and-Sound Rubric

The first ten quality indicators focus on speech-related elements, while the last five address sound in general. Table 5 includes detailed descriptors, some of which may overlap due to their interconnected nature. The descriptors are not meant to be measured individually; they are intended to further explain and clarify each quality indicator. Despite not requiring in-depth scrutiny on a micro level, as with the textual parameters, this rubric attempts to provide a sufficient level of detail to facilitate the identification and resolution of glitches or issues.

The rubric is applicable to both dubs and voice-overs, making some quality indicators more relevant than others based on the modality. For example, lip sync, voice transitions, and background murmur may not apply to voice-overs, whereas timing and vocal delivery would. As outlined earlier, a score of 10 can be assigned to quality indicators not applicable to a particular product, provided this approach is used across all voice-over products. This ensures consistency and enables comparative analysis and quality benchmarking across different product types and modalities within the same company. It is important to recognize that voice-over products may not solely involve traditional voice-over narration or UN voice-over style but often also encompass hybrid modalities, such as combining both voice-over and lip-sync dubbing within a single product.

The relevance of specific quality indicators can also differ depending on whether the output is human-generated or AI-generated. For instance, descriptors related to naturalness, emotion, warmth, and consistency in voice and style tend to be more critical for AI-generated output compared to human recordings. Regardless of whether the dub is produced by humans or machines, the goals of achieving credibility, authenticity, and creative intent remain the same. Therefore, the same rubric can be applied to both, with features that address specific challenges and issues more frequently encountered in AI dubs, as revealed in the model's pilot application by the researcher in a didactic context. These include speech-to-body correspondence, narrative cohesion, contextual understanding, continuity, coherence and consistency in dialogue, characterization, and narrative flow.

For a better understanding of the quality indicators, a brief description of each one is provided below:

- ***Voice Symbiosis*** emphasizes the appropriate attribution of voice qualities to speakers in the original product. Voices must align with age, gender identity, physique du role, and characterization. Matching voice qualities implies a suitable voice range (e.g., mid-to-low, high, mid-range, low range), pitch, timbre, and adequate volume rises.
- ***Vocal Output*** refers to the degree of naturalness and credibility achieved, whether the voice conveys warm or natural human tones, or if synthetic attributes emerge (unless deliberate). Synthetic qualities could arise from AI voice synthesis or intentional editing (e.g., pitch shifting when adults dub children).
- ***Performance and Delivery*** involve voice interpretation, relying heavily on the actor's ability to embody the character through their voice. This includes conveying dramatic and emotional dynamics, dependent on natural intonation, speech melody (rising and falling tones), emphasis and also adequate voice projection. Clear diction (pronunciation and articulation) is crucial for comprehension. The ability to apply specific accents or flavors significantly contributes to the performance. Enriching the performance with necessary reactions and non-verbal sounds per the visuals is equally important.
- ***Body Language*** refers to the semantic and synchronous correspondence between facial expressions, body gestures, and the uttered target-language speech.
- ***Timing*** is crucial, implying appropriate levels of speech cueing and duration based on the specific modality. In voice-over, slightly delaying the voice ensures technical accuracy. Timing depends on various professional roles, including script adapters, voice talents, and sound technicians. Issues in timing can be traced back to the script, performance, pace, or technical glitches during voice track placement and movement.
- ***Lip sync*** refers to the lip articulatory movements, generally entailing matching bilabial consonants, fricatives, lip-rounded vowels, and semiconsonants. Ensuring the same mouth flap recurrence is paramount and depends on the internal speech tempo of every phrase.

The next three quality indicators are heavily script-dependent, though still focused on the reception and perception of the viewer when watching the final dubbed product. If most text-related parameters do not achieve a sufficient acceptability score, the Script Rubric, which centres on a micro-analysis of textual parameters, can help pinpoint specific issues and address them.

- ***Narrative cohesion*** is a crucial quality indicator, especially when considering AI-generated dubs, to ensure a seamless and engaging viewing experience. It hinges on maintaining narrative integrity and a holistic storytelling experience through consistent plot progression. This involves facilitating a cohesive flow between dialogue exchanges and character interactions, ensuring smooth and well-connected dialogue events or narration. A logical sequence of utterances with minimal ambiguity or disjointed events is vital for maintaining coherence. Additionally, maintaining consistency and continuity across scenes or episodes within the same serial production, including consistent characterization, plays a pivotal role. Understanding the context and determining what to render in the target or carry over in the original language is essential. Moreover, suitable attribution of multilingual utterances among different characters or within the speech of the same character contributes to the overall narrative cohesion.

- The ***translation*** aspect emphasizes staying true to the original creative intent by rendering the content in a way that is both understandable and meaningful in the new cultural context. Depending on the product and genre, a high level of accuracy and equivalence may be necessary, or otherwise, a certain degree of adaptation may be needed to achieve the desired impact and offer culturally appropriate solutions. Special attention should be given to the precise use of specialized terminology. It is also crucial to employ sensitive and inclusive language that suits the plot, context, and target audience, while at the same time preserving the original creative essence. The evaluation of this quality indicator is based solely on the target output, adopting a perceived quality approach, and thus reflecting the viewer experience. Reference to the original source text or video can be made only if and when necessary but not as a default comparative approach throughout that would significantly slow down the overall assessment process.
- ***Language*** concentrates on the technical elements of language, ensuring correctness, sensitivity, and stylistic appropriateness. It considers linguistic precision in terms of register, grammar, slang, colloquialisms, and use of dialects (where needed). Some dubbing cultures also value phonaesthetics, aiming to avoid discordant sounds, repetitive patterns, or unintended rhymes. This quality indicator also focuses on how believable and authentic the dialogue sounds. While a natural-sounding intonation relies primarily on the actors' performance, achieving authentic-sounding dialogue is primarily dependent on the quality of the script. In the realm of fiction, viewers typically expect a level of naturalness that may not align with everyday speech patterns in real life. This expectation, ingrained through viewers' habitual consumption of media, sits within the acceptable range of dubbed content, which may inherently carry an artificial register, the so-called dubbese register. A carefully weighed balance of dubbese features is essential for the dialogue to sound natural and find the right position on the spoken-written continuum. Even in its original form, film dialogue reflects a scripted orality, designed for spoken delivery, a prefabricated orality (Baños-Piñero and Chaume, 2009; Baños-Piñero, 2024). This consideration may hold less significance for genres like documentaries, live TV, interviews, or reality shows, where techniques like voice-over or phrase-sync dubbing are commonly employed.
- The final aspect to consider in the speech category is the wider ***visual and aural context***, which plays a role in ensuring alignment between dubbed speech and the overall visual elements. This alignment extends beyond just matching body language which is a separate parameter; it includes reactions of other characters and elements displayed on screen, such as on-screen graphics or forced narratives, or any other elements in the images. Discrepancies, such as missing speech, can disrupt the harmony between visual and audio components, impacting the audience's viewing experience. These discrepancies fall outside the realm of translation, as they can also be attributed to technical glitches rather than linguistic choices, while still influencing how the content is perceived by viewers. Additionally, this rubric focuses on the viewers' perspective, therefore missing speech is perceived as a lack of visual and aural correspondence.

The ***sound component*** focuses on post-recording elements, at least in the case of in-studio dubbing, whilst in AI-generated dubs there may not be a recording process depending on the type of tool and its features. Meticulous attention is given to several key parameters to ensure a seamless aural and viewing experience.

- **Volume balancing** is paramount, requiring a harmonious blend of volume levels not only within newly recorded tracks but also between these tracks and the original voice recordings. This careful balancing also applies to overlapping speech and voice-over or lector tracks. All this entails careful management while also preventing unwanted original dialogue interference.
- Additionally, adjustments must be made based on ***camera shots***, considering varying distances (e.g., long shots versus close-ups) and angles (e.g., over-the-shoulder shots) to optimize voice positioning and volume levels.
- ***Background murmur*** and general audio depth play a crucial role, necessitating a balanced mix of ambient sounds, crowd murmur density, and a coherent blend of background dialogue and noise with music and effects tracks.
- ***Room tones and effects*** further contribute to the immersive experience, grounding voices in their spatial context and establishing credibility through specific ambiance effects. Adapting to different room types or outdoor/indoor settings demands versatility while avoiding a sterile studio sound, instead incorporating appropriate levels of echo and reverberation. Filters and effects also need to be applied to simulate varied spatial contexts like TV, radio, or also physical barriers (e.g., characters talking through a glass door), enhancing the narrative's realism by giving depth to the audio track. Eliminating unwanted studio noise or interference that could detract from the final product is also essential.
- ***Voice transitions*** (where applicable) add another layer of complexity, requiring seamless track shifts between different voices for multilingual utterances by the same character.

By meticulously addressing volume nuances, spatial considerations, ambiance authenticity, and continuity in voice delivery, the sound component in dubbing plays a crucial role in crafting a cohesive and engaging audiovisual experience for audiences.

4. Application and integration into workflows, tools and training

Exploring the practical applications of the SSS model and its rubrics reveals its potential use in various areas.

4.1. Professional workflows

The SSS Model could serve as a resource for professionals within the dubbing industry, including translators, adapters, reviewers, QC specialists, project managers and software developers. It can be used to integrate QC features into a professional workflow or platform since it offers a structured framework to identify areas requiring improvement. The ultimate goal is to uphold the necessary quality standards required to engage an audience. The scoring systems provide a tool to measure dubbing quality, setting benchmarks for project acceptability and facilitating comparisons across different projects. This comparative analysis can unveil areas necessitating enhancements, potentially linked to team competency, the need for ongoing training, client-specific requirements, or nuances related to different project types. Furthermore, the model can serve as a ranking system for recruitment purposes, aiding in the evaluation of adapters, translators, and other professionals.

Within professional workflows, the Script Rubric can optimize QC processes by introducing early script-related quality assessment before recording or producing the AI voice-over or dub. In the case of generative AI output, depending on the tool used and the types of post-QC adjustments implemented, regenerating the dubbed content can lead to significantly different outcomes. This scenario would mandate a thorough review check to ensure that any glitches newly appearing in other areas, previously considered satisfactory, are promptly addressed.

Integrating the Script Rubric within cloud-based dubbing platforms can facilitate the review process, enabling efficient error identification and quantification by reviewers or project managers. The individual parameters or their tags can be integrated (possibly as a drop-down menu) in the individual dialogue events to rapidly flag issues in their specific location. That said, both rubrics can also be used as manual easy-to-use templates or mere checklists.

Conversely, the Speech-and-Sound Rubric can be applied once the product has been recorded and dubbed, but before finalization and delivery. This reflects common current quality control (QC) workflows, which conduct technical and linguistic evaluations at this stage. Although this approach may not be the most time and cost-effective, it ensures a more accurate assessment. Introducing an additional linguistic and technical script assessment (Script Rubric) prior to recording or machine generation could potentially reduce the occurrence of errors and subsequent adjustments. This could prevent scenarios where actors must return to the studio to re-record dialogue, which can incur extra time and costs and potentially strain delivery deadlines. Incorporating QC twice in the workflow naturally involves additional roles and processes; however, if it mitigates time and cost burdens further down the chain, it may prove worthwhile. A challenge with pre-output script assessment is the need for highly specialized evaluation skills, as discussed in the next section.

Another (more convoluted) alternative aimed at optimizing time and cost-effective QC processes is to break down the Speech-and-Sound Rubric and distribute its various quality indicators among the pertinent professional roles involved. This would offer the possibility of a *pre-output process-oriented QC process* carried out by multiple players in the workflow of a project, each one responsible for an individual quality indicator or more than one. For instance, the performance and synchronization-related parameters could be controlled by the dubbing directors, creative leads, or assistants, the script-related parameters could be controlled by linguists or adapters while the sound-related and vocal-related parameters could be controlled by sound technicians, each group using their relevant rubric quality indicators as a checklist once that specific stage in the workflow has been completed.

4.2. Evaluation techniques

The Speech-and-Sound Rubric involves conducting random sampling of the product, depending on its duration, and can focus solely on the target language version to speed up the process and reflect the viewers' experience. Incorporating multiple reviewers can mitigate subjective biases, considering the perceived quality approach. On the other hand, in the case of the Script Rubric and its textual parameters, a thorough check across the entire dialogue list against the original content is recommended. Given that the Script Rubric is ideally intended to assess the pre-dubbed adaptation, the key difficulty lies in honing the skills required to perform both linguistic and technical assessments of the script independently of the recording, particularly in identifying synchronization issues by relying only on the script. In this case, one effective technique involves having the script reviewer or QC specialist test the target language-adapted speech by reciting it alongside the original video (while varying the volume levels), simulating the process as an adapter. This requires highly specialized skills, therefore providing training for QC specialists, script editors, or post-editors to develop these skills, especially if they are not script adaptation professionals, is crucial. For cloud-based systems that host and facilitate the script adaptation process, a potential workaround involves granting script adapters and reviewers access to the platform's remote recording tool used by voice actors for individual dialogue events. This will enhance their ability to evaluate and fine-tune scripts effectively. The

SSS model and its rubrics can serve as guiding tools to support any QC process, regardless of whether they are used as scoring tools.

4.3. Training

In educational and training settings, the proposed model offers a pedagogical tool to develop QC, script editing, and speech-and-sound post-editing skills essential for emerging roles in AI-driven dubbing environments. Drawing on the SSS model, *speech-and-sound post-editing* refers to modifying, refining, rewriting, reworking, or recreating the speech and sound components of AI-generated dubs, thus incorporating both technical and linguistic revisions. Trainees and students in audiovisual translation studies can practice analyzing and rating AI-generated dubs using the Speech-and-Sound Rubric and its scoring system as part of their training. Comparing ratings of the same output could serve as a useful exercise. Specifically for script translation and adaptation, trainees can use the Script Rubric as a checklist and a tool for self-evaluation or self-revision. The rubrics' versatility extends to in-studio training and corporate contexts, aligning educational initiatives with industry demands for adaptable skill sets required to navigate the rapidly changing landscape of dubbing technologies. The SSS model offers audiovisual translation students the opportunity to broaden their skill set to include the ability to critically analyze and assess the entire product, not just the linguistic aspects typically concerning them. This know-how and supporting didactic tools are increasingly relevant in an industry reshaped by AI technologies, which require traditional roles to evolve and pave the way for new profiles demanding versatile skills and a broader knowledge base. The SSS model has already been applied and tested by the researcher in a university training setting and the findings from the experiment will be presented in a separate paper.

5. Conclusions

The Script, Speech and Sound (SSS) Quality Assessment Model recognizes the importance of evaluating not only individual script adaptation elements, but also the overall impact of the dubbed product on the viewer. It incorporates two distinct yet complementary scoring rubrics. The Script Rubric focuses on textual parameters, employing a granular, error-based approach for a thorough assessment against the original script, and is outlined in Section 2. The Speech-and-Sound Rubric takes a viewer-centered perspective, employing an acceptability score system to evaluate the quality of speech and sound elements in the dubbed product.

These rubrics serve as versatile tools applicable to both in-studio dubs, voice-overs, and AI-generated outputs, as they all pursue a common goal of ensuring quality. The rubrics have been developed through meticulous analysis, drawing on first-hand experience in dubbing and voice-over, combined with research and evaluation of both in-studio and AI-generated dubbed content. This led to the identification of prevalent patterns, areas of improvement, strengths, and weaknesses that were methodically organized, labeled, and incorporated into the rubrics. Valuable data was generated by systematically comparing outputs from in-studio and AI-generated dubbing processes for the same content, as well as from the first pilot attempts at applying the model in training contexts. As previously noted, this data and the findings will be shared in future publications.

It is essential to note that this quality assessment model has its limitations and is not intended as a definite or exhaustive solution, but rather an evolving framework that will require further refinement in response to its application as well as ongoing industry developments, particularly within the rapidly changing landscape of AI dubbing. As the AI revolution continues to unfold, revealing new advancements and challenges, professionals, trainers, and researchers will need to adapt accordingly. Meanwhile, this model is presented as a flexible tool for industry practitioners, educators, and learners in the field. It offers customization options and the potential for continuous enhancements as industry practices evolve and new insights emerge.

6. References

Baños-Piñero, R. & Chaume, F. (2009). Prefabricated orality: A challenge in audiovisual translation. In M. Giorgio Marrano, G. Nadiani & C. Rundle (Eds.), *The translation of dialects in multimedia* (Special Issue). *inTralinea*. http://www.intralinea.org/specials/article/1714

Baños, R. (2021). Creating credible and natural-sounding dialogue in dubbing: Can it be taught? *The Interpreter and Translator Trainer, 15*(1), 13-33. https://doi.org/10.1080/1750399X.2021.1880262

Baños-Piñero, R. (2024). *La oralidad prefabricada en la ficción audiovisual original y doblada: Siete vidas y Friends*. Publicacions de la Universitat Jaume I.

Bolaños García-Escribano, A. (2025). *Practices, education and technology in audiovisual translation*. Routledge.

Caillé, P. F. (1960). Cinéma et traduction: le traducteur devant l'écran. *Babel*, *6*(3), 103–109. https://doi.org/10.1075/babel.6.3.01cai

Chaume, F. (2007). Quality standards in dubbing: A proposal. *TradTerm 13,* 71–89. http://www.revistas.usp.br/tradterm/article/view/47466/51194

Chaume, F. (2008). Teaching synchronisation in a dubbing course. In J. Díaz-Cintas (Ed.), *The didactics of audiovisual translation* (pp. 129–140). Benjamins.

Chaume, F. (2012). *Audiovisual translation: Dubbing*. St. Jerome Publishing.

Chaume, F. & De los Reyes-Lozano, J. (2021). El doblaje en la nube: la última revolución en la localización de contenidos audiovisuales. In B. Reverter-Oliver, J. J. Martínez-Sierra, D. Gonzales-Pastor & J. F. Carrero-Martín (Eds.), *Modalidades de traducción audiovisual: Completando el espectro* (pp. 1-15). Comares.

Mayoral, R., Kelly, D., & Gallardo, N. (1988). Concept of constrained translation. Non-linguistic perspectives of translation. *Meta*, *33*(3), 356–367. https://doi.org/10.7202/003608ar

Künzli, A. (2017). *Die Untertitelung — Von der Produktion zur Rezeption*. Frank & Timme.

Künzli, A. (2021). From inconspicuous to flow — the CIA model of subtitle quality. *Perspectives*, *29*(3), 326-338. https://doi.org/10.1080/0907676X.2020.1733628

Martínez Sierra, J. J. (2008). *Humor y traducción: Los Simpson cruzan la frontera*. Castellón: Universitat Jaume I.

Pedersen, J. (2017). The FAR model: Assessing quality in interlingual subtitling. *The Journal of Specialised Translation*, *28*, 210-229. https://jostrans.soap2.ch/issue28/art_pedersen.php

Romero-Fresco, P. & Martínez Pérez, J. (2015). Accuracy rate in subtitling: The NER model. In R. Baños Piñero & J. Díaz Cintas (Eds.), *Audiovisual translation in a global context: Mapping an ever-changing landscape* (pp. 28-50). Palgrave Macmillan.

Romero-Fresco, P. & Pöchhacker, F. (2017). Quality assessment in live interlingual subtitling: A new challenge. *Linguistica Antverpiensia NS: Themes in Translation Studies*, *16*, 149-167. https://doi.org/10.52034/lanstts.v16i0.454

Sánchez-Mompeán, S. (2023). Engaging English audiences in the dubbing experience: A matter of quality or habituation? *Íkala, Revista De Lenguaje Y Cultura*, *28*(2), 1-18.

Spiteri Miggiani, G. (2019). *Dialogue writing for dubbing. An insider's perspective*. Palgrave Macmillan.

Spiteri Miggiani, G. (2021a). English-language dubbing: Challenges and quality standards of an emerging localisation trend. *The Journal of Specialised Translation*, *36*, 2-25. https://jostrans.soap2.ch/issue40/art_spiteri.pdf

Spiteri Miggiani, G. (2021b). Exploring applied strategies for English-language dubbing. *Journal of Audiovisual Translation*, *4*(1), 137-156. https://doi.org/10.47476/jat.v4i1.2021.166

Spiteri Miggiani, G. (2022a). Measuring quality in translation for dubbing: A quality assessment model proposal for trainers and stakeholders. *XLINGUAE, 15*(2), 85 102. https://xlinguae.eu/files/XLinguae2_2022_7.pdf

Spiteri Miggiani, G. (2022b). The dubbing metamorphosis: Where do we go from here? *EST Newsletter*, *60*, 10.

Spiteri Miggiani, G. (2023). Quality in translation and adaptation for dubbing: Applied research in a professional setting. *The Journal of Specialised Translation*, *40*, 297-321. https://www.jostrans.soap2.ch/issue40/art_spiteri.pdf

Spiteri Miggiani, G. (2024). Translators and adapters in training: Addressing quality issues in dubbing scripts. In M. L. Poveda-Balbuena, V. Marrahi-Gómez & C. Mangiron (Eds.), *Traducción y lingüística de corpus: Avances en la era digital* (pp. 71-98). Tirant Lo Blanch.

Titford, C. (1982). Subtitling constrained translation. *Lebende Sprachen*, *27*(3), 113–116.

Whitman-Linsen, C. (1992). *Through the dubbing glass*. Peter Lang.

Yang, Y., Shillingford, B., Assel, Y., Wang, M., Liu, W., Chen, Y., Zhang, Y., Sezener, E., Cobo, L. C., Denil, M., Aytar, Y., & de Freitas, N. (2020). Large-scale multilingual audio visual dubbing. https://doi.org/10.48550/arXiv.2011.03530

Zabalbeascoa, P. (1993). *Developing translation studies to better account for audiovisual texts and other new forms of text production* (Unpublished PhD dissertation). University of Lleida. http://hdl.handle.net/10803/818

Webpages

Netflix (last update 2024), Introduction to Netflix quality control, https://partnerhelp.netflixstudios.com/hc/en-us/articles/115000353211-Introduction-to-Netflix-Quality-Control-QC

Deepmind, https://deepmind.google/

Dubly, https://dubly.ai/

Flawless, https://www.flawlessai.com/truesync

Lenseup, https://www.lenseup.com/en/

Lipdub, https://www.lipdub.ai/

Giselle Spiteri Miggiani
University of Malta
Room 308B
Old Humanities Building,
University of Malta
Msida MSD 2080
Malta
giselle.spiteri-miggiani@um.edu.mt

Biography: Giselle Spiteri Miggiani, PhD, is a tenured Senior Lecturer in the Department of Translation, Terminology, and Interpreting Studies at the University of Malta, where she introduced Audiovisual Translation as an area of study and where she teaches and coordinates this postgraduate specialization stream. She is also a professional audiovisual translator and adapter of media content since 2006, specializing mainly in dubbing, and has worked on numerous productions broadcast on RAI and Mediaset. She is invited regularly as a guest speaker, trainer, and visiting lecturer at other European universities and delivers consultancy and training to EU Institutions and leading global media localization stakeholders. She authored the book *Dialogue Writing for Dubbing – An Insider's Perspective* (Palgrave Macmillan, 2019), among other journal articles and book chapters.